

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA ELEKTROTECHNICKÁ



BAKALÁŘSKÁ PRÁCE

Klasifikace CTG záznamů

Praha, 2009

Autor: Martin Čurda

Prohlášení

Prohlašuji, že jsem svou bakalářskou práci vypracoval samostatně a použil jsem pouze podklady (literaturu, projekty, SW atd.) uvedené v příloženém seznamu.

V Praze dne 24.8.2009



podpis

Poděkování

Děkuji především vedoucímu bakalářské práce Ing. Václavu Chudáčkovi za poskytnutí odborných informací, za přátelské a ochotné jednání.

Abstrakt

Cílem práce je navrhnout řešení pro klasifikaci kardiokografických dat pomocí programu MATLAB s použitím nástroje PRTools (Toolbox for Pattern Recognition). Vybral jsem 3 druhy klasifikátorů. K-NN (K-Nearest Neighbour), stromový klasifikátor a poslední SVM (Support Vector Machine). Pro selekci příznaků jsou použity funkce *featselb* a *featselp* s použitím parametrů určených ke každému klasifikátoru. Snížení dimenze je provedeno metodou PCA (Principal Component Analysis), pouze pro experimentální využití.

Pro každé nastavení klasifikátoru a výběrové funkce je provedena křížová validace s dělením na 10 složek a s 10-ti násobným opakováním. Pouze u stromového klasifikátoru je použito právě jedno opakování.

Výsledné hodnoty jsem vynesl v tabulkách. Vybral jsem nejlepší řešení, které je uvedeno v závěru této práce.

Klíčová slova: CTG, kardiokograf, knn, knnc, svm, svc, treec, featselp, featselb, pca, testc, PRTools

Abstract

The whole point of this work is to find out a possible solution for classification of cardiographic datas using MATLAB programme and its PRTools. I have chosen three types of classifiers. K-NN (K-Nearest Neighbour), the decision tree classifier and SVM (Support Vector Machine). The functions *featselb* and *featselp* having parameters for each classifiers are used for the features selection. Transformation to lower dimension is realized by PCA. It is used for an experimental purposes only.

For each configurations of the classifier and selected functions is used cross validation that divides in to 10 folders and having 10 numbers of repetitions. The only exception comes with the decision tree classifier type which uses just one repetition.

I have chosen the best solution for this task and I have written results in conclusion of the bachelor theses.

The keywords: CTG, cardiocograph, knn, knnc, svm, svc, treec, featselp, featselb, pca, testc, PRTools

České vysoké učení technické v Praze
Fakulta elektrotechnická

Katedra kybernetiky

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Student: Martin Čurda
Studijní program: Elektrotechnika a informatika (bakalářský), strukturovaný
Obor: Kybernetika a měření
Název tématu: Klasifikace CTG záznamů


Pokyny pro vypracování:

1. Seznamte se s problematikou klasifikace CTG záznamů.
2. Na základě dodaných příznaků popisujících CTG záznamy klasifikujte tyto do tříd.
3. Použijte nejméně tři metody klasifikace - zvažte využití jejich kombinace.
4. Výsledky kriticky zhodnoťte a navrhněte pokračování/vylepšení Vašeho postupu.

Seznam odborné literatury: Dodá vedoucí práce.

Vedoucí bakalářské práce: Ing. Václav Chudáček

Platnost zadání: do konce zimního semestru 2009/2010


prof. Ing. Vladimír Mařík, DrSc.
vedoucí katedry




doc. Ing. Boris Šimák, CSc.
děkan

V Praze dne 23. 2. 2009

Obsah

Seznam obrázků	xi
Seznam tabulek	xiii
1 Úvod	1
2 CTG	3
2.1 Hypoxie plodu	3
2.1.1 Životodárný kyslík	3
2.1.2 Aerobní metabolismus	3
2.1.3 Hypoxie [12]	4
2.2 Způsoby měření	5
2.2.1 Externí monitoring	6
2.2.2 Interní monitoring	6
2.2.3 Kardiokrogram	8
2.2.3.1 Kontrakční činnost dělohy	8
2.2.3.2 Dlouhodobé frekvenční jevy	8
2.2.3.3 Střednědobé frekvenční jevy	9
2.2.3.4 Krátkodobé frekvenční jevy	9
2.2.4 FIGO 1986	10
2.3 Monitor srdeční činnosti plodu - STAN S31 TM	10
3 Teoretický rozbor	13
3.1 Strojové učení	13
3.2 Předzpracování, snížení dimenze	14
3.2.1 PCA (Principal Components Analysis)	14
3.3 Modelování	16
3.3.1 Základní popis klasifikátorů	16

3.3.2	<i>k</i> -NN (<i>k</i> -Nearest Neighbour)	18
3.3.3	Rozhodovací stromy	19
3.3.3.1	Vytváření rozhodovacího stromu	19
3.3.4	Support Vector Machine (SVM)	20
3.4	Vyhodnocení	22
3.4.0.1	Senzitivita, specifická a výkonost klasifikátorů	22
3.4.0.2	Křivka ROC	23
3.4.0.3	Křížová validace	24
4	PRTools	25
4.1	Struktura funkcí toolboxu	25
5	Praktická realizace a experimenty	27
5.1	Zpracování	27
5.1.1	Použité výpočetní zařízení	27
5.1.2	Anotace	27
5.1.3	Selekce, snížení dimenze	28
6	Vyhodnocení	31
6.1	Vypočtené hodnoty jednotlivých klasifikátorů	31
6.1.1	Klasifikátor <i>knn</i>	31
6.1.2	Klasifikátor <i>svc</i>	33
6.1.3	Klasifikátor <i>treec</i>	35
6.2	Vypočtené hodnoty pro kombinace klasifikátorů	37
6.2.1	Základní kombinace	37
6.2.2	Kombinace po výběru nejlepšího klasifikátoru	39
7	Závěr	43
	Literatura	46
A	Kompletní tabulky výpočtů	I
A.0.3	KNN	I
A.1	SVM	VI
A.2	Rozhodovací stromy	X
A.3	Základní kombinace	XIII
A.4	Kombinace nejlepších klasifikátorů	XVII

B	Výpis použitých funkcí a jednoduchá demonstrace	XXI
B.1	dataset()	XXI
B.2	pca()	XXI
B.3	featselb()	XXII
B.4	featselp()	XXII
B.5	knnc()	XXII
B.6	svc()	XXIII
B.7	kombinace pomocí mean a median	XXIII
B.8	treec()	XXIII
B.9	testc()	XXIV
B.10	testauc()	XXIV
C	Monitor srdeční činnosti plodu - STAN S31™	XXV
C.1	Funkce a technické parametry	XXV
C.1.1	Funkce zahrnuté v CTG platformě	XXV
C.1.2	Základní technické parametry	XXVI
D	Obsah přiloženého CD	XXIX

Seznam obrázků

2.1	Externí monitoring. Obrázek z článku 9	6
2.2	Interní monitoring. Obrázek z článku [9]	7
2.3	Skalповá elektroda (el. vpravo od společnosti Neoventa [7])	7
2.4	Parametry kardiogramu [10]	8
2.5	STAN S31 TM od společnosti Neoventa [7]	10
3.1	Schéma procesu vytěžování dat dle metody CRISP-DM	13
3.2	Principal Component Analysis	15
3.3	Zjednodušený model klasifikátoru	16
3.4	Příklad 2D prostoru s třídami ”+” a ”*”	17
3.5	Zobrazení prostoru 3D do 2D	18
3.6	Lineární SVM	21
3.7	Křivka ROC	23
5.1	Experiment s histogramem pH	28

Seznam tabulek

2.1	Tabulka reakcí orgánů na hypoxii	4
2.2	Interpretace fetálního CTG	10
3.1	Matice záměn	22
5.1	Technická specifikace použitého výpočetního zařízení	27
5.2	Společná tabulka příznaků a selekcí pro všechny metody.	29
5.3	Dodatková tabulka příznaků pro selekční metody s parametrem <i>knnc</i> . . .	30
5.4	Dodatková tabulka příznaků pro selekční metody s parametrem <i>svc</i>	30
6.1	<i>k</i> -NN bez použití selekce	31
6.2	<i>k</i> -NN s použitím funkce <i>featselp</i> , parametr <i>knnc</i>	32
6.3	<i>k</i> -NN s použitím funkce <i>featselp</i> , bez parametru	32
6.4	<i>k</i> -NN s použitím funkce <i>featselb</i> , bez parametru	32
6.5	<i>k</i> -NN s použitím funkce <i>featselb</i> , parametr <i>knnc</i>	32
6.6	<i>k</i> -NN, <i>pca</i> bez parametru	32
6.7	<i>k</i> -NN snížení dimenze funkcí <i>pca</i> na 2-D	33
6.8	<i>k</i> -NN snížení dimenze funkcí <i>pca</i> na 5-D	33
6.9	<i>k</i> -NN snížení dimenze funkcí <i>pca</i> na 10-D	33
6.10	SVM bez použití selekce	33
6.11	SVM s použitím funkce <i>featselb</i> , bez parametru	33
6.12	SVM s použitím funkce <i>featselb</i> , parametr <i>svc</i>	34
6.13	SVM s použitím funkce <i>featselp</i> , bez parametru	34
6.14	SVM s použitím funkce <i>featselp</i> , parametr <i>svc</i>	34
6.15	SVM, snížení dimenze funkcí <i>pca</i> , bez parametru	34
6.16	SVM, snížení dimenze funkcí <i>pca</i> na 2D	34
6.17	SVM, snížení dimenze funkcí <i>pca</i> na 5D	35
6.18	SVM, snížení dimenze funkcí <i>pca</i> na 10D	35

6.19	Rozhodovací strom bez použití selekce	35
6.20	Rozhodovací strom s použitím funkce <i>featselb</i> , bez parametru	35
6.21	Rozhodovací strom s použitím funkce <i>featselp</i> , bez parametru	36
6.22	Rozhodovací strom s použitím funkce <i>featselb</i> , parametr <i>treec</i>	36
6.23	Rozhodovací strom, snížení dimenze funkcí <i>pca</i> , bez parametru	36
6.24	Rozhodovací strom, snížení dimenze funkcí <i>pca</i> na $2D$	36
6.25	Rozhodovací strom, snížení dimenze funkcí <i>pca</i> na $5D$	37
6.26	Rozhodovací strom, snížení dimenze funkcí <i>pca</i> na $10D$	37
6.27	Kombinace bez použití selekce	37
6.28	Kombinace s použitím funkce <i>featselb</i> , bez parametru	38
6.29	Kombinace s použitím funkce <i>featselp</i> , bez parametru	38
6.30	Kombinace, snížení dimenze funkcí <i>pca</i> , bez parametru	38
6.31	Kombinace, snížení dimenze funkcí <i>pca</i> na $2D$	38
6.32	Kombinace, snížení dimenze funkcí <i>pca</i> na $5D$	39
6.33	Kombinace, snížení dimenze funkcí <i>pca</i> na $10D$	39
6.34	Kombinace nejlepších klas. bez použití selekce	39
6.35	Kombinace nejlepších klas. s použitím funkce <i>featselb</i> , bez parametru . .	40
6.36	Kombinace nejlepších klas. s použitím funkce <i>featselp</i> , bez parametru . .	40
6.37	Kombinace nejlepších klas., snížení dimenze funkcí <i>pca</i> , bez parametru .	40
6.38	Kombinace nejlepších klas., snížení dimenze funkcí <i>pca</i> na $2D$	41
6.39	Kombinace nejlepších klas., snížení dimenze funkcí <i>pca</i> na $5D$	41
6.40	Kombinace nejlepších klas., snížení dimenze funkcí <i>pca</i> na $10D$	41
A.1	k -NN bez použití selekce	I
A.2	k -NN s použitím funkce <i>featselp</i> , parametr <i>knnc</i>	II
A.3	k -NN s použitím funkce <i>featselp</i> , bez parametru	II
A.4	k -NN s použitím funkce <i>featselb</i> , bez parametru	III
A.5	k -NN s použitím funkce <i>featselb</i> , parametr <i>knnc</i>	III
A.6	k -NN, <i>pca</i> bez parametru	IV
A.7	k -NN snížení dimenze funkcí <i>pca</i> na $2-D$	IV
A.8	k -NN snížení dimenze funkcí <i>pca</i> na $5-D$	V
A.9	k -NN snížení dimenze funkcí <i>pca</i> na $10-D$	V
A.10	SVM bez použití selekce	VI
A.11	SVM s použitím funkce <i>featselb</i> , bez parametru	VI
A.12	SVM s použitím funkce <i>featselb</i> , parametr <i>svc</i>	VII

A.13 SVM s použitím funkce <i>featselp</i> , bez parametru	VII
A.14 SVM s použitím funkce <i>featselp</i> , parametr <i>svc</i>	VIII
A.15 SVM, snížení dimenze funkcí <i>pca</i> , bez parametru	VIII
A.16 SVM, snížení dimenze funkcí <i>pca</i> na $2D$	IX
A.17 SVM, snížení dimenze funkcí <i>pca</i> na $5D$	IX
A.18 SVM, snížení dimenze funkcí <i>pca</i> na $10D$	X
A.19 Rozhodovací strom bez použití selekce	X
A.20 Rozhodovací strom s použitím funkce <i>featselb</i> , bez parametru	XI
A.21 Rozhodovací strom s použitím funkce <i>featselp</i> , bez parametru	XI
A.22 Rozhodovací strom s použitím funkce <i>featselb</i> , parametr <i>treec</i>	XI
A.23 Rozhodovací strom, snížení dimenze funkcí <i>pca</i> , bez parametru	XII
A.24 Rozhodovací strom, snížení dimenze funkcí <i>pca</i> na $2D$	XII
A.25 Rozhodovací strom, snížení dimenze funkcí <i>pca</i> na $5D$	XII
A.26 Rozhodovací strom, snížení dimenze funkcí <i>pca</i> na $10D$	XIII
A.27 Kombinace bez použití selekce	XIII
A.28 Kombinace s použitím funkce <i>featselb</i> , bez parametru	XIV
A.29 Kombinace s použitím funkce <i>featselp</i> , bez parametru	XIV
A.30 Kombinace, snížení dimenze funkcí <i>pca</i> , bez parametru	XV
A.31 Kombinace, snížení dimenze funkcí <i>pca</i> na $2D$	XV
A.32 Kombinace, snížení dimenze funkcí <i>pca</i> na $5D$	XVI
A.33 Kombinace, snížení dimenze funkcí <i>pca</i> na $10D$	XVI
A.34 Kombinace nejlepších klas. bez použití selekce	XVII
A.35 Kombinace nejlepších klas. s použitím funkce <i>featselb</i> , bez parametru . .	XVII
A.36 Kombinace nejlepších klas. s použitím funkce <i>featselp</i> , bez parametru . .	XVIII
A.37 Kombinace nejlepších klas., snížení dimenze funkcí <i>pca</i> , bez parametru .	XVIII
A.38 Kombinace nejlepších klas., snížení dimenze funkcí <i>pca</i> na $2D$	XIX
A.39 Kombinace nejlepších klas., snížení dimenze funkcí <i>pca</i> na $5D$	XIX
A.40 Kombinace nejlepších klas., snížení dimenze funkcí <i>pca</i> na $10D$	XX

Kapitola 1

Úvod

Život a smrt. Dva pojmy týkající se každé bytosti. Antonyma, jejichž význam dělí tenká hranice. Život je od prvopočátku soubor se smrtí a všechny jeho etapy jsou jím provázeny.

Zadání bakalářské práce vychází z jednoho z největších souborů mezi životem a smrtí, kdy se již zmíněná hranice nebezpečně přibližuje - POROD. Při porodu totiž může dojít k nepřehlednému množství komplikací. Jednou z nich je nedostatek kyslíku v organismu plodu, odborně hypoxie. Pokud k hypoxii dojde, je plod vystaven riziku, kdy mohou nastat trvalé patofyziologické změny, dokonce i úmrtí.

V současné době máme k dispozici měřicí a predikční technologii, která otevírá bránu k dalším možnostem. Zařízení používané pro snímání dat se nazývá CTG, neboli kardiotokograf.

Příchod nových technologií však stále zvyšuje nároky uživatelů. V našem případě lékařů. S plynoucím časem se používané technologie stávají více běžnou záležitostí a je potřeba jejich zdokonalení. Nyní dokážeme měřit a zpracovávat ohromné množství dat. Nesou však všechny plnohodnotnou informaci o zkoumaném objektu? Dokážeme nalézt řešení, které v reálném čase s jistotou rozpozná aktuální stav nenarozeného dítěte?

Cílem práce je co nejpřesněji klasifikovat stavy, kdy dochází k nedostatku kyslíku plodu a eliminovat tím některá rizika a zbytečné chirurgické zákroky. K dispozici mám databázi ze 188 rizikových porodů, každý se 17 příznaky. Pokusím se tyto příznaky částečně eliminovat a provést klasifikaci. Ke klasifikaci používám metody umělé inteligence. Pomocí nástroje PRTools chci nalézt optimální selekční a klasifikační metody, které dokáží s největší přesností vyhodnotit vstupní data.

Ve druhé kapitole se zaměřím na popis CTG jak z pohledu lékaře, technika, tak neopomenu ani samotný proces zpracování dat od signálu, až po samotné výsledky.

Třetí kapitola teoreticky pojednává o použitých metodách selekce příznaků a klasifikátorech. Jako hlavní metodu selekce jsem vybral PCA (Principal Component Analysis), neboli metodu analýzy hlavních složek, dále tzv. Pudil's floating forward feature selection a Backward feature selection. Použité klasifikátory jsou k NN (Nearest Neighbour, tj. metoda nejbližšího souseda), druhý je binární stromový klasifikátor a poslední SVM (Support Vector Machine, což je druh neuronové sítě).

Ve čtvrté kapitole se zabývám nástrojem zvaným PRTools k programu MATLAB, poskytnutým od vývojářů z univerzity technologií v holandském Delftu.

Pátá a další kapitoly jsou zaměřeny na samotnou praktickou realizaci, experimenty, zhodnocení a návrh jiných řešení.

Kapitola 2

CTG

2.1 Hypoxie plodu

Co se rozumí termínem “hypoxie“? Čeho se vlatně hypoxie týká? Jaké má důsledky pro lidské tělo? Tyto otázky se pokusím zodpovědět v následující kapitole.

2.1.1 Životodárný kyslík

Kyslík je druhá nejčastější látka vyskytující se na Zemi. Pro život je natolik důležitý, že při jeho okamžitém vymizení, by společně i s ním vymizelo vše živé do několika minut. V atmosféře je kyslík zastoupen 21 %. Je přítomen ve vodě (kde se váže se dvěma molekulami vodíku), nachází se téměř ve všech horninách, tříatomový kyslík (ozon) najdeme ve stratoféře zemské atmosféry, kde chrání pozemský život před škodlivým ultrafialovým zářením. A konečně, jeho reakčností je lidstvo fascinováno již od dávných dob (hoření).

Kyslík je jedním z nejdůležitějších biogenních prvků, a z toho vyplývá i vážnost situace, kdy tak důležitá látka chybí. Nedostatkem kyslíku v lidském těle vzniká hypoxie. Pokud přidáme i jeden z nejsložitějších fyziologických procesů v lidském životě, uvědomíme si, do jakých důsledků může hypoxie plodu dojít.

2.1.2 Aerobní metabolismus

Lidské tělo pro svou činnost (pohyb, myšlení.....), potřebuje energii. Tu získává velmi složitou biochemickou cestou, tzv. aerobním metabolismem. V podstatě se jedná o rozklad živin (glukózy) v součinnosti s kyslíkem a vznikem „buněčného paliva“ - sloučeniny

adenosintrifosfátu (ATP). V této makroergní sloučenině si lidské tělo uchovává energii do doby, než ji využije. Celý tento proces může proběhnout pouze v přítomnosti kyslíku.

2.1.3 Hypoxie [12]

Distres plodu (hypoxie), je velmi závažný stav, způsobený nejčastěji poruchou transplacentární výměny plynů. To může mít za následek těžké poškození centrálního nervového systému (hypoxicko – ischemická encefalopatie), v těžkých případech i smrt plodu.

V podstatě se jedná o nedostatek O_2 v krvi (hypoxémie) a nadbytek CO_2 (hyperkapnie).

Tabulka 2.1: Tabulka reakcí orgánů na hypoxii

	Reakce
plíce	perzistující plicní hypertenze aspirace mekoniuma pokles sulfaktantu
ledviny	oligurie ischemické poškození či nekróza proximálních tubulů (tj. akutní renální selhání)
kardiovaskulární systém	snížený srdeční výdej hypotenze šok
gastrointestinální	nekrotizující enterokolitida hepatální dysfunkce
metabolické	metabolická acidóza hypoglykémie hypokalcémie hyponatrémie
hematologické	trombocytopenie

Prvními důsledky hypoxie jsou zejména snížená spotřeba kyslíku, bradykardie, snížení až vymizení pohybů plodu, anaerobní glykolýza (v jejím průběhu se ukládá látka zvaná laktát, čímž dochází ke vzniku metabolické acidózy). Všechny tyto důsledky mají za úkol jedno, a to udržet neměnnou hodnotu homeostázy. Tyto reakce, uskutečněné podnětem

ze vznikající hypoxie se nazývají kompenzační mechanismy plodu.

Pokud je hypoxie natolik silná, že i kompenzační mechanismy plodu nedokáží zachytit, klesá průtok krve všemi orgány. Je-li tento stav dlouhodobějšího rázu, začínají se objevovat patologické změny v orgánové struktuře.

Největší poškození však hypoxie způsobuje v mozku. Jedná se o hypoxicko – ischemickou encefalopatii. Vzniká edém mozku, dále dochází k primární nekróze neuronů – ta je způsobena zejména nahromaděnou kyselinou mléčnou v důsledku anaerobního metabolismu. Sekundární nekróza neuronů vzniká poškozením buněčných membrán (přebytečné kalcium aktivuje enzymy štěpící bílkoviny a tuky - protéázy a lipázy).

Dalším stadiem hypoxie plodu je bohužel už jen smrt.

2.2 Způsoby měření

Jak již bylo řečeno v úvodu, těhotenství i porod patří mezi rizikové momenty v životě daného jedince (v tomto případě jedinců dvou – matky a dítěte). Proto je nutné, aby matka docházela na pravidelné kontroly na gynekologické oddělení, kde zkontrolují stav matky i dítěte.

V dnešní době se k vyšetření rodiček používá dosti přesných přístrojů, bez kterých by se lékaři neobešli. Dovolím si krátce pohovořit o vyšetřeních v průběhu porodu.

Při příjmu rodičky se jednak provádí zevní a vnitřní gynekologické vyšetření, ale i vstupní CTG s asi dvacetiminutovým záznamem. Lékař sleduje stav rodičky i stav plodu. Rodičku je nutné kontrolovat v průměru každé dvě až tři hodiny ošetřujícím lékařem, aby byla nastavena správná medikace.

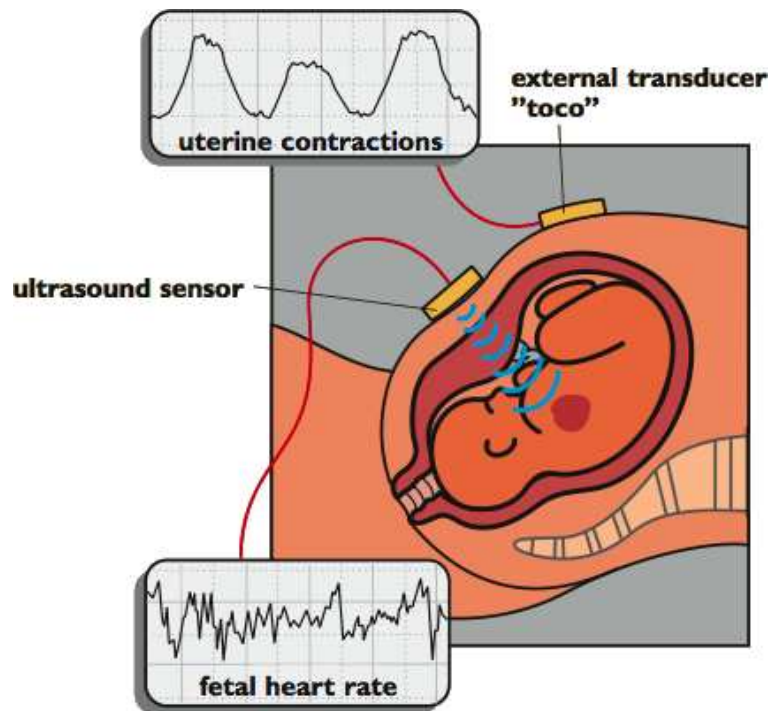
V dnešní době je kardiokografie zcela běžné vyšetření rodičky, které umožní sledovat srdeční frekvenci plodu (fetal heart rate), ale i stahy a četnost stahů dělohy matky (uterine contractions).

K vyšetření se používá přístroj zvaný kardiokograf – jeho křivka je kardiokogram. Tu můžeme rozdělit do dvou skupin – kardiokogram (srdeční činnost plodu) - hodnotíme frekvenční jevy dlouhodobé, střednědobé a krátkodobé - a tokogram (křivka děložní činnosti)

2.2.1 Externí monitoring

Neinvazivní CTG metoda užívaná obvykle kolem 36. týdne gravidity (informace z gynekologicko-porodnického oddělení, Nemocnice České Budějovice) se provádí jako preventivní vyšetření. Přesné termíny určuje lékař. Používá se širokopásmového ultrazvuku pro snímání srdeční tepové frekvence (ultrasound sensor) a externí sondy, mechanicko elektrického převodníku pro snímání stahů (external transducer "toco"), citlivé na změny napětí břišního svalstva (obrázek 2.1). Citlivost vnějšího převodníku (TOCO) je úměrná tloušťce břišní stěny rodičky.

Bohužel je zde mnoho problémů s kvalitním záznamem. Například nedopatřením dojde k dislokaci elektrody, což vede k nesprávnému snímání plodu. V případě větších artefaktů je nutné opakovat měření, což znamená větší časovou náročnost. Při porodu dvou a více plodů, je možné přímo rozeznat o který plod se jedná, použitím dvou snímačů.

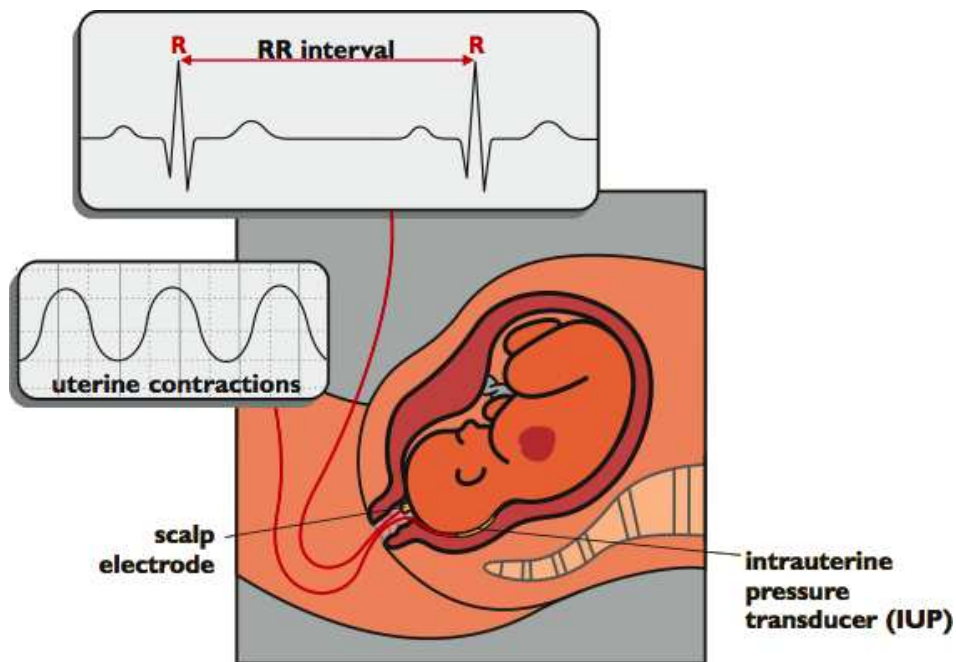


Obrázek 2.1: Externí monitoring. Obrázek z článku 9

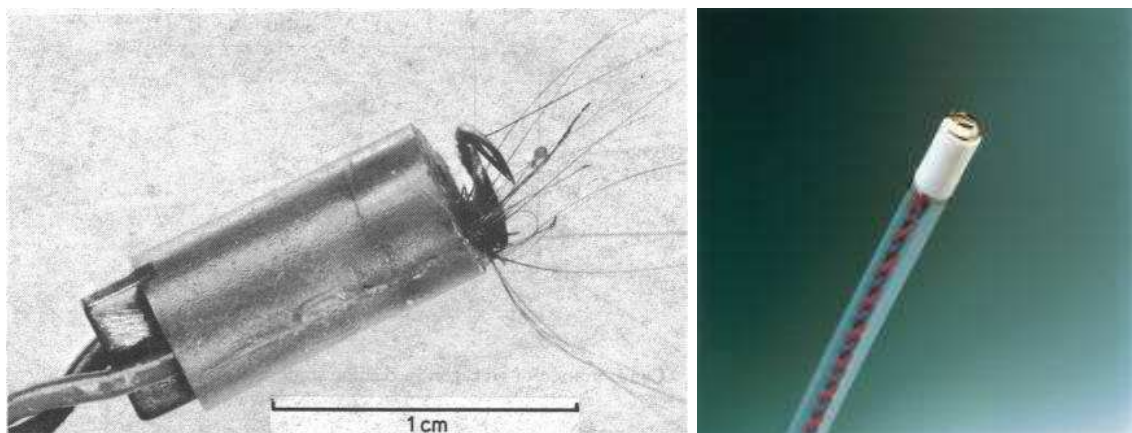
2.2.2 Interní monitoring

Mnohem přesnější metoda je interní monitoring, který se ale může použít pouze v případě, kdy plodový obal je již perforovaný a děložní hrdlo je širší než 1 až 2 cm. Jde

o zavedení skalpové elektrody (scalp electrode, obrázek 2.3) a nitroděložního tlakového snímače (IUP). Skalповá, bipolární elektroda s jemným kovovým závitem se "šroubuje" do hlavičky plodu. V případech kdy je plod otočený, můžeme tuto elektrodu uchytit do zadečku. Pro zavedení nitroděložního tlakového snímače se používá katetr. Hlavní nevýhodou metody je fakt, že hrozí riziko infekce vlivem poranění v průběhu zavádění elektrod a snímačů. Proto se používá jen tehdy, kdy vyžadujeme vysokou přesnost a kde riziko je převáženo přínosem. Tato situace se týká předpokládaných rizikových porodů.



Obrázek 2.2: Interní monitoring. Obrázek z článku [9]

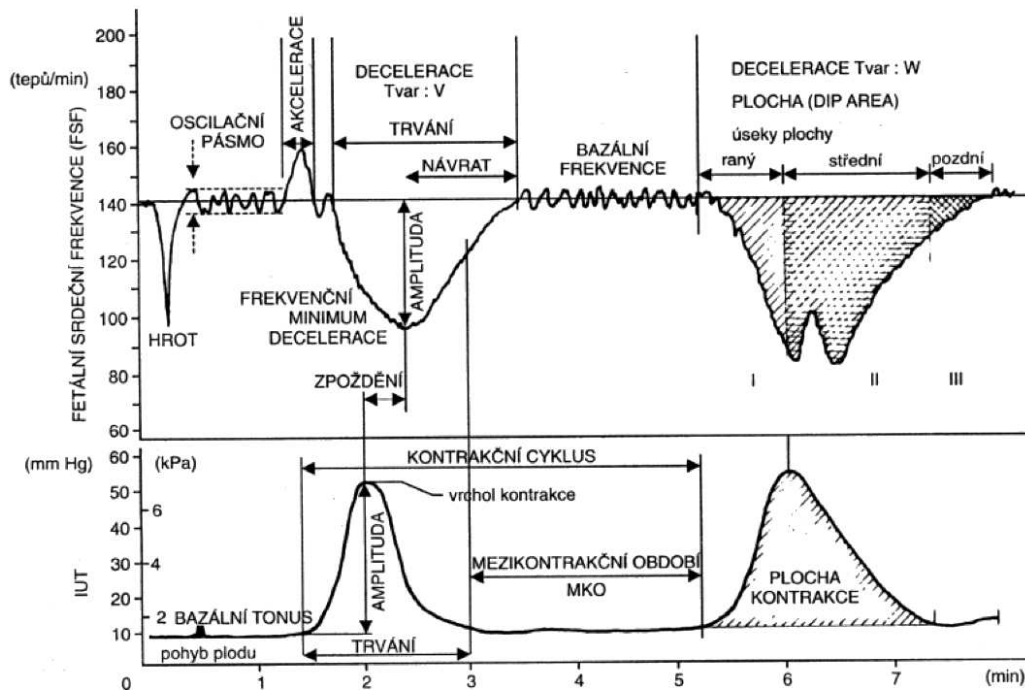


Obrázek 2.3: Skalповá elektroda (el. vpravo od společnosti Neoventa [7])

2.2.3 Kardiokogram

- tokogram - kontrakční činnost dělohy
- kardiogram - frekvenční jevy dlouhodobé, střednědobé a krátkodobé

(pokud neuvedu jinak, budu vycházet z článku [11])



Obrázek 2.4: Parametry kardiokogramu [10]

2.2.3.1 Kontrakční činnost dělohy

Hodnotíme bazální tonus dělohy, délku kontrakčního cyklu, délku mezikontrakčního období, amplitudu kontrakce (nenese informaci o intenzitě kontrakcí) a trvání kontrakce.

2.2.3.2 Dlouhodobé frekvenční jevy

Charakterizují difference bazální srdeční frekvence (dále jen BF, průměrná hodnota srdeční frekvence bez decelerací a akcelerací)

- Normokardie: BF 110-150/min
- Tachykardie

- lehká: zvýšení BF na 150-170/min
- těžká: BF > 170/min
- Bradykardie - délka záznamu pro určení bradykardie minimálně 3 minuty
 - lehká: pokles BF na 110 - 100/min
 - těžká: BF pod 100/min

2.2.3.3 Střednědobé frekvenční jevy

Jedná se o přechodné frekvenční zrychlení, akceleraci, nebo zpomalení, deceleraci.

- akcelerace: zvýšení BF o 15/min po dobu 15s až 3min.
 - reaktivní - pohyb plodu, mechanické, akustické a termické podněty
 - periodické - přítomnost zrychlení spolu se sekvencí minimálně tří kontrakcí
 - jako součást variabilní decelerace - podmíněny kontrakcí pupečníku při kontrakci
- decelerace: snížení BF o 15/min po dobu 15s až 3min.
 - sporadické - nezávislé na kontrakcích. Spike(hrot) < 10s
 - periodické - závislé na kontrakční zátěži plodu
 - * rané
 - * pozdní
 - * variabilní

2.2.3.4 Krátkodobé frekvenční jevy

Variabilita je charakterizována amplitudou oscilací kolem BF. Vyhodnocujeme průchody nulovým bodem bazální frekvence. Normální stav je více než 6 oscilací/minutu při 10-20 úderech/minutu. Pokud je 2-5 průchodů nulou 2-5/min, 10-15 úderů/minutu a křivka má tvar sinusoidy, bavíme se o zvláštním druhu variability.

Jedná se o chybějící kontrolu CNS, chonickou hypoxii a terminální stav plodu. Vyžaduje okamžitý zákrok porodníka.

2.2.4 FIGO 1986

Frekvenční jevy	Fyziologický záznam	Suspektní záznam	Patologický záznam
Bazální frekvence	110 – 150	110 – 100 150 – 170	< 100 > 170
Oscilační pásmo	10 – 15	5 – 10 > 25	< 5 sinusoida
Průchody "nulou" /min	6	2 – 6	< 2
Akcelerace/20min	2 a více	periodické	žádná
Decelerace	žádná (spike)	sporadické	periodické

Tabulka 2.2: Interpretace fetálního CTG

2.3 Monitor srdeční činnosti plodu - STAN S31™

STAN S31™ (obrázek 2.5) slouží pro intrapartální monitoring srdečních aktivit plodu při porodu a je vybaven všemi funkcemi pro standardní FHR. Tento systém je vyvíjen na velmi flexibilní platformě, která poskytuje celou řadu možností. Funkce zajišťují co možná nejlepší řešení pro fetální a maternální sledování.



Obrázek 2.5: STAN S31™ od společnosti Neoventa [7]

K dispozici máme jednu jedinečnou funkci, ST analýzu. ST analýza je založena na změnách ve fetální vlně EKG a poskytuje další informace o stavu plodu v průběhu porodu jako doplněk k záznamu z FHR.

Tato jedinečná kombinace ST analýzy a CTG poskytuje přesnější informace o plodu a pomáhá při identifikaci plodů s nedostatkem kyslíku.

Funkce a technické parametry přístroje jsou uvedeny v příloze C.

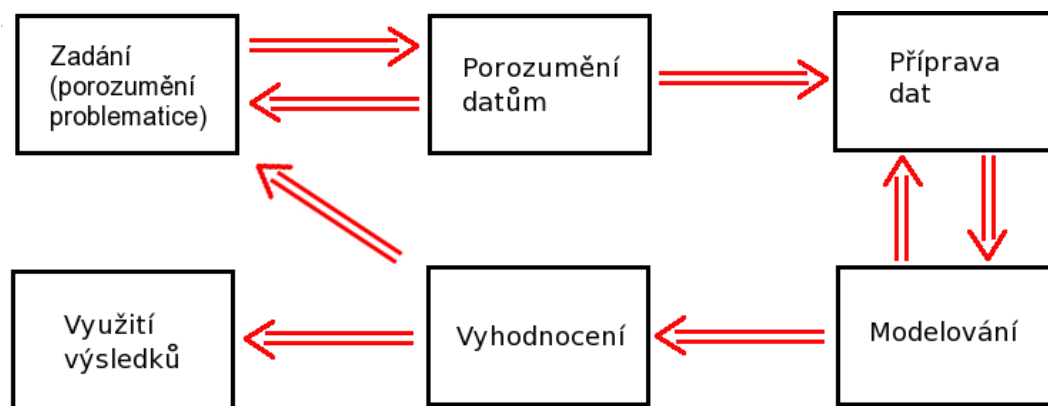
Kapitola 3

Teoretický rozbor

3.1 Strojové učení

V kapitole strojového učení se budu zabývat jednotlivými kroky procesu vytěžování dat a pokusím se alespoň částečně vysvětlit vybrané metody.

V roce 1997-1999 byl zaveden standard konsorciem evropských firem v rámci projektu vedeného k návržení průmyslového standardu metodiky vytěžování dat¹ (www.crisp-dm.org, obrázek 3.1).



Obrázek 3.1: Schéma procesu vytěžování dat dle metody CRISP-DM

V prvním kroku je nejprve potřeba obdržet zadání projektu, porozumět dané problematice a nejlépe si určit, co očekáváme ve výsledcích.

Blok druhý je zaměřen na samotné porozumění datům, které je velice důležité pro kvalitativní vyhodnocení. Jinými slovy se snažíme předem nalézt taková data, která by

¹značen CRISP-DM (CRoss Industry Standard Process for Data Mining)

mohla nést potřebnou informaci k dané problematice.

Příprava dat, nebo-li předzpracování. Vstupní data (měřený signál) se upraví do potřebného tvaru (například vypočteme střední hodnotu, teplotu, fraktální dimenze a další). Odstraní se artefakty a nepoužitelné části signálů. Předzpracování je považováno za nejnáročnější část procesu. Je důležité, aby v této části byly vyselektovány nevýznamné atributy. Odstraněním těchto dat získáme menší časovou náročnost v následujícím zpracování a zvýšíme tím výslednou kvalitu.

Modelování pracuje s předzpracovanými daty. Využívá metody umělé inteligence. Tyto metody rozdělujeme do dvou skupin. První vychází z učení z klasifikovaných příkladů, neboli učení s učitelem. Druhá skupina je založena na učení z neklasifikovaných příkladů, nazývá se učení bez učitele.

V předposledním bloku, vyhodnocení, se snažíme o určení úspěšnosti klasifikátorů. Definujeme tzv. matici záměn, která je popsána v kapitole 3.1, ze které vypočteme například celkovou správnost, celkovou chybu, senzitivitu, specificitu a výkonostní křivku ROC [3.4].

Dostali jsme se až k poslednímu bloku, který je pro zadavatele nejdůležitější. Jde o samostatné využití výsledků v praxi. [6]

3.2 Předzpracování, snížení dimenze

Dimenzi stavového prostoru můžeme snížit několika způsoby. Jedním z nich je použití vhodné transformace, dalším je například odstranění nevhodných příznaků tzv. metodou selekce příznaků. Pro experimenty se snížením dimenze jsem použil PCA.

Ostatní selekční metody, použité v této práci, jsem vybral zcela empiricky.

3.2.1 PCA (Principal Components Analysis)

Principal Component Analysis, neboli Analýza hlavních komponent (složek). Matematicky jde o lineární ortogonální transformaci. Cílem metody je tedy transformace k -dimenzionálního prostoru (k je počet příznaků) do prostoru s dimenzí mnohem nižší, ale s možností zpětné rekonstrukce \mathbf{z} , kde \mathbf{z} je přesně rekonstruovatelné z \mathbf{y} .

$$\mathbf{y} = \mathbf{W}_D(\mathbf{z} - \bar{\mathbf{z}}) \quad (3.1)$$

\mathbf{W}_D je transformací z N -dimenzionálního prostoru měřeného prostoru do D -dimenzionálního

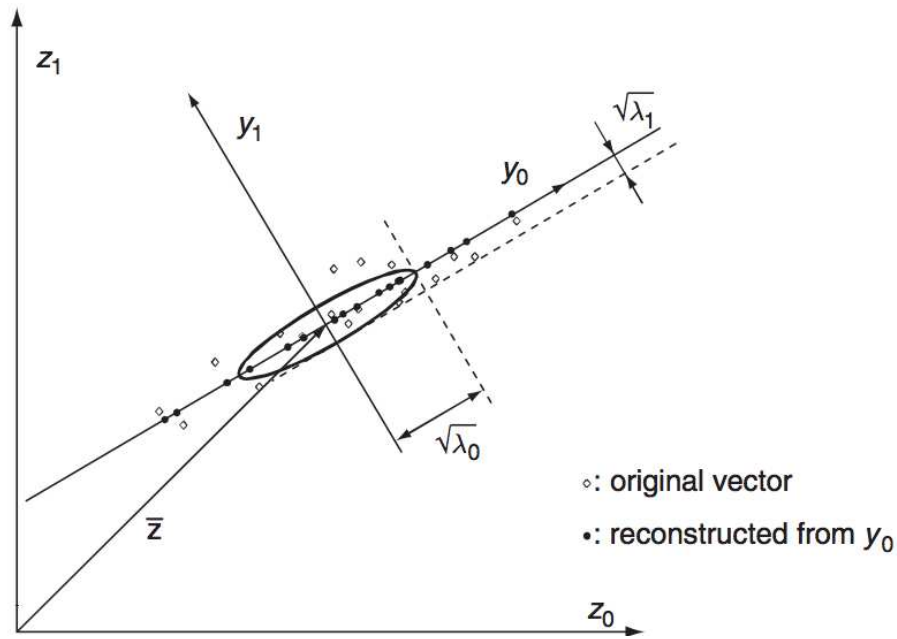
příznakového prostoru. Hledáme ortonormální lineární transformaci, umožňující rekonstruovat vektory s co nejmenší střední kvadratickou odchylkou (MMSE - minimum mean square error).

$$\mathbf{W}_D = \underset{w}{\operatorname{argmin}} \{E[|\bar{\mathbf{z}}_{\text{IMMSE}}(\mathbf{y}) - \mathbf{z}|^2]\}, \quad \mathbf{y} = \mathbf{W}\mathbf{z} \quad (3.2)$$

kde

$$\bar{\mathbf{z}}_{\text{IMMSE}}(\mathbf{y}) = \mathbf{W}_D^T \mathbf{y} = \mathbf{W}_D^T \mathbf{W}_D \mathbf{z} \quad (3.3)$$

Každá složka nese hodnoty rozdílné variance. Ve směru největší variance se položí první komponenta, kolmo k ní ve směru další největší variance je položena komponenta druhá. Takto se proces opakuje do vyčerpání všech variancí (obrázek 3.2)



Obrázek 3.2: Principal Component Analysis

Výběr se provede pomocí kovarianční matice ze vstupní matice (\mathbf{C}_y musí být diagonální matice)

$$\mathbf{C}_y = \mathbf{W}_D \mathbf{C}_z \mathbf{W}_D^T \quad (3.4)$$

a určí se její vlastní čísla. Vlastní čísla vypovídají o varianci jednotlivých komponent. Jsou seřazena sestupně, podle velikosti. V dalším kroku se spočítají vlastní vektory tvořící bázi. Snažíme se snížit dimenzi příznakového vektoru, z toho vypývá, že vybereme pouze několik prvních komponent podle hodnot vlastních čísel.

PCA může být použita jako první krok ke snížení dimenze vybraného prostoru. V praxi to znamená, že kovarianční matice je často nahrazena vzorem kovariance odhadnuté z trénovací množiny.

Bohužel, PCA může být contra-produktivní v klasifikační a odhadové problematice. Kritérium PCA zvolí podprostor z příznakového prostoru takový, že rozdíly variancí \mathbf{z} jsou co možná největší. Bohužel, bez ohledu na třídy. Podprostor s velkými variancemi není nutně jedním ze tříd, které jsou dobře odděleny.

Druhá nevýhoda PCA je, že výsledky nejsou invariantní ke konkrétní volbě fyzikálních jednotek. Každý element ze \mathbf{z} je jednotlivě změněn podle jeho jednotky, ve které je vyjádřen. [2]

3.3 Modelování

3.3.1 Základní popis klasifikátorů

U CTG využíváme pevně danou množinu příznaků vstupující do klasifikátoru. Příznak je sloupcový vektor $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$, jehož hodnoty reprezentují jednotlivá měření (jednotlivé porody). U klasifikace můžeme použít všechny poskytnuté příznaky nebo pro zvýšení přesnosti využijeme pouze vybrané pomocí některé z dostupných metod selekce, jak již bylo popsáno v předcházející kapitole. Každý příznak si můžeme představit jako vstupní signál přicházející na vstup klasifikátoru. Na výstupu očekáváme pouze konkluzi o zařazení příznaku do jisté třídy. Obrázek 3.3

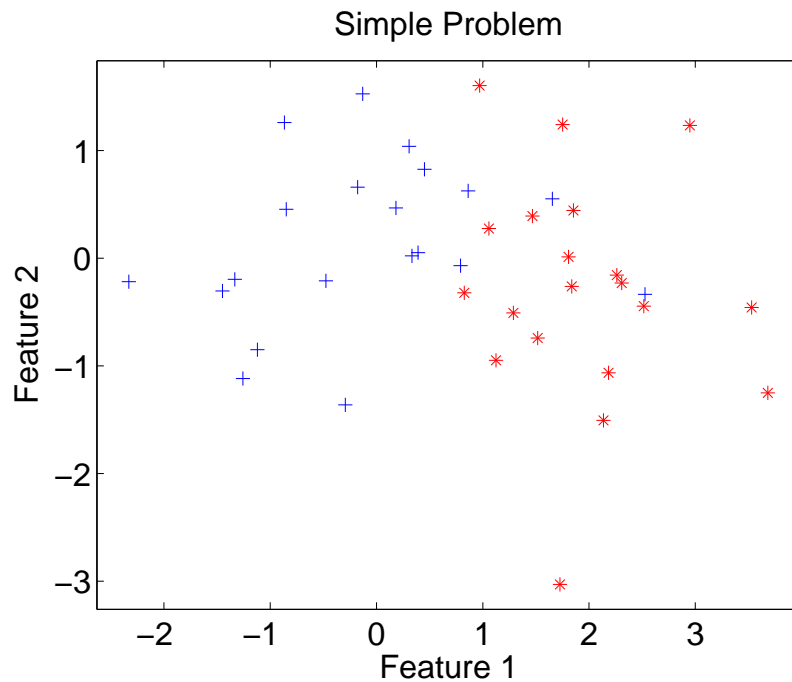


Obrázek 3.3: Zjednodušený model klasifikátoru

V našem případě dělíme data pouze do dvou tříd, dle anotace. Třídění do R^2 se nazývá dichotomie. Klasifikace se provádí na celém prostoru R^n , kde n představuje počet

vstupních příznaků. Pro jednoduchost si na obrázku 3.4 ukážeme příklad 2-dimenzionálního prostoru.

```
A=gendats([20 20]);      % vytvoří dataset se dvěma třídami, každá o 20 prvcích
scatterd(A)              % vykreslí dataset do 2D
```



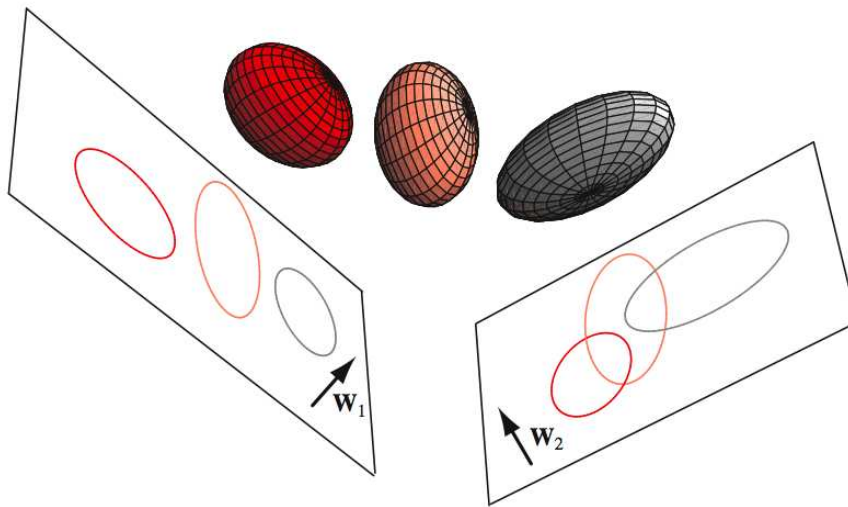
Obrázek 3.4: Příklad 2D prostoru s třídami "+" a "*"

U více rozměrného prostoru se potýkáme s problémem, kdy nemůžeme graficky zobrazit všechny příznaky do jednoho obrazu. Analogicky k tomu dochází již u zobrazení 3D těles pravoúhlým promítáním, kde nevidíme obraz axonometricky, ale jen z jedné strany. Proto zde graficky nezobrazím výsledek celkové klasifikace.

Klasifikaci dělíme podle metody učení z klasifikovaných a neklasifikovaných příkladů. Setkáváme se se dvěma základními metodami. První vychází z učení z klasifikovaných příkladů a nazývá se učení s učitelem. Druhá je založena na učení z neklasifikovaných příkladů a říkáme jí učení bez učitele.[5]

Všechny tři mnou použité klasifikátory spadají do skupiny metod učení s učitelem. Musím tedy provést rozdělení do dvou množin, trénovací a testovací (někdy se dělí na množiny tři, kde třetí množina je validační). Nejprve necháme klasifikátor učit z množiny trénovací. V dalším kroku provedeme otestování natrénovaného klasifikátoru na množině

testovací. K rozdělení dat na trénovací a testovací množinu použijí křížovou validaci s 10 složkami tak, že 9 složek vždy přiřadím do trénovací množiny a zbylou jednu složku do množiny testovací[3.4.0.3]. Dělení provádíme z toho důvodu, že pokud provedeme trénování a testování na stejných množinách, získáme příliš optimistické hodnoty. Máme-li dostatečně velkou trénovací množinu, je pravděpodobné, že dosáhneme relativně vysoké přesnosti klasifikace. Samozřejmě také záleží na použitých datech. Musíme dát pozor, aby nedocházelo k přeučení klasifikátoru, což způsobuje nežádoucí snížení přesnosti.



Obrázek 3.5: Zobrazení prostoru 3D do 2D

3.3.2 k -NN (k -Nearest Neighbour)

Nejjednodušší klasifikátor z naší trojice. Jeho princip je založen na predikci neznámé hodnoty z příkladů, kde tuto hodnotu známe.

- Nejprve nalezneme k nejbližších sousedů, které uzavřeme do jedné množiny.
- Zjistíme, do jakých tříd spadají sousedé v námi definované množině.
- Nakonec se náš zkoumaný prvek zařadí do takové třídy, která má vyšší zastoupení v nalezené množině sousedů.

U klasifikátorů s dichotomickým dělením se doporučuje volba lichého počtu k , aby ne-nastávaly nejisté stavy řazení do příslušné třídy a výsledné hlasování bylo vždy nerovnoměrné.

Pro kvantitativní ohodnocení podobnosti dvou příkladů používáme různé vzdálenostní metriky.

- Euklidovská vzdálenost - hyperkoule
- Mahalanobisova vzdálenost - hyperelipsoidy

[3] Hodnota závislé veličiny je poté určena z *jádrové funkce*. Nejjednodušší jádrovou fcí je aritmetický průměr.

$$\bar{y}(i) = \frac{1}{k} \sum_{j=1}^k y(j) \quad (3.5)$$

Predikce pro i -tý prvek je průměrnou hodnotou závislé veličiny nalezené u k nejbližších sousedů v trénovacích datech.

3.3.3 Rozhodovací stromy

Algoritmy rozhodovacích stromů použitých v nástroji PRTools jsou variabilní podle použitého kritéria. Nebudu popisovat všechna kritéria, ale omezím se na základní princip TDIDT² popsany v [6].

Mezi nejoblíbenější rozhodovací stromy patří TDIDT a ID3. Metoda TDIDT má dvě fáze

- vytváření rozhodovacího stromu
- a tzv. prořezávání

3.3.3.1 Vytváření rozhodovacího stromu

- V prvním kroku nalezneme atribut, který má nejmenší entropii (největší množství informace).
- Zvolíme jej kořenem stromu.
- Rozdělíme množinu příkladů na podmnožiny dle hodnot kořenového atributu.
- Každá podmnožina nese příklady s právě jednou hodnotou tohoto atributu.

²Top-Down Induction of Decision Trees - indukce rozhodovacích stromů shora dolů

- Prohledáme každou z podmnožin a nalezneme v každé z nich další nejvýznamnější atribut.
- Pokračujeme rekurzivně, dokud máme dostatek atributů a příkladů.

Množství informace se měří entropií. Dle Shannonovy věty platí, že:

$$H_y = -p_1 \log_2 p_1 - p_2 \log_2 p_2 \quad (3.6)$$

kde p_1 — p_2 je poměr pozitivních — negativních případů v j -té podmnožině k celkovému množství prvků podmnožiny. H_j je kladné, neboť p_1, p_2 náležejí intervalu $< 0; 1 >$, tzn. $\log_2 p_x < 0$. Celkovou entropii spočteme jako vážený součet entropií daných podmnožin

$$\mathbf{H}_y = \sum_{j=1}^K P_j H_j \quad (3.7)$$

kde K je počet vytvořených podmnožin daným atributem, P_j poměr velikosti j -té podmnožiny k množině všech příkladů. [5]

3.3.4 Support Vector Machine (SVM)

SVM vybere jedno partikulární řešení. Rozděluje třídy s maximálním rozpětím (margin). Margin je definována jako největší možná šíře mezi vzory, která tvoří rozhodovací hranice (obrázek 3.6). To může mít za následek, že partikulární řešení má největší generalizační schopnost.

Mějme trénovací vzory $\mathbf{z}_n, n = 1, \dots, N_S$. Každý z nich označme $c_n \in \{1, -1\}$, to rozhodne, do které ze dvou tříd byl vzor přiřazen. Poté lineární klasifikátor $g(\mathbf{z}) = \mathbf{w}^T \mathbf{z} + b$ je nalezen tak, že:

$$\begin{aligned} \mathbf{w}^T \mathbf{z}_n + b &\geq 1 && \text{pokud } c_n = +1 \\ \mathbf{w}^T \mathbf{z}_n + b &\leq -1 && \text{pokud } c_n = -1 \end{aligned} \quad (3.8)$$

pro každé n

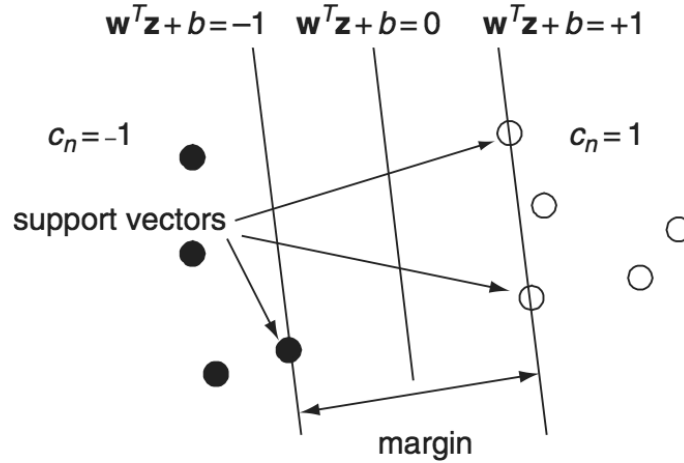
Tato dvě omezení mohou být přepsána do jedné nerovnosti:

$$c_n \mathbf{w}^T \mathbf{z}_n + b \geq 1 \quad (3.9)$$

Rozdílovým vektorem $g(\mathbf{z})$ je \mathbf{w} . Proto je kvadrát rozpětí inverzně úměrný $\|\mathbf{w}\|^2 = \mathbf{w}^T \mathbf{w}$. Pro maximalizaci rozpětí musíme snížit $\|\mathbf{w}\|^2$. Použitím *Lagrangeových multiplikátorů*

můžeme začlenit 3.9 do minimalizace:

$$L = \frac{1}{2} \|\mathbf{w}\|^2 + \sum_{n=1}^{N_S} \alpha_n (c_n [\mathbf{w}^T \mathbf{z}_n + b] - 1), \quad \alpha_n \geq 0 \quad (3.10)$$



Obrázek 3.6: Lineární SVM

L by měl být minimalizován podle \mathbf{w} a b a maximalizován podle *Lagrangeových multiplikátorů* α_n . Položíme parciální derivaci L podle \mathbf{w} a b rovnu nule:

$$\begin{aligned} \mathbf{w} &= \sum_{n=1}^{N_S} \alpha_n c_n \mathbf{z}_n \\ \sum_{n=1}^{N_S} c_n \alpha_n &= 0 \end{aligned} \quad (3.11)$$

Resubstitucí do (3.10) získáme tzv. *duální formu*:

$$L = \sum_{n=1}^{N_S} \alpha_n - \frac{1}{2} \sum_{n=1}^{N_S} \sum_{m=1}^{N_S} c_n c_m \alpha_n \alpha_m \mathbf{z}_n^T \mathbf{z}_m, \quad \alpha_n \geq 0 \quad (3.12)$$

L by mělo být maximalizováno podle α_n . Jde o kvadratický optimalizační problém, pro který jsou standardní SW přizpůsobeny. Po optimalizaci jsou α_n z (3.11) použité pro nalezení \mathbf{w} . Typicky se problém obchází tak, že většinu α_n položíme rovno nule.

Vzory, pro které platí, že $\alpha_n = 0$, se nepoužijí při výpočtu \mathbf{w} . Zbylé vzory \mathbf{z}_n , pro které $\alpha_n > 0$ nazýváme *support vectors*.

(citace [2], strany 168-170)

3.4 Vyhodnocení

Po klasifikaci se zabýváme vyhodnocením a ověřením přesností klasifikátorů na testovacích datech. Máme data dělená do dvou tříd. Třída pozitivní (naše třída 1) a druhé negativní (0). Vytvoříme si tzv. matici záměn [3.1]. Sloupce nesou informaci o předpovídané pozitivitě či negativitě v kombinaci s řádky nesoucími skutečný stav. V posledním sloupci tabulky jsou relativní četnosti těchto kombinací.

- TP (True Positive) - správně zařazených příkladů do pozitivní třídy
- FN (False Negative) - nesprávně zařazených příkladů do negativní třídy
- TN (True Negative) - správně zařazených příkladů do negativní třídy
- FP (False Positive) - nesprávně zařazených příkladů do pozitivní třídy

Tabulka 3.1: Matice záměn

	Předpovídaný pozitivní	Předpovídaný negativní	Relativní četnost
Skutečně pozitivní	a	b	$TP = \frac{a}{a+b}, FN = \frac{b}{a+b}$
Skutečně negativní	c	d	$TN = \frac{d}{c+d}, FP = \frac{c}{c+d}$

3.4.0.1 Senzitivita, specificita a výkonnost klasifikátorů

V biomedicíně nás nejvíce zajímají hodnoty zvané specificita a senzitivita. Specificita odpovídá relativní četnosti TN a senzitivita naopak relativní četnosti TP. Jinými slovy je specificita hodnota představující poměr správně ohodnocených případů s hypoxií. Senzitivita je poměr správně označených plodů s normálním stavem, tedy bez hypoxie. Vždy z celkových dat.

Výkonnost klasifikátoru je charakterizována hodnotami celková správnost[3.14] a celková chyba[3.13]

$$Celkova\ chyba = \frac{FP + FN}{TP + TN + FP + FN} \quad (3.13)$$

$$Celkova\ spravnost = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.14)$$

3.4.0.2 Křivka ROC

ROC (Receiver Operating Characteristic) pochází z oboru radiologie. Grafické zobrazení průběhu křivky nám napomáhá pro nalezení neoptimálnějšího klasifikátoru s největšími hodnotami senzitivity a specificity. Na osu x vynášíme hodnoty specificity a na osu y hodnoty senzitivity. Snažíme se maximalizovat plochu pod křivkou. Ideální stav je, když dosáhneme jednotkové plochy. Máme-li k dispozici více klasifikátorů, použijeme ten s největší plochou.

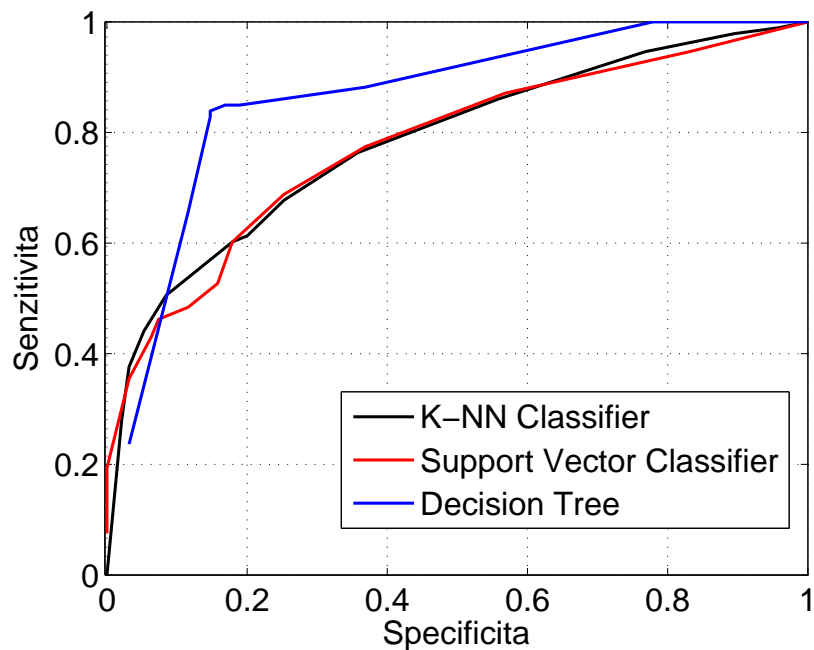
```
e=roc(A,W) % vypočet ROC křivky
e.error=ones(size(e.error))-e.error %uprava os
plot(e) % vykreslení
```

```
E=A*W*testauc %vypočte plochy pod křivkou
E=ones(size(E))-cell2mat(E) % upravi osy
```

Vypočtené plochy jsou:

- K -Nearest Neighbour = 0.7831
- Support Vector Machine = 0.7845
- Decision Tree = 0.8795

Podle obrázku 3.7 a vypočtených hodnot vybereme klasifikátor Decision Tree.



Obrázek 3.7: Křivka ROC

3.4.0.3 Křížová validace

V originálním znění Cross Validation.

Princip:

- Data jsou nahodně rozdělena do určitého počtu složek, podmnožin (dle publikací je plně dostačující dělení na 10 podmnožin).
- Každá z těchto složek je právě jednou použita jako testovací množina. Zbylý počet se použije pro trénování klasifikátoru.
- Proces se stále opakuje, dokud nejsou vyčerpány všechny kombinace.

Zvýšením počtu opakování a průměrováním chybových odhadů, zvýšíme přesnost. Pro krosvalidaci využíváme funkci *crossval* z toolboxu PRTools (popsáno v kapitole 4).

(literatura 5)

Kapitola 4

PRTools

Jak jsem již zmínil, jde o mocný nástroj k programu MATLAB, vyvíjený na Delftské univerzitě v Holandsku. Toolbox je pro akademické účely poskytnut zcela zdarma na jeho domovských stránkách <http://prtools.org/> [4]. V sekci Documentation naleznete manuál k poslední verzi. Velice užitečná literatura, na kterou se odkazuje ze stránek toolboxu nese název Classification, Parameter Estimation and State Estimation [2] a provází čtenáře problematikou přímo s ukázkami kódu MATLABU.

Toolbox nabízí obrovské množství funkcí, využitelné snad ve všech odvětvích rozpoznávání.

4.1 Struktura funkcí toolboxu

- Datasetsy
- Datové soubory
- Mapování
- Generování dat
- Klasifikátory
- Selektce příznaků
- Testovací funkce
- Regrese

- Zpracování obrazů
- Selektce příznaků z obrazů
- Shlukování (clustering)
- Vykreslovací funkce
- Variabilní testy a podpora rutinních operací
- Příklady

Některé funkce, které jsem použil, jsou uvedené v příloze B.

Kapitola 5

Praktická realizace a experimenty

5.1 Zpracování

5.1.1 Použité výpočetní zařízení

Pro všechny výpočty jsem použil počítač s parametry v tabulce 5.1. Specifikaci uvádím pro porovnání časové náročnosti výpočtů.

Hardware a software	Popis
Model	Apple MacBook 13,3" unibody
Processor	Intel Core 2 duo, frekvence 2GHz, cache 3MB
Paměť	velikost 2GB, rychlost 1066MHz, DDR3
OS	Mac OS X Leopard, verze 10.5.8
Matlab	verze 7.7.0.471 (R2008b)
PRTools	verze 4.1.4

Tabulka 5.1: Technická specifikace použitého výpočetního zařízení

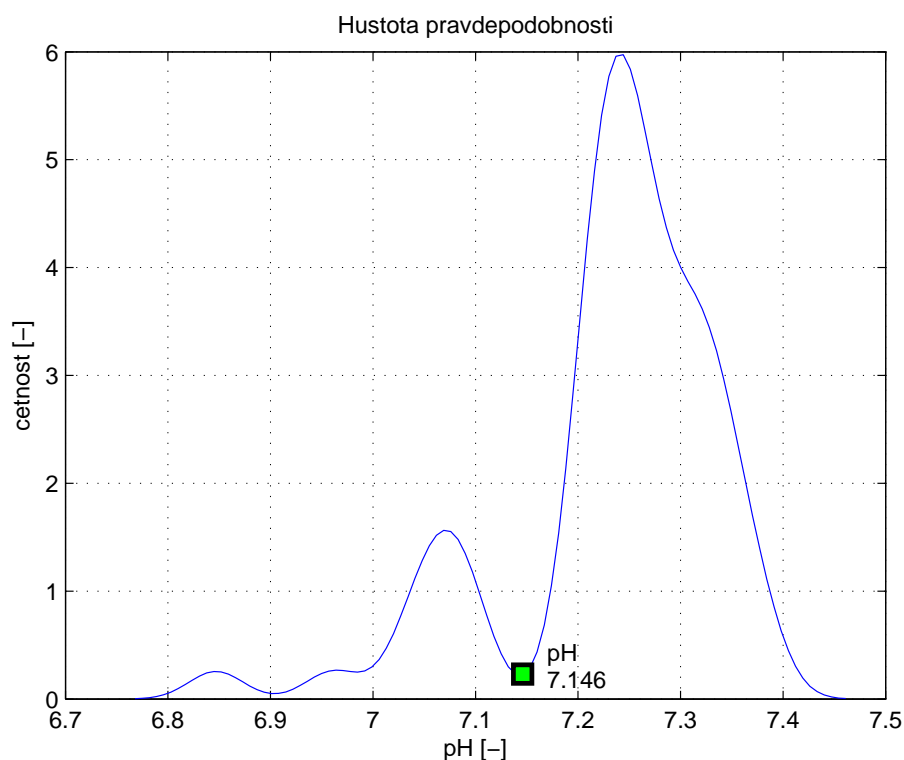
5.1.2 Anotace

Pro klasifikaci je potřeba rozdělit data do několika tříd. V našem případě využíváme dichotomické rozdělení. Anotace představuje pH krve plodu. Pokud je pH menší než 7.15, jedná se o hypoxii, tedy třídu 0. V jiném případě je stav považován za normální a řadíme jej do třídy 1. Hranici pH volí lékař, většinou v rozmezí 7.15 nebo 7.20.

Do rukou se mi dostala databáze příznaků z Řecka (konkrétně z 59 sledovaných porodů)

bez rozdělení do tříd. Dovolil jsem si tedy provést experiment, který by mohl objasnit lékařem stanovenou hranici.

Na obrázku 5.1 je zobrazen histogram souboru. Hledám lokální minimum mezi hodnotami pH 7.10 a 7.25 (± 0.05 od předpokládaných hodnot). Hodnota ukazuje na 7.146, což odpovídá tvrzení lékařů, že hranice pH 7.15 je optimální.



Obrázek 5.1: Experiment s histogramem pH

5.1.3 Selekcce, snížení dimenze

Pro výběr příznaků jsem použil funkce *featselp* a *featselb*. Vždy nejprve obecné nastavení a poté s parametrem příslušného klasifikátoru. U kombinací a rozhodovacho stromu jsem používal pouze selekční funkce bez parametru.

```
[vyber,matic] = featselp(A, []); % obecné nastavení
[vyber,matic] = featselp(A, knnc); % s použitím pro klasifikátor knnc
[vyber,matic] = featselp(A, svc); % s použitím pro klasifikátor svc

A = featsel(A, vyber); % vyber příznaku z datasetu
```

Druhá metoda je PCA, ze které bohužel nezískáme klasifikované příznaky. Pouze transformuje dataset A do zvolené dimenze.

```
w=pca(A);           % obecné nastavení
w=pca(A,pocetDimenzi); % parametr pocetDimenzi určuje výstupní dimenzi

A=A*w;           % aplikace w na vstupní dataset
```

Příznaky	featselb	featselp
'ApEn2015'	'ApEn2015'	'ApEn2015'
'ApEn202'	'fdVarianceD'	'fdVarianceD'
'SampEn2015'	'fdVarianceDl'	'fdVarianceDl'
'SampEn202'	'fdHiguchiDs'	'lzc2'
'lzc2'	'fdHiguchiDl'	'SampEn202'
'fdVarianceD'		'ApEn202'
'fdVarianceDs'		'fdBoxCountD'
'fdVarianceDl'		
'fdBoxCountD'		
'fdBoxCountDs'		
'fdBoxCountDl'		
'fdHiguchiD'		
'fdHiguchiDs'		
'fdHiguchiDl'		
'fdDfaD'		
'fdDfaDs'		
'fdDfaDl'		

Tabulka 5.2: Společná tabulka příznaků a selekcí pro všechny metody.

featselb(A,knnc)	featselp(A,knnc)
'ApEn2015'	'fdHiguchiDs'
'fdVarianceDs'	'fdHiguchiDl'
	'fdVarianceDs'
	'fdDfaD'
	'fdBoxCountDl'
	'fdBoxCountDs'
	'fdHiguchiD'
	'fdVarianceDl'
	'SampEn202'

Tabulka 5.3: Dodatková tabulka příznaků pro selekční metody s parametrem knnc.

featselb(A,svc)	featselp(A,svc)
'SampEn2015'	'fdHiguchiDs'
'SampEn202'	'fdBoxCountDs'
'fdVarianceDs'	'SampEn2015'
'fdVarianceDl'	'fdDfaDl'
'fdBoxCountDs'	'fdVarianceD'
'fdBoxCountDl'	'ApEn202'
'fdHiguchiDl'	'fdHiguchiD'

Tabulka 5.4: Dodatková tabulka příznaků pro selekční metody s parametrem svc.

Pro každé nastavení klasifikátoru a výběrové funkce je provedena křížová validace s dělením na 10 složek a s 10-ti násobným opakováním. Pouze u stromového klasifikátoru je použito právě jedno opakování. Výsledné hodnoty jsem vynesl v tabulkách v následující kapitole. (kompletní tabulky jsou v příloze A). Vybral jsem nejlepší řešení, které je uvedeno v závěru.

Kapitola 6

Vyhodnocení

V této kapitole jsem vypsal tabulky pouze nejúspěšnějších nastavení daného klasifikátoru (kompletní tabulky jsou uvedeny v příloze A). V popisku tabulek naleznete potřebné informace o použité selekční metodě a parametrech. Nastavení klasifikátorů je vždy v prvním sloupci této tabulky. Pokud nad tabulkou uvedu přibližnou časovou náročnost výpočtu, jedná se o délku běhu kódu celé tabulky, nikoliv pro jeden klasifikátor. Při použití rozhodovacích stromů uvedu navíc přesnější hodnotu časové náročnosti (velké rozdíly u použitého nastavení). Používám zde křížovou validaci, která dělí data na 10 složek, při 10 opakováních. Vyjimku tvoří rozhodovací stromy, kde je nastaveno pouze jedno opakování z důvodu vysoké časové náročnosti úlohy (přesné informace naleznete v souborech *.mat v podsložkách Table a Tables).

6.1 Výpočtené hodnoty jednotlivých klasifikátorů

6.1.1 Klasifikátor *knn*

Přibližná časová náročnost výpočtu: 3 minuty/tabulku

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
<code>knn([],3)</code>	72.2	69.4	76.1	70.8	29.2

Tabulka 6.1: *k*-NN bez použití selekce

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
<code>knnc([],1)</code>	76.1	67.5	77.1	71.8	28.2

Tabulka 6.2: k -NN s použitím funkce *featselp*, parametr *knnc*

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
<code>knnc([],1)</code>	83.1	72.7	79.6	77.9	22.1
<code>knnc([])</code>	78.6	71.6	76.0	75.1	24.9

Tabulka 6.3: k -NN s použitím funkce *featselp*, bez parametru

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
<code>knnc([],1)</code>	77.8	71.3	78.9	74.5	25.5
<code>knnc([])</code>	71.1	72.4	75.4	71.7	28.3

Tabulka 6.4: k -NN s použitím funkce *featselb*, bez parametru

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
<code>knnc([],3)</code>	74.7	73.8	76.2	74.3	25.7

Tabulka 6.5: k -NN s použitím funkce *featselb*, parametr *knnc*

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
<code>knnc([],11)</code>	65.8	74.1	73.6	70.0	30.0

Tabulka 6.6: k -NN, *pca* bez parametru

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc([],1)	72.8	69.1	77.4	71.0	29.0

Tabulka 6.7: k -NN snížení dimenze funkcí *pca* na 2-D

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc([])	62.5	80.8	73.5	71.6	28.4

Tabulka 6.8: k -NN snížení dimenze funkcí *pca* na 5-D

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc([])	62.5	80.8	73.5	71.6	28.4

Tabulka 6.9: k -NN snížení dimenze funkcí *pca* na 10-D

6.1.2 Klasifikátor *svc*

Přibližná časová náročnost výpočtu: 15 minut/tabulku

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 's')	65.8	77.1	75.7	71.5	28.5

Tabulka 6.10: SVM bez použití selekce

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 'e')	64.7	79.0	78.6	71.8	28.2

Tabulka 6.11: SVM s použitím funkce *featselb*, bez parametru

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([])	65.0	77.5	74.4	71.3	28.7
svc([], 'p')	65.0	77.5	74.4	71.3	28.7
svc([], 'h')	65.0	77.5	74.4	71.3	28.7

Tabulka 6.12: SVM s použitím funkce *featselb*, parametr *svc*

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 'e')	66.9	76.7	76.8	71.8	28.2

Tabulka 6.13: SVM s použitím funkce *featselp*, bez parametru

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 'e')	63.9	77.9	76.8	70.9	29.1

Tabulka 6.14: SVM s použitím funkce *featselp*, parametr *svc*

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 's')	65.8	76.7	77.0	71.3	28.7

Tabulka 6.15: SVM, snížení dimenze funkcí *pca*, bez parametru

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 's')	66.7	74.5	75.2	70.6	29.4

Tabulka 6.16: SVM, snížení dimenze funkcí *pca* na 2D

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 's')	65.3	76.7	76.3	71.0	29.0

Tabulka 6.17: SVM, snížení dimenze funkcí *pca* na 5D

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 's')	66.7	77.1	76.3	71.9	28.1

Tabulka 6.18: SVM, snížení dimenze funkcí *pca* na 10D

6.1.3 Klasifikátor *treec*

Časová náročnost výpočtu tabulky 6.19: 19 minut/tabulku

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
treec	67.5	66.3	73.3	66.9	33.1

Tabulka 6.19: Rozhodovací strom bez použití selekce

Časová náročnost výpočtu tabulky 6.20: 38 minut/tabulku

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
treec([], 'fishcrit', NaN)	59.7	72.0	69.9	65.9	34.1

Tabulka 6.20: Rozhodovací strom s použitím funkce *featselb*, bez parametru

Časová náročnost výpočtu tabulky 6.21: 19 minut/tabulku

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
treec	64.2	60.2	69.0	62.2	37.8

Tabulka 6.21: Rozhodovací strom s použitím funkce *featselp*, bez parametru

Časová náročnost výpočtu tabulky 6.22: 23 minut/tabulku

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
treec([], 'infcrit', NaN)	60.6	65.6	60.8	63.1	36.9

Tabulka 6.22: Rozhodovací strom s použitím funkce *featselb*, parametr *treec*

Časová náročnost výpočtu tabulky 6.23: 23 minut/tabulku

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
treec([], 'fishcrit', NaN)	62.8	70.2	70.6	66.5	33.5

Tabulka 6.23: Rozhodovací strom, snížení dimenze funkcí *pca*, bez parametru

Časová náročnost výpočtu tabulky 6.24: 23 minut/tabulku

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
treec([], 'infcrit', NaN)	61.9	72.2	66.3	67.1	32.9

Tabulka 6.24: Rozhodovací strom, snížení dimenze funkcí *pca* na 2D

Časová náročnost výpočtu tabulky 6.25: 22 minut/tabulku

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
treec([],NaN,NaN)	68.1	66.8	76.7	67.4	32.6

Tabulka 6.25: Rozhodovací strom, snížení dimenze funkcí *pca* na 5D

Časová náročnost výpočtu tabulky 6.26: 22 minut/tabulku

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
treec([],NaN,NaN)	68.1	66.8	76.7	67.4	32.6

Tabulka 6.26: Rozhodovací strom, snížení dimenze funkcí *pca* na 10D

6.2 Vypočtené hodnoty pro kombinace klasifikátorů

6.2.1 Základní kombinace

Přibližná časová náročnost výpočtu: 9 minut/tabulku

- K = knnc
- S = svc
- T = treec

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc - treec /meanc/	67.5	72.0	78.3	69.7	30.3
svc - treec /medianc/	67.5	72.0	78.3	69.7	30.3
knnc - treec /meanc/	66.4	73.1	76.1	69.7	30.3
knnc - treec /medianc/	66.4	73.1	76.1	69.7	30.3

Tabulka 6.27: Kombinace bez použití selekce

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc	73.1	72.6	79.1	72.8	27.2
knnc - svc /meanc/	64.2	77.8	80.3	71.0	29.0
knnc - svc /medianc/	64.2	77.8	80.3	71.0	29.0

Tabulka 6.28: Kombinace s použitím funkce *featselb*, bez parametru

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc	85.0	73.4	78.7	79.2	20.8
knnc - svc - treec /medianc/	68.1	76.3	78.4	72.2	27.8
knnc - svc /meanc/	62.2	80.8	82.0	71.5	28.5

Tabulka 6.29: Kombinace s použitím funkce *featselp*, bez parametru

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc	61.1	79.3	75.0	70.2	29.8
knnc - svc - treec /medianc/	54.4	83.7	73.2	69.1	30.9
knnc - svc - treec /meanc/	60.0	72.3	71.3	66.2	33.8

Tabulka 6.30: Kombinace, snížení dimenze funkcí *pca*, bez parametru

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc - svc /meanc/	63.3	80.5	75.7	71.9	28.1
knnc - svc /medianc/	63.3	80.5	75.7	71.9	28.1

Tabulka 6.31: Kombinace, snížení dimenze funkcí *pca* na 2D

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc	62.8	77.5	73.9	70.1	29.9
knnc - svc - treec /meanc/	56.9	78.6	74.6	67.8	32.2
knnc - svc - treec /medianc/	53.9	83.7	74.7	68.8	31.2

Tabulka 6.32: Kombinace, snížení dimenze funkcí *pca* na 5D

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc	62.2	78.6	75.1	70.4	29.6
knnc - svc - treec /meanc/	60.3	81.9	74.1	71.1	28.9
knnc - svc - treec /medianc/	59.2	86.3	75.1	72.8	27.2

Tabulka 6.33: Kombinace, snížení dimenze funkcí *pca* na 10D

6.2.2 Kombinace po výběru nejlepšího klasifikátoru

Časová náročnost výpočtu tabulky 6.34: 8 minut/tabulku

K = knnc([],3); S = svc([], 's'); T = treec

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
K - S - T /meanc/	66.7	73.0	77.1	69.8	30.2
S	65.8	78.3	74.5	72.0	28.0
K - S - T /medianc/	64.7	79.0	76.7	71.8	28.2

Tabulka 6.34: Kombinace nejlepších klas. bez použití selekce

Časová náročnost výpočtu tabulky 6.35: 33 minut/tabulku

K = knnc([],1); S = svc([], 'e'); T = treec([], 'fishcrit', str2num('NaN'))

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
K	74.7	79.3	80.6	77.0	23.0
K - S /meanc/	69.4	80.5	81.3	75.0	25.0
K - S /medianc/	69.4	80.5	81.3	75.0	25.0

Tabulka 6.35: Kombinace nejlepších klas. s použitím funkce *featselb*, bez parametru

Časová náročnost výpočtu tabulky 6.36: 9 minut/tabulku

K = knnc([],1); S = svc([], 'e'); T = treec

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
K	80.8	74.6	78.2	77.7	22.3
K - S /meanc/	68.1	77.1	80.8	72.6	27.4
K - S /medianc/	68.1	77.1	80.8	72.6	27.4
K - S - T /medianc/	70.3	73.1	76.8	71.7	28.3
K - S - T /meanc/	68.1	73.1	79.8	70.6	29.4

Tabulka 6.36: Kombinace nejlepších klas. s použitím funkce *featselp*, bez parametru

Časová náročnost výpočtu tabulky 6.37: 30 minut/tabulku

K = knnc([],11); S = svc([], 's'); T = treec([], 'fishcrit', str2num('NaN'))

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
K - S - T /meanc/	68.9	73.5	77.6	71.2	28.8
K - S - T /medianc/	64.7	75.7	77.3	70.2	29.8

Tabulka 6.37: Kombinace nejlepších klas., snížení dimenze funkcí *pca*, bez parametru

Časová náročnost výpočtu tabulky 6.38: 37 minut/tabulku

K = knnc([],1); S = svc([], 's'); T = treec([], 'infcrit', str2num('NaN'))

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
K - S /meanc/	70.6	77.5	78.9	74.0	26.0
K - S /medianc/	70.6	77.5	78.9	74.0	26.0

Tabulka 6.38: Kombinace nejlepších klas., snížení dimenze funkcí *pca* na 2D

Časová náročnost výpočtu tabulky 6.39: 106 minut/tabulku

K = knnc; S = svc([], 's'); T = treec([], str2num('NaN'), str2num('NaN'))

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
S	66.9	77.9	76.2	72.4	27.6
K - S - T /medianc/	63.6	78.7	75.7	71.1	28.9
K - S - T /meanc/	66.9	71.3	74.3	69.1	30.9

Tabulka 6.39: Kombinace nejlepších klas., snížení dimenze funkcí *pca* na 5D

Časová náročnost výpočtu tabulky 6.40: 98 minut/tabulku

K = knnc; S = svc([], 's'); T = treec([], str2num('NaN'), str2num('NaN'))

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
S	66.9	77.9	76.2	72.4	27.6
K - S - T /medianc/	63.6	78.7	75.7	71.1	28.9
K - S - T /meanc/	66.9	71.3	74.3	69.1	30.9

Tabulka 6.40: Kombinace nejlepších klas., snížení dimenze funkcí *pca* na 10D

Kapitola 7

Závěr

Vzhledem k časové náročnosti a výsledné přesnosti jsou za daných okolností klasifikátory support vector machine a rozhodovací stromy méně účinné než k -NN.

Konfigurace klasifikátoru k -NN, kde k je vybrané zcela automaticky a s použitím selekční metody *featselp* (bez parametru), vykazuje hodnoty senzitivity 85% a specifcity 73.4%, s celkovou správností 79.2% (výsledek klasifikace při experimentu s kombinací, tabulka 6.29).

Druhý klasifikátor (nejčastěji dosahuje nejlepších výsledků) vykazuje hodnotu senzitivity 83.1%, specifcity 72.9%, s celkovou správností 77.9% (tabulka 6.3). Klasifikátor je testován na stejné množině příznaků jako klasifikátor předchozí (totožná selekční funkce *featselp*, bez parametru). Využívá pevné nastavení k . Hledá pouze jednoho "souseda" ($k=1$).

Kombinace klasifikátorů nepřinesla žádné efektivní zlepšení výsledků.

Do budoucna je vhodné více prozkoumat možnosti konfigurace i dalších klasifikátorů a selekčních metod, které nabízí nástroj PRTools. Je potřeba hledat další příznaky, které by více rozlišily jednotlivé třídy.

Přestože jsem u klasifikátoru k -NN dosáhl nejvyšších hodnot, jsou tyto hodnoty do optimálních sta procent ještě příliš vzdálené.

Literatura

- [1] Z. KOTEK & V. CHALUPA, I. BRŮHA, J. JELÍNEK: *Adaptivní a učící se systémy*, Praha: Vydalo SNTL spolu s n. p. Alfa.(1980)
- [2] F. VAN DER HEIDEN, R.P.W. DUIN, D. DE RIDDER & D.M.J. TAX: *Classification, Parameter Estimation and State Estimation*, Canada:John Wiley & Sons Inc. (2004)
- [3] RICHARD O. DUDA, PETER E. HART, DAVID G. STORK: *Pattern Classification, 2nd edition*, Canada:John Wiley & Sons, Inc. (2001)
- [4] R.P.W. DUIN, P. JUSZCZAK, P. PACLIK, E. PEKALSKA, D. DE RIDDER, D.M.J. TAX, S. VERZAKOV: *PRTools4: A Matlab Toolbox for Pattern Recognition, version 4.1*, Delft Pattern Recognition Research, Fakulty EWI - ICT, Delft University of Technology. (August 2007)
<http://prtools.org/>
- [5] VLADIMÍR MAŘÍK, OLGA ŠTĚPÁNKOVÁ, JIŘÍ LAŽANSKÝ & KOLEKTIV: *Umělá inteligence 1*, Praha: Academia. (1993)
- [6] VLADIMÍR MAŘÍK, OLGA ŠTĚPÁNKOVÁ, JIŘÍ LAŽANSKÝ & KOLEKTIV: *Umělá inteligence 4*, Praha: Academia. (2003)
- [7] NEOVENTA MEDICAL AB: Společnost vyvíjející biomedicínské monitorovací a řídicí nástroje pro zlepšení zdravotní péče v perinatálním období. Švédsko
<http://www.neoventa.com/>
- [8] PARADIGM: Biomedicínské systémy <http://www.paradigmmedicalsyste.ms.com/>
- [9] ANNA KARIN SUNDSTÖRM, DAVID ROSÉN, K. G. ROSÉN: *Fetal Surveillance*, Göteborg(2003)

- [10] ČECH, E.; HÁJEK, Z.; SRP, B.: *Porodnictví*, 2.nd ed. Praha : Olympia, (2006). ISBN 80-247-1303-9.
- [11] PETRA PAPÍRNÍKOVÁ *Moderní babictví 10, (2006), Gynekologicko-porodnická klinika VFN a 1. LF UK, Praha (2003)*
<http://www.levret.cz/publikace/casopisy/mb/2006-10/clanek.php?c=1>
- [12] MUDR. PHDR. PAVEL ČEPICKÝ, CSC : Webové stránky určené pro nelékařskou veřejnost <http://www.levret.cz/>

Příloha A

Kompletní tabulky výpočtů

A.0.3 KNN

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc([],3)	72.2	69.4	76.1	70.8	29.2
knnc([],1)	69.4	65.4	74.0	67.4	32.6
knnc([],9)	63.6	71.3	73.0	67.4	32.6
knnc([],13)	64.7	74.6	73.0	69.7	30.3
knnc([],11)	63.6	74.2	73.9	68.9	31.1
knnc([],15)	62.5	74.6	72.7	68.6	31.4
knnc([],19)	61.4	75.7	72.8	68.6	31.4
knnc([],17)	61.7	77.9	72.2	69.8	30.2
knnc([],21)	60.3	78.7	73.0	69.5	30.5
knnc([])	60.3	79.8	72.1	70.0	30.0
knnc([],5)	60.6	68.7	73.6	64.6	35.4
knnc([],7)	58.3	69.0	70.7	63.7	36.3

Tabulka A.1: k -NN bez použití selekce

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc([],1)	76.1	67.5	77.1	71.8	28.2
knnc([],21)	64.2	78.3	74.3	71.2	28.8
knnc([],7)	68.3	66.5	75.0	67.4	32.6
knnc([],5)	66.4	68.3	74.8	67.3	32.7
knnc([],9)	65.3	68.3	74.8	66.8	33.2
knnc([])	66.4	71.3	73.5	68.8	31.2
knnc([],11)	65.3	70.6	74.9	67.9	32.1
knnc([],15)	64.2	73.1	73.7	68.6	31.4
knnc([],19)	64.2	74.9	75.4	69.5	30.5
knnc([],17)	63.1	73.8	73.8	68.4	31.6
knnc([],3)	71.9	62.0	73.7	67.0	33.0
knnc([],13)	60.8	71.7	72.7	66.3	33.8

Tabulka A.2: k -NN s použitím funkce *featselp*, parametr *knnc*

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc([],1)	83.1	72.7	79.6	77.9	22.1
knnc([])	78.6	71.6	76.0	75.1	24.9
knnc([],9)	69.2	69.7	73.9	69.4	30.6
knnc([],11)	69.2	67.9	72.0	68.5	31.5
knnc([],13)	68.1	70.1	71.9	69.1	30.9
knnc([],15)	66.9	71.5	71.1	69.2	30.8
knnc([],7)	71.1	66.4	74.0	68.8	31.2
knnc([],3)	73.1	63.1	73.7	68.1	31.9
knnc([],17)	66.9	68.2	70.6	67.6	32.4
knnc([],19)	63.9	70.4	70.9	67.1	32.9
knnc([],21)	63.9	71.5	71.3	67.7	32.3
knnc([],5)	72.5	60.9	74.8	66.7	33.3

Tabulka A.3: k -NN s použitím funkce *featselp*, bez parametru

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc([],1)	77.8	71.3	78.9	74.5	25.5
knnc([],5)	70.0	69.1	76.5	69.6	30.4
knnc([])	71.1	72.4	75.4	71.7	28.3
knnc([],9)	68.9	72.8	74.9	70.8	29.2
knnc([],13)	68.6	74.6	75.1	71.6	28.4
knnc([],21)	65.6	76.8	73.4	71.2	28.8
knnc([],19)	64.4	76.4	74.8	70.4	29.6
knnc([],7)	67.8	68.7	74.9	68.3	31.7
knnc([],11)	65.6	69.4	74.5	67.5	32.5
knnc([],15)	64.4	73.9	74.6	69.2	30.8
knnc([],17)	63.3	75.7	74.5	69.5	30.5
knnc([],3)	76.7	60.6	75.7	68.6	31.4

Tabulka A.4: k -NN s použitím funkce *featselb*, bez parametru

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc([],3)	74.7	73.8	76.2	74.3	25.7
knnc([],1)	72.5	69.7	74.1	71.1	28.9
knnc([],21)	65.6	73.8	77.1	69.7	30.3
knnc([],17)	65.6	73.8	77.1	69.7	30.3
knnc([],13)	66.4	74.9	76.6	70.7	29.3
knnc([],15)	64.4	74.9	76.3	69.7	30.3
knnc([],7)	67.5	68.3	76.2	67.9	32.1
knnc([],9)	65.3	70.2	77.2	67.7	32.3
knnc([],5)	71.7	65.3	76.4	68.5	31.5
knnc([],11)	63.3	72.0	76.1	67.7	32.3
knnc([],19)	64.4	73.8	76.9	69.1	30.9
knnc([])	61.1	76.7	73.8	68.9	31.1

Tabulka A.5: k -NN s použitím funkce *featselb*, parametr *knnc*

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc([],13)	64.7	71.9	73.1	68.3	31.7
knnc([],11)	65.8	74.1	73.6	70.0	30.0
knnc([],15)	63.6	73.0	72.3	68.3	31.7
knnc([],19)	63.6	75.2	71.7	69.4	30.6
knnc([],21)	62.5	74.1	73.3	68.3	31.7
knnc([],17)	62.5	74.1	72.6	68.3	31.7
knnc([])	58.3	81.9	72.3	70.1	29.9
knnc([],5)	62.5	65.6	74.7	64.1	35.9
knnc([],3)	62.2	66.7	71.9	64.5	35.5
knnc([],9)	64.7	69.7	72.5	67.2	32.8
knnc([],7)	62.5	68.6	71.4	65.5	34.5
knnc([],1)	69.4	63.1	72.8	66.3	33.7

Tabulka A.6: k -NN, pca bez parametru

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc([],1)	72.8	69.1	77.4	71.0	29.0
knnc([],19)	64.4	78.7	72.4	71.5	28.5
knnc([],9)	64.4	79.4	72.5	71.9	28.1
knnc([],21)	63.3	79.8	73.1	71.5	28.5
knnc([],13)	62.2	80.9	71.8	71.5	28.5
knnc([],3)	64.2	72.8	72.0	68.5	31.5
knnc([],17)	62.5	79.4	71.1	70.9	29.1
knnc([],11)	61.9	79.4	71.4	70.7	29.3
knnc([],15)	62.2	80.5	70.8	71.3	28.7
knnc([])	57.8	81.6	69.9	69.7	30.3
knnc([],5)	61.1	69.0	69.9	65.1	34.9
knnc([],7)	60.0	70.2	69.6	65.1	34.9

Tabulka A.7: k -NN snížení dimenze funkcí pca na 2-D

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc([],15)	65.8	76.0	73.0	70.9	29.1
knnc([])	62.5	80.8	73.5	71.6	28.4
knnc([],11)	64.7	72.0	73.3	68.4	31.6
knnc([],19)	64.4	73.8	74.2	69.1	30.9
knnc([],13)	64.7	74.2	73.5	69.5	30.5
knnc([],21)	62.5	76.0	74.0	69.3	30.7
knnc([],17)	61.4	75.3	73.5	68.4	31.6
knnc([],3)	65.6	66.4	70.0	66.0	34.0
knnc([],5)	63.6	64.9	73.9	64.3	35.7
knnc([],1)	67.8	64.6	75.5	66.2	33.8
knnc([],9)	62.5	70.2	72.2	66.3	33.7
knnc([],7)	61.7	70.9	74.0	66.3	33.7

Tabulka A.8: k -NN snížení dimenze funkcí *pca* na 5-D

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc([],15)	65.8	76.0	73.0	70.9	29.1
knnc([])	62.5	80.8	73.5	71.6	28.4
knnc([],11)	64.7	72.0	73.3	68.4	31.6
knnc([],19)	64.4	73.8	74.2	69.1	30.9
knnc([],13)	64.7	74.2	73.5	69.5	30.5
knnc([],21)	62.5	76.0	74.0	69.3	30.7
knnc([],17)	61.4	75.3	73.5	68.4	31.6
knnc([],3)	65.6	66.4	70.0	66.0	34.0
knnc([],5)	63.6	64.9	73.9	64.3	35.7
knnc([],1)	67.8	64.6	75.5	66.2	33.8
knnc([],9)	62.5	70.2	72.2	66.3	33.7
knnc([],7)	61.7	70.9	74.0	66.3	33.7

Tabulka A.9: k -NN snížení dimenze funkcí *pca* na 10-D

A.1 SVM

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 's')	65.8	77.1	75.7	71.5	28.5
svc([], 'e')	62.8	80.1	78.7	71.4	28.6
svc([], 'r')	54.2	83.4	77.8	68.8	31.2
svc([])	48.9	85.6	78.6	67.3	32.7
svc([], 'p')	48.9	85.6	78.6	67.3	32.7
svc([], 'h')	48.9	85.6	78.6	67.3	32.7
svc([], 'c')	4.2	92.9	23.9	48.5	51.5
svc([], 'm')	4.2	92.9	23.9	48.5	51.5
svc([], 'd')	4.2	92.9	23.9	48.5	51.5
svc([], 'o')	4.2	94.3	24.7	49.2	50.8

Tabulka A.10: SVM bez použití selekce

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 'e')	64.7	79.0	78.6	71.8	28.2
svc([], 'r')	59.4	82.6	75.2	71.0	29.0
svc([])	58.3	82.6	75.7	70.5	29.5
svc([], 'p')	58.3	82.6	75.7	70.5	29.5
svc([], 'h')	58.3	82.6	75.7	70.5	29.5
svc([], 's')	58.3	84.5	75.7	71.4	28.6
svc([], 'o')	4.2	92.9	24.7	48.5	51.5
svc([], 'd')	4.2	92.9	24.6	48.5	51.5
svc([], 'c')	4.2	92.9	24.5	48.5	51.5
svc([], 'm')	4.2	92.9	24.5	48.5	51.5

Tabulka A.11: SVM s použitím funkce *featselb*, bez parametru

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([])	65.0	77.5	74.4	71.3	28.7
svc([], 'p')	65.0	77.5	74.4	71.3	28.7
svc([], 'h')	65.0	77.5	74.4	71.3	28.7
svc([], 'e')	59.7	79.8	75.8	69.7	30.3
svc([], 's')	58.6	83.5	74.7	71.1	28.9
svc([], 'r')	58.6	79.8	74.5	69.2	30.8
svc([], 'd')	2.5	95.0	26.8	48.8	51.3
svc([], 'o')	1.7	95.0	25.4	48.3	51.7
svc([], 'c')	1.7	95.0	25.3	48.3	51.7
svc([], 'm')	1.7	95.0	25.3	48.3	51.7

Tabulka A.12: SVM s použitím funkce *featselb*, parametr *svc*

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 'e')	66.9	76.7	76.8	71.8	28.2
svc([])	49.7	83.7	75.8	66.7	33.3
svc([], 'p')	49.7	83.7	75.8	66.7	33.3
svc([], 'h')	49.7	83.7	75.8	66.7	33.3
svc([], 'r')	49.7	84.8	75.8	67.3	32.7
svc([], 's')	46.4	87.1	75.6	66.7	33.3
svc([], 'd')	5.0	91.4	24.4	48.2	51.8
svc([], 'o')	4.2	92.1	25.6	48.2	51.8
svc([], 'c')	4.2	92.1	24.5	48.2	51.8
svc([], 'm')	4.2	92.1	24.5	48.2	51.8

Tabulka A.13: SVM s použitím funkce *featselp*, bez parametru

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 'e')	63.9	77.9	76.8	70.9	29.1
svc([])	57.2	82.3	75.8	69.8	30.2
svc([], 'p')	57.2	82.3	75.8	69.8	30.2
svc([], 'h')	57.2	82.3	75.8	69.8	30.2
svc([], 's')	51.9	89.0	75.9	70.5	29.5
svc([], 'r')	57.2	81.2	75.6	69.2	30.8
svc([], 'd')	5.0	92.1	24.3	48.6	51.4
svc([], 'c')	5.0	92.1	24.3	48.6	51.4
svc([], 'm')	5.0	92.1	24.3	48.6	51.4
svc([], 'o')	5.0	92.1	23.7	48.6	51.4

Tabulka A.14: SVM s použitím funkce *featselp*, parametr *svc*

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 's')	65.8	76.7	77.0	71.3	28.7
svc([], 'e')	61.4	79.0	79.4	70.2	29.8
svc([], 'r')	51.9	84.1	77.9	68.0	32.0
svc([])	48.9	85.6	77.7	67.2	32.8
svc([], 'p')	48.9	85.6	77.7	67.2	32.8
svc([], 'h')	48.9	85.6	77.7	67.2	32.8
svc([], 'c')	5.0	92.1	21.8	48.6	51.4
svc([], 'm')	5.0	92.1	21.8	48.6	51.4
svc([], 'o')	4.2	92.1	23.8	48.2	51.8
svc([], 'd')	4.2	92.1	23.3	48.2	51.8

Tabulka A.15: SVM, snížení dimenze funkcí *pca*, bez parametru

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 's')	66.7	74.5	75.2	70.6	29.4
svc([], 'e')	61.1	76.7	75.3	68.9	31.1
svc([], 'r')	52.2	83.0	73.4	67.6	32.4
svc([])	51.4	81.9	74.1	66.6	33.4
svc([], 'p')	51.4	81.9	74.1	66.6	33.4
svc([], 'h')	51.4	81.9	74.1	66.6	33.4
svc([], 'o')	3.3	92.1	24.9	47.7	52.3
svc([], 'c')	3.3	92.1	24.8	47.7	52.3
svc([], 'm')	3.3	92.1	24.8	47.7	52.3
svc([], 'd')	3.3	92.1	24.7	47.7	52.3

Tabulka A.16: SVM, snížení dimenze funkcí *pca* na 2D

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 's')	65.3	76.7	76.3	71.0	29.0
svc([], 'e')	60.8	76.7	77.5	68.8	31.2
svc([], 'r')	51.9	83.0	76.7	67.5	32.5
svc([])	47.8	86.0	77.4	66.9	33.1
svc([], 'p')	47.8	86.0	77.4	66.9	33.1
svc([], 'h')	47.8	86.0	77.4	66.9	33.1
svc([], 'o')	2.5	92.1	25.5	47.3	52.7
svc([], 'd')	2.5	92.1	24.4	47.3	52.7
svc([], 'c')	2.5	92.1	23.4	47.3	52.7
svc([], 'm')	2.5	92.1	23.4	47.3	52.7

Tabulka A.17: SVM, snížení dimenze funkcí *pca* na 5D

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc([], 's')	66.7	77.1	76.3	71.9	28.1
svc([], 'e')	62.2	78.7	78.9	70.4	29.6
svc([], 'r')	50.0	83.8	77.4	66.9	33.1
svc([])	48.9	88.3	76.5	68.6	31.4
svc([], 'p')	48.9	88.3	76.5	68.6	31.4
svc([], 'h')	48.9	88.3	76.5	68.6	31.4
svc([], 'o')	3.3	92.9	23.9	48.1	51.9
svc([], 'd')	3.3	92.9	23.4	48.1	51.9
svc([], 'c')	3.3	93.6	22.5	48.5	51.5
svc([], 'm')	3.3	93.6	22.5	48.5	51.5

Tabulka A.18: SVM, snížení dimenze funkcí *pca* na 10D

A.2 Rozhodovací stromy

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
treec([], 'maxcrit', NaN)	73.1	61.9	74.0	67.5	32.5
treec	67.5	66.3	73.3	66.9	33.1
treec([], 'inferit', NaN)	64.2	62.4	70.9	63.3	36.7
treec([], 'fishcrit', NaN)	59.7	69.0	69.5	64.3	35.7
treec([], NaN, NaN)	55.3	61.2	66.7	58.2	41.8

Tabulka A.19: Rozhodovací strom bez použití selekce

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
<code>treec([], 'fishcrit', NaN)</code>	59.7	72.0	69.9	65.9	34.1
<code>treec([], 'infcrit', NaN)</code>	64.2	65.4	69.2	64.8	35.2
<code>treec([], 'maxcrit', NaN)</code>	63.6	59.1	65.1	61.4	38.6
<code>treec([], NaN, NaN)</code>	58.6	63.2	64.2	60.9	39.1
<code>treec</code>	59.7	59.8	64.8	59.8	40.2

Tabulka A.20: Rozhodovací strom s použitím funkce *featselb*, bez parametru

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
<code>treec</code>	64.2	60.2	69.0	62.2	37.8
<code>treec([], 'maxcrit', NaN)</code>	64.2	60.2	69.0	62.2	37.8
<code>treec([], 'fishcrit', NaN)</code>	67.5	58.7	73.8	63.1	36.9
<code>treec([], 'infcrit', NaN)</code>	61.4	62.3	69.0	61.8	38.2
<code>treec([], NaN, NaN)</code>	61.9	59.8	69.4	60.9	39.1

Tabulka A.21: Rozhodovací strom s použitím funkce *featselp*, bez parametru

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
<code>treec([], 'fishcrit', NaN)</code>	60.8	62.7	61.1	61.8	38.2
<code>treec([], 'infcrit', NaN)</code>	60.6	65.6	60.8	63.1	36.9
<code>treec([], NaN, NaN)</code>	60.6	65.6	60.8	63.1	36.9
<code>treec</code>	57.5	52.4	56.5	54.9	45.1
<code>treec([], 'maxcrit', NaN)</code>	56.4	51.3	52.2	53.8	46.2

Tabulka A.22: Rozhodovací strom s použitím funkce *featselb*, parametr *treec*

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
treec([], 'fishcrit', NaN)	62.8	70.2	70.6	66.5	33.5
treec([], NaN, NaN)	59.7	62.4	67.2	61.1	38.9
treec([], 'infcrit', NaN)	61.9	56.8	66.0	59.4	40.6
treec	60.3	55.3	60.7	57.8	42.2
treec([], 'maxcrit', NaN)	55.8	53.1	57.5	54.5	45.5

Tabulka A.23: Rozhodovací strom, snížení dimenze funkcí *pca*, bez parametru

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
treec([], 'infcrit', NaN)	61.9	72.2	66.3	67.1	32.9
treec([], NaN, NaN)	59.7	72.2	67.4	66.0	34.0
treec([], 'fishcrit', NaN)	52.8	80.5	70.1	66.6	33.4
treec	63.9	63.0	68.8	63.5	36.5
treec([], 'maxcrit', NaN)	65.0	63.0	68.3	64.0	36.0

Tabulka A.24: Rozhodovací strom, snížení dimenze funkcí *pca* na 2D

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
treec([], NaN, NaN)	68.1	66.8	76.7	67.4	32.6
treec	63.1	64.2	66.8	63.6	36.4
treec([], 'maxcrit', NaN)	63.1	64.2	66.3	63.6	36.4
treec([], 'infcrit', NaN)	63.1	64.5	70.5	63.8	36.2
treec([], 'fishcrit', NaN)	60.3	66.8	73.3	63.6	36.4

Tabulka A.25: Rozhodovací strom, snížení dimenze funkcí *pca* na 5D

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
treec([],NaN,NaN)	68.1	66.8	76.7	67.4	32.6
treec	63.1	64.2	66.8	63.6	36.4
treec([], 'maxcrit', NaN)	63.1	64.2	66.3	63.6	36.4
treec([], 'infcrit', NaN)	63.1	64.5	70.5	63.8	36.2
treec([], 'fishcrit', NaN)	60.3	66.8	73.3	63.6	36.4

Tabulka A.26: Rozhodovací strom, snížení dimenze funkcí *pca* na 10D

A.3 Základní kombinace

- K = knnc
- S = svc
- T = treec

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
svc - treec /meanc/	67.5	72.0	78.3	69.7	30.3
svc - treec /medianc/	67.5	72.0	78.3	69.7	30.3
knnc - treec /meanc/	66.4	73.1	76.1	69.7	30.3
knnc - treec /medianc/	66.4	73.1	76.1	69.7	30.3
knnc - svc - treec /meanc/	63.1	76.4	77.9	69.7	30.3
knnc	59.2	81.9	74.1	70.5	29.5
knnc - svc - treec /medianc/	57.2	86.7	78.3	72.0	28.0
treec	66.4	73.1	72.5	69.7	30.3
knnc - svc /meanc/	50.8	84.5	74.9	67.7	32.3
knnc - svc /medianc/	50.8	84.5	74.9	67.7	32.3
svc	48.6	87.5	77.4	68.0	32.0

Tabulka A.27: Kombinace bez použití selekce

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc	73.1	72.6	79.1	72.8	27.2
knnc - svc /meanc/	64.2	77.8	80.3	71.0	29.0
knnc - svc /medianc/	64.2	77.8	80.3	71.0	29.0
knnc - svc - treec /medianc/	60.0	80.0	76.2	70.0	30.0
knnc - treec /meanc/	66.1	69.4	78.5	67.7	32.3
knnc - treec /medianc/	66.1	69.4	78.5	67.7	32.3
knnc - svc - treec /meanc/	57.2	76.0	78.8	66.6	33.4
svc	58.6	82.2	73.7	70.4	29.6
treec	62.8	68.3	73.1	65.5	34.5
svc - treec /meanc/	61.7	69.4	74.3	65.5	34.5
svc - treec /medianc/	61.7	69.4	74.3	65.5	34.5

Tabulka A.28: Kombinace s použitím funkce *featselb*, bez parametru

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc	85.0	73.4	78.7	79.2	20.8
knnc - svc - treec /medianc/	68.1	76.3	78.4	72.2	27.8
knnc - svc /meanc/	62.2	80.8	82.0	71.5	28.5
knnc - svc /medianc/	62.2	80.8	82.0	71.5	28.5
knnc - svc - treec /meanc/	65.0	74.1	80.4	69.6	30.4
knnc - treec /meanc/	65.8	68.6	77.2	67.2	32.8
knnc - treec /medianc/	65.8	68.6	77.2	67.2	32.8
svc	50.6	83.7	77.2	67.1	32.9
svc - treec /meanc/	61.7	70.4	75.7	66.0	34.0
svc - treec /medianc/	61.7	70.4	75.7	66.0	34.0
treec	62.5	66.3	69.4	64.4	35.6

Tabulka A.29: Kombinace s použitím funkce *featselp*, bez parametru

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc	61.1	79.3	75.0	70.2	29.8
knnc - svc - treec /medianc/	54.4	83.7	73.2	69.1	30.9
svc	48.1	84.8	76.7	66.4	33.6
knnc - svc - treec /meanc/	60.0	72.3	71.3	66.2	33.8
knnc - svc /meanc/	48.1	82.6	76.9	65.3	34.7
knnc - svc /medianc/	48.1	82.6	76.9	65.3	34.7
svc - treec /meanc/	57.2	60.5	70.1	58.8	41.2
svc - treec /medianc/	57.2	60.5	70.1	58.8	41.2
knnc - treec /meanc/	55.3	58.3	68.8	56.8	43.2
knnc - treec /medianc/	55.3	58.3	68.8	56.8	43.2
treec	55.3	58.3	61.2	56.8	43.2

Tabulka A.30: Kombinace, snížení dimenze funkcí *pca*, bez parametru

Klasifikátor	Senzitivita [%]	Specifická [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc - svc /meanc/	63.3	80.5	75.7	71.9	28.1
knnc - svc /medianc/	63.3	80.5	75.7	71.9	28.1
knnc	61.4	80.1	73.8	70.7	29.3
knnc - svc - treec /meanc/	62.5	75.2	73.6	68.9	31.1
knnc - svc - treec /medianc/	56.9	79.7	75.7	68.3	31.7
svc	54.4	81.9	75.5	68.2	31.8
knnc - treec /medianc/	63.6	67.5	71.4	65.5	34.5
knnc - treec /meanc/	63.6	67.5	71.4	65.5	34.5
svc - treec /meanc/	58.1	67.9	72.3	63.0	37.0
svc - treec /medianc/	58.1	67.9	72.3	63.0	37.0
treec	63.3	61.2	67.5	62.3	37.7

Tabulka A.31: Kombinace, snížení dimenze funkcí *pca* na 2D

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc	62.8	77.5	73.9	70.1	29.9
knnc - svc - treec /meanc/	56.9	78.6	74.6	67.8	32.2
knnc - svc - treec /medianc/	53.9	83.7	74.7	68.8	31.2
knnc - svc /meanc/	50.6	83.0	77.1	66.8	33.2
knnc - svc /medianc/	50.6	83.0	77.1	66.8	33.2
svc	48.3	85.2	78.4	66.8	33.2
svc - treec /meanc/	56.7	70.8	73.6	63.7	36.3
svc - treec /medianc/	56.7	70.8	73.6	63.7	36.3
treec	62.2	59.4	65.9	60.8	39.2
knnc - treec /meanc/	62.2	59.0	71.6	60.6	39.4
knnc - treec /medianc/	62.2	59.0	71.6	60.6	39.4

Tabulka A.32: Kombinace, snížení dimenze funkcí *pca* na 5D

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
knnc	62.2	78.6	75.1	70.4	29.6
knnc - svc - treec /meanc/	60.3	81.9	74.1	71.1	28.9
knnc - svc - treec /medianc/	59.2	86.3	75.1	72.8	27.2
knnc - svc /meanc/	52.8	82.6	78.0	67.7	32.3
knnc - svc /medianc/	52.8	82.6	78.0	67.7	32.3
svc	48.9	88.2	77.7	68.5	31.5
svc - treec /meanc/	62.5	68.7	73.8	65.6	34.4
svc - treec /medianc/	62.5	68.7	73.8	65.6	34.4
knnc - treec /meanc/	63.6	62.0	71.7	62.8	37.2
knnc - treec /medianc/	63.6	62.0	71.7	62.8	37.2
treec	63.6	60.9	66.6	62.2	37.8

Tabulka A.33: Kombinace, snížení dimenze funkcí *pca* na 10D

A.4 Kombinace nejlepších klasifikátorů

Tabulka A.34: $K = \text{knnc}([],3)$; $S = \text{svc}([], 's')$; $T = \text{treec}$

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
K - S - T /meanc/	66.7	73.0	77.1	69.8	30.2
K - S /meanc/	63.6	76.0	76.5	69.8	30.2
K - S /medianc/	63.6	76.0	76.5	69.8	30.2
S	65.8	78.3	74.5	72.0	28.0
K - S - T /medianc/	64.7	79.0	76.7	71.8	28.2
T	65.8	66.0	72.4	65.9	34.1
S - T /meanc/	64.7	66.0	74.9	65.4	34.6
S - T /medianc/	64.7	66.0	74.9	65.4	34.6
K - T /meanc/	64.7	67.1	75.0	65.9	34.1
K - T /medianc/	64.7	67.1	75.0	65.9	34.1
K	64.2	69.0	71.9	66.6	33.4

Tabulka A.34: Kombinace nejlepších klas. bez použití selekce

Tabulka A.35: $K = \text{knnc}([],1)$; $S = \text{svc}([], 'e')$; $T = \text{treec}([], 'fishcrit', \text{str2num}('NaN'))$

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
K	74.7	79.3	80.6	77.0	23.0
K - S /meanc/	69.4	80.5	81.3	75.0	25.0
K - S /medianc/	69.4	80.5	81.3	75.0	25.0
K - S - T /meanc/	61.1	75.3	81.0	68.2	31.8
S	63.1	78.3	77.4	70.7	29.3
K - S - T /medianc/	63.1	78.3	77.8	70.7	29.3
K - T /meanc/	63.1	71.3	77.8	67.2	32.8
K - T /medianc/	63.1	71.3	77.8	67.2	32.8
S - T /meanc/	61.1	76.8	77.5	69.0	31.0
S - T /medianc/	61.1	76.8	77.5	69.0	31.0

Tabulka A.35: Kombinace nejlepších klas. s použitím funkce *featselb*, bez parametru

Tabulka A.36: K = knnc([],1); S = svc([], 'e'); T = treec

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
K	80.8	74.6	78.2	77.7	22.3
K - S /meanc/	68.1	77.1	80.8	72.6	27.4
K - S /medianc/	68.1	77.1	80.8	72.6	27.4
K - S - T /medianc/	70.3	73.1	76.8	71.7	28.3
K - S - T /meanc/	68.1	73.1	79.8	70.6	29.4
S - T /meanc/	64.7	76.0	78.1	70.4	29.6
S - T /medianc/	64.7	76.0	78.1	70.4	29.6
S	63.6	78.3	78.0	70.9	29.1
T	64.7	69.8	73.5	67.2	32.8
K - T /meanc/	63.6	69.4	78.1	66.5	33.5
K - T /medianc/	63.6	69.4	78.1	66.5	33.5

Tabulka A.36: Kombinace nejlepších klas. s použitím funkce *featselp*, bez parametru

Tabulka A.37: K = knnc([],11); S = svc([], 's'); T = treec([], 'fishcrit', str2num('NaN'))

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
K - S - T /meanc/	68.9	73.5	77.6	71.2	28.8
S - T /meanc/	64.2	74.6	78.2	69.4	30.6
S - T /medianc/	64.2	74.6	78.2	69.4	30.6
S	66.9	77.9	76.3	72.4	27.6
K - S - T /medianc/	64.7	75.7	77.3	70.2	29.8
K - S /meanc/	64.7	76.8	76.3	70.8	29.2
K - S /medianc/	64.7	76.8	76.3	70.8	29.2
K - T /meanc/	66.4	72.4	75.5	69.4	30.6
K - T /medianc/	66.4	72.4	75.5	69.4	30.6
T	65.3	71.3	73.0	68.3	31.7
K	63.6	71.3	73.0	67.4	32.6

Tabulka A.37: Kombinace nejlepších klas., snížení dimenze funkcí *pca*, bez parametru

Tabulka A.38: K = knnc([],1); S = svc([], 's'); T = treec([], 'infcrit', str2num('NaN'))

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
K - S /meanc/	70.6	77.5	78.9	74.0	26.0
K - S /medianc/	70.6	77.5	78.9	74.0	26.0
K	71.7	70.5	77.0	71.1	28.9
K - S - T /meanc/	67.2	76.0	78.5	71.6	28.4
S	65.8	75.7	76.0	70.8	29.2
K - T /meanc/	67.2	68.3	75.1	67.7	32.3
K - T /medianc/	67.2	68.3	75.1	67.7	32.3
K - S - T /medianc/	63.1	74.2	76.7	68.6	31.4
S - T /meanc/	59.4	72.8	74.6	66.1	33.9
S - T /medianc/	59.4	72.8	74.6	66.1	33.9
T	60.6	63.9	69.2	62.2	37.8

Tabulka A.38: Kombinace nejlepších klas., snížení dimenze funkcí *pca* na 2D

Tabulka A.39: K = knnc; S = svc([], 's'); T = treec([], str2num('NaN'), str2num('NaN'))

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
S	66.9	77.9	76.2	72.4	27.6
K - S - T /medianc/	63.6	78.7	75.7	71.1	28.9
K - S - T /meanc/	66.9	71.3	74.3	69.1	30.9
K - S /meanc/	60.3	77.9	75.4	69.1	30.9
K - S /medianc/	60.3	77.9	75.4	69.1	30.9
K	55.8	82.0	73.5	68.9	31.1
S - T /meanc/	63.3	67.9	74.5	65.6	34.4
S - T /medianc/	63.3	67.9	74.5	65.6	34.4
K - T /meanc/	61.1	64.6	73.4	62.9	37.1
K - T /medianc/	61.1	64.6	73.4	62.9	37.1
T	61.7	58.7	72.0	60.2	39.8

Tabulka A.39: Kombinace nejlepších klas., snížení dimenze funkcí *pca* na 5D

Tabulka A.40: K = knnc; S = svc([], 's'); T = treec([], str2num('NaN'), str2num('NaN'))

Klasifikátor	Senzitivita [%]	Specificita [%]	Plocha ROC [%]	Celková správnost [%]	Celková chyba [%]
S	66.9	77.9	76.2	72.4	27.6
K - S - T /medianc/	63.6	78.7	75.7	71.1	28.9
K - S - T /meanc/	66.9	71.3	74.3	69.1	30.9
K - S /meanc/	60.3	77.9	75.4	69.1	30.9
K - S /medianc/	60.3	77.9	75.4	69.1	30.9
K	55.8	82.0	73.5	68.9	31.1
S - T /meanc/	63.3	67.9	74.5	65.6	34.4
S - T /medianc/	63.3	67.9	74.5	65.6	34.4
K - T /meanc/	61.1	64.6	73.4	62.9	37.1
K - T /medianc/	61.1	64.6	73.4	62.9	37.1
T	61.7	58.7	72.0	60.2	39.8

Tabulka A.40: Kombinace nejlepších klas., snížení dimenze funkcí *pca* na

10D

Příloha B

Výpis použitých funkcí a jednoduchá demonstrace

Pro podrobnější informace využijte nápovedy v programu Matlab.
(napr.: help dataset)

B.1 dataset()

```
A=dataset(data,anotace);
```

Vstup

data - matice příznaků

anotace - vektor rozdělení do tříd 0 a 1 dle hodnoty pH

Výstup

A - dataset

B.2 pca()

```
w=pca(A,dimenze)
```

Vstup

A - dataset

XXIIPŘÍLOHA B. VÝPIS POUŽITÝCH FUNKCÍ A JEDNODUCHÁ DEMONSTRACE

dimenze - číselná hodnota požadované dimenze

Výstup

w - affíní PCA mapping

B.3 `featselb()`

```
[vyber,malice]=featselb(A,klasifikator)
```

Vstup

A - dataset

klasifikator - netrenovaný klasifikátor

Výstup

vyber - vybrané příznaky

matice - krok za krokem při rozhodování

B.4 `featselp()`

```
[vyber,malice]=featselp(A,klasifikator)
```

Vstup

A - dataset

klasifikator - netrenovaný klasifikátor

Výstup

vyber - vybrané příznaky

matice - krok za krokem při rozhodování

B.5 `knnc()`

```
W=knnc(A,k)
```

Vstup

A - dataset

k - počet nejbližších sousedů

Výstup

W - k -NN klasifikace

B.6 svc()

`W=svc(A)`

Vstup

A - dataset

Výstup

W - SVC klasifikace

B.7 kombinace pomocí mean a median

```
w={knn(A),svc(A),treec(A)} % Kombinace klasifikátorů
W=medianc(V);
W=meanc(V);
```

Vstup

V - množina klasifikátorů

Výstup

W - median / mean vstupních klasifikátorů

B.8 treec()

`W=treec(A, kriterium, prorezavani)`

Vstup

A - dataset

kriterium *prorezavani* Výstup

W - klasifikace binárním stromem

B.9 testc()

Nejdůležitější funkce, která dokáže spočítat všechny potřebné hodnoty.

```
[FN,FP]=testc(Test*W,'FN', trida); %
[TP,TN]=testc(Test*W,'TP', trida);

[Senzitivita,Specificita]=testc(Test*W,'sensitivity', trida);
[Presnost_POSITIVE,Uplnost_POSITIVE]=testc(Test*W,'precision', trida); %
```

Vstup

Test - dataset testovací množiny

W - natrénovaný klasifikátor

'*FN*' - algoritmická vlastnost matice záměn

'*TP*' - algoritmická vlastnost matice záměn

'*sensitivity*' - algoritmická vlastnost pro výpočet senzitivity a specificity

'*precision*' - algoritmická vlastnost pro určení přesnosti

Výstup

FN, FP, TP, TN - relativní četnosti

Senzitivita, Specificita - hodnoty relativní četnosti

Presnost_POSITIVE, Uplnost_POSITIVE - hodnoty přesnosti a úplnosti pro

řazení do pozitivní třídy

B.10 testauc()

Výpočet plochy pod křivkou ROC

```
E = testauc(A*W)
```

Vstup

A - dataset ke klasifikaci

W - klasifikátor

Výstup

E - Chyba, plocha pod křivkou ROC

Příloha C

Monitor srdeční činnosti plodu - STAN S31TM

C.1 Funkce a technické parametry

C.1.1 Funkce zahrnuté v CTG platformě

- Záznam vnějšího a vnitřního FHR
- Vnější TOCO s nastavitelnou citlivostí
- Dvojitá detekce FHR
- Záznam skutečné tepové frekvence (beat-to-beat) vnitřního FHR
- Velká, dotyková a "user friendly" obrazovka
- Upgradujte počítač platformy
- Komunikace s laserovými tiskárnami
- Unikátní digitální archivace záznamů
- Speciální zvukové a vizuální alarmy
- Kompatibilní s několika centrálními monitorovacími systémy
- Náhledí celého real-time záznamu CTG

Volitelné funkce k dispozici

- Unikátní ST analýza vln EKG plodu
- Integrovaní vitálních funkce rodičky
- Přesné čtečky čárových kódů pro pacienta ID nahrávání
- Pokročilé síťové konektivity
- Exkluzivní síťové archivace nahrávek
- Didaktický STAN Viewer software pro retrospektivní hodnocení nahrávek
- Real-time termální tiskové řešení
- Bezdrátové sledování kompatibilní s RimkusTM telemetrie
- Duální převodník pro externí FHR záznam dvojčat
- Flexibilní montážní řešení: vozík, montáž na zeď nebo stolní stojan

C.1.2 Základní technické parametry

Záznam fEKG:

- Maximální elektrodový rozdíl napětí $\pm 0,75$ VDC
- Vstupní rozsah ± 10 mV
- Pre-amplifier pásma 0.05 – 100 Hz (-3 dB)
- Potlačení síťové frekvence > 40 dB
- Beat-to-beat HR detekce
- HR v rozmezí 30 – 240 BPM
- HR s přesností ± 1 BPM (< 180 BPM), ± 2 BPM (> 180 BPM)
- průměrování 30-tepů EKG
- rozsah T / QRS poměru $-0,30$ až $+0,90$
- přesnost T / QRS $\pm 0,02$

- Bifázická indikace ST segmentů stupně 1, 2 a 3

Záznam US:

- Provozní frekvence 1,0 a 1,5 MHz
- Intenzita $< 20 \text{ mW / cm}^2$?
- HR v rozmezí 50 – 240 BPM
- HR s přesností $\pm 2 \text{ BPM}$ ($< 180 \text{ BPM}$), $\pm 3 \text{ BPM}$ ($> 180 \text{ BPM}$)

Záznam TOCO:

- Maximální zatížení 500 g
- Rozmezí v 0 – 100 jednotek
- Citlivost 0.32 jednotek / g (nominální, nastavitelné)
- Manuální výstup offset 5 jednotek
- Rozlišení 1 jednotka
- Přesnost $\pm 10\%$ udávané hodnoty

IUP ZAZNAMENÁVÁNÍ:

- Rozmezí 0 – 100 mmHg (0 – 13,33 kPa)
- Rozlišení 1 mmHg (0,13 kPa)
- Přesnost $\pm 5\%$ čtení

(oficiální popis produktu od společnosti Neoventa [7])

Příloha D

Obsah přiloženého CD

K této práci je přiloženo CD, na kterém jsou uloženy zdrojové kódy.

- Tables: složka se soubory tabulek(*.mat), pro jednotlivé klasifikátory
- TablesMini: složka zmenšených tabulek ze složky Tables
- data: složka použitých vstupních dat
- vypocty.m: hlavní program pro generování tabulek
- anotace.m: program pro zobrazení grafu anotace
- klasifikaceCTG.pdf: elektronická verze bakalářské práce