

Redukcia a klasifikácia húb

Rastislav Kamenický

Michal Jarábek

Michal Piovarči

Obsah

- Databáza
- Redukčné algoritmy
- Klasifikátory
- ROC krivky
- Vyhodnotenie

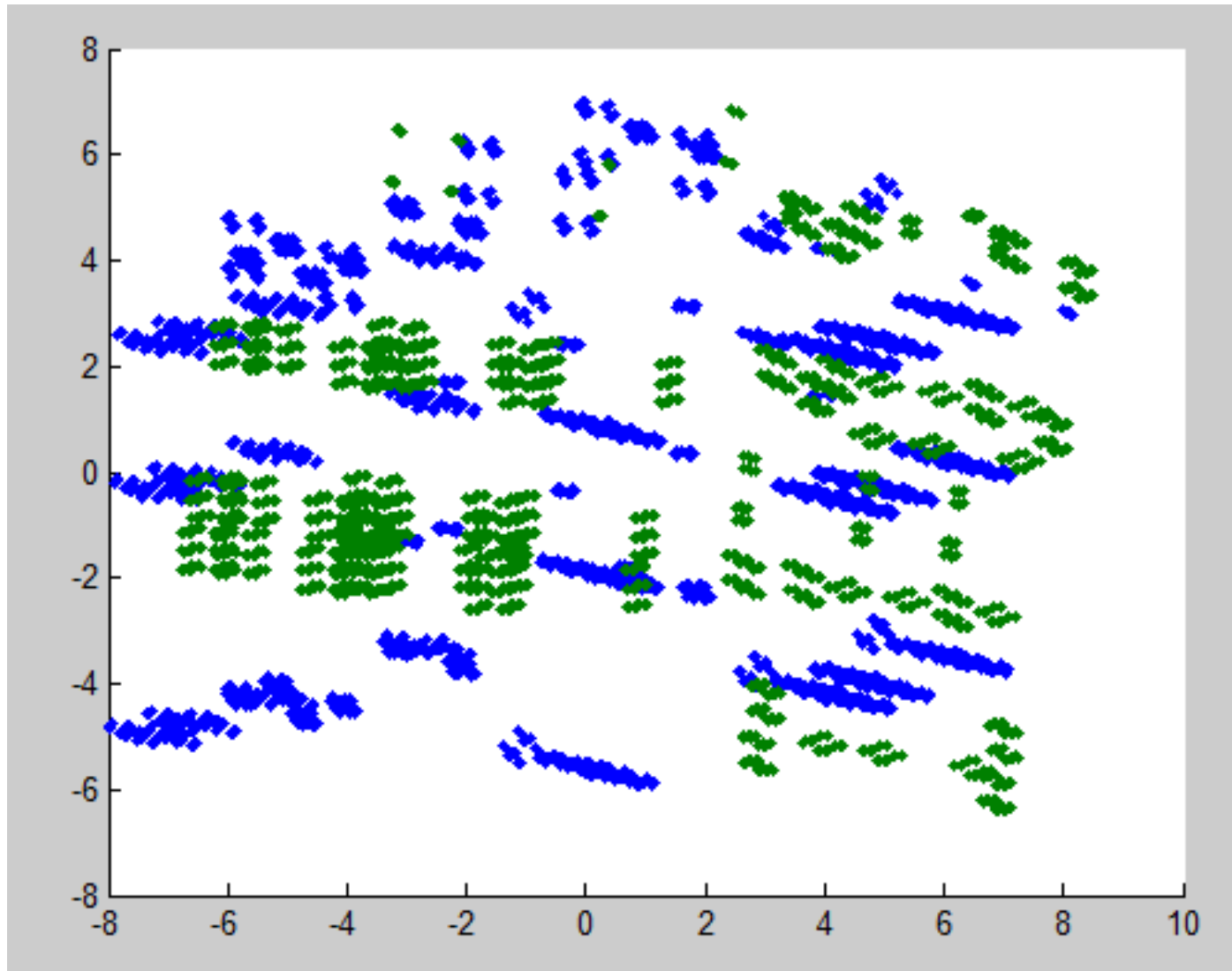
Databáza

- Pôvodná databáza
 - 8124 hríbov
 - 23 príznakov
- Hríby s chýbajúcim príznakom sme odstránili
 - ostalo 5936 hríbov
- Zisťované príznaky
 - Jedlosť -2 stavy
 - Nálezisko -7 stavov

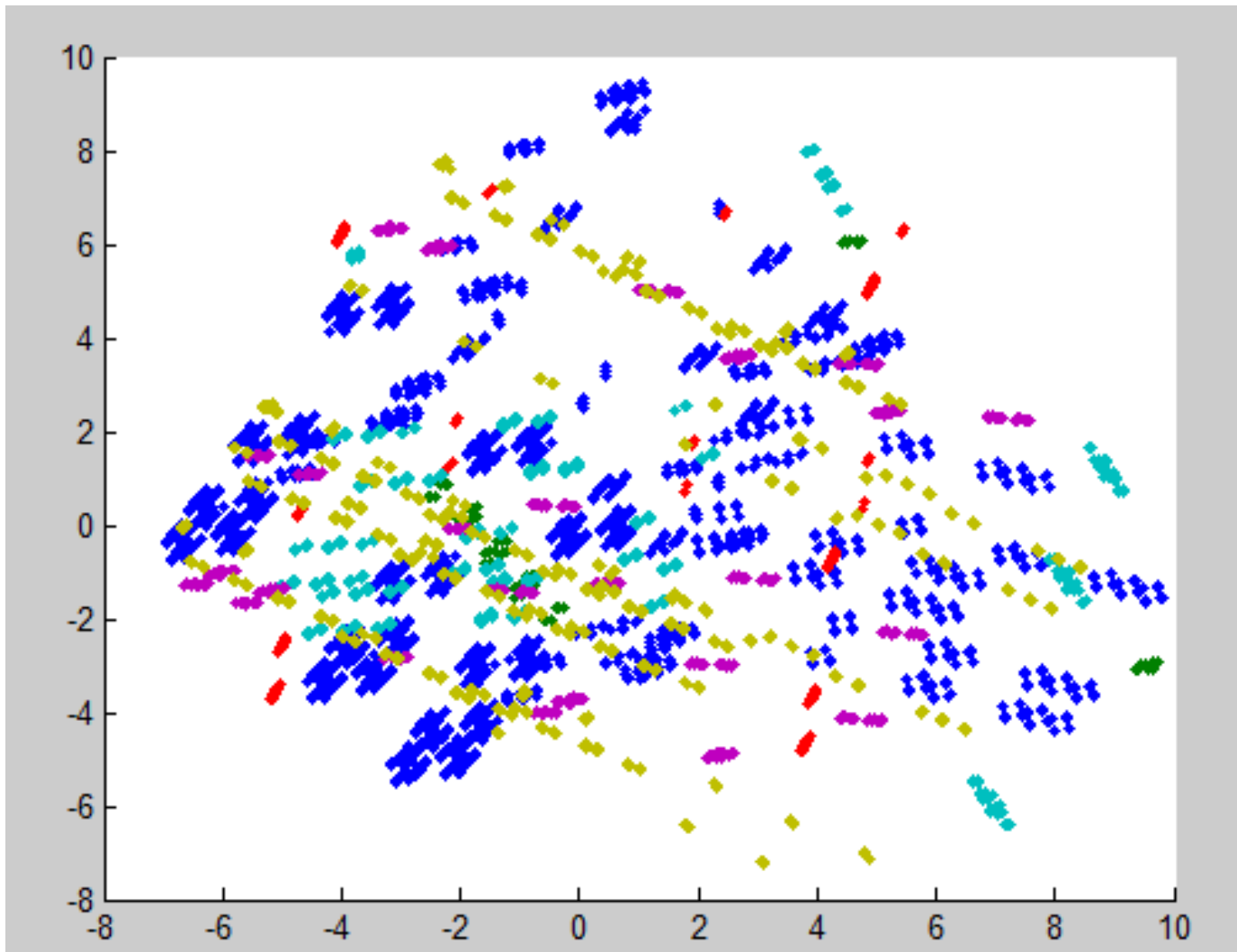
Redukčné algoritmy

- PCA
 - Povinný algoritmus
- ICA
 - Vybrali sme ho lebo bol kontrastom k PCA

Zredukované dáta



Zredukované dáta



Klasifikátory

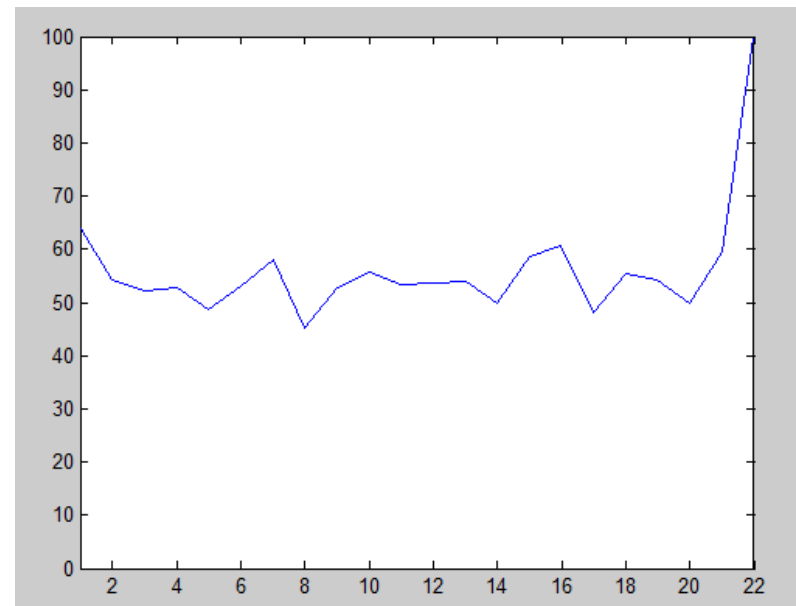
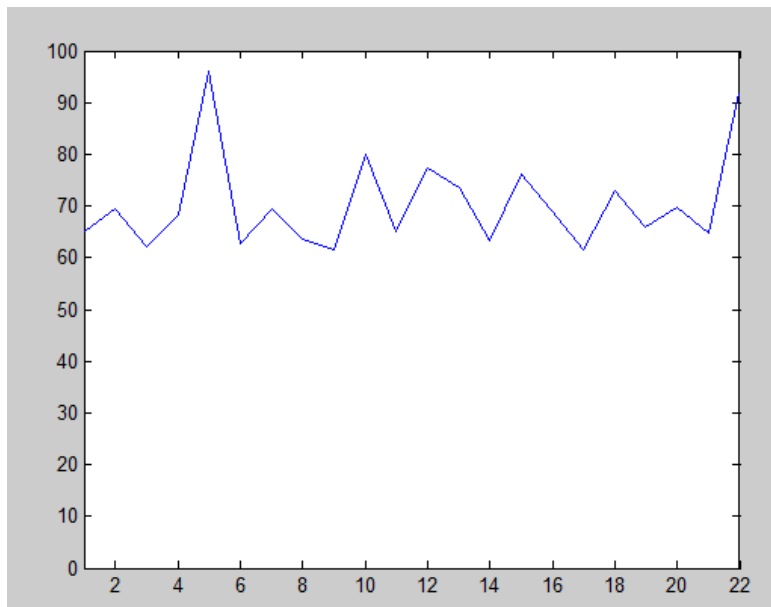
- Lineárny klasifikátor
- Neurónové siete
- SOM
 - Všetky 3 povinné
- Rozhodovací strom
 - Vybrali sme lebo sme mysleli, že vetvením bude vedieť dobre rozhodnúť o jedlosti

ROC krivky

- Pre jednotlivé algoritmy a zisťované príznaky budú ROC krivky uvedené naľavo s PCA na pravo s ICA redukciou

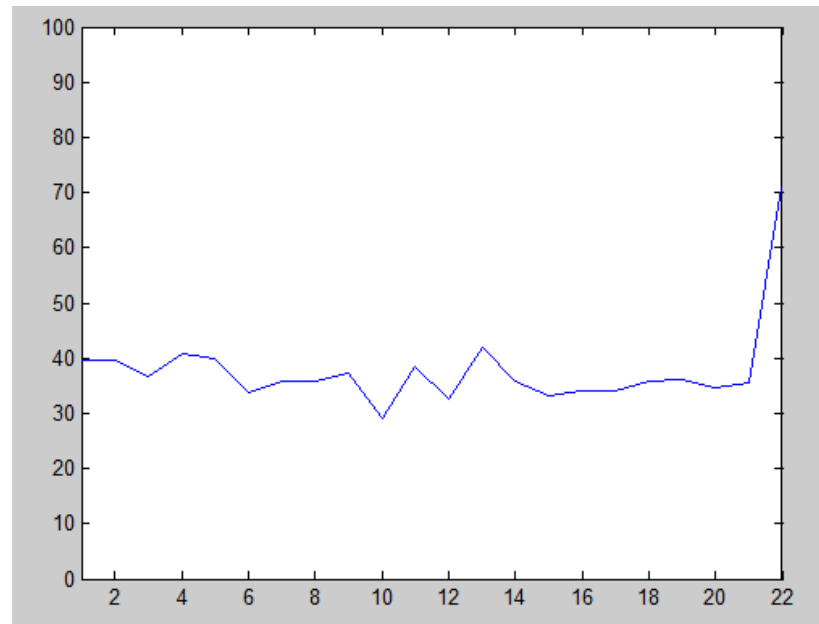
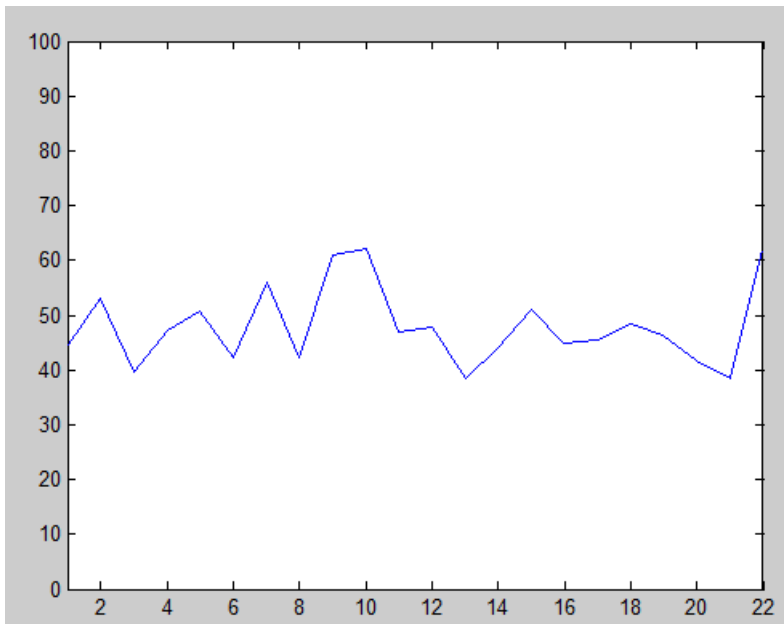
ROC krivky

- Neurónové siete pre jedlosť



ROC krivky

- Neurónové siete pre nálezisko



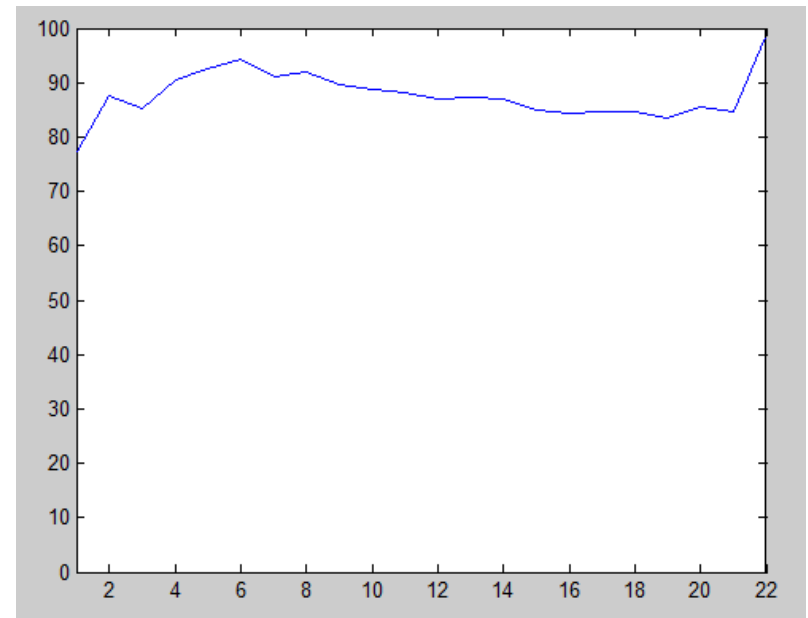
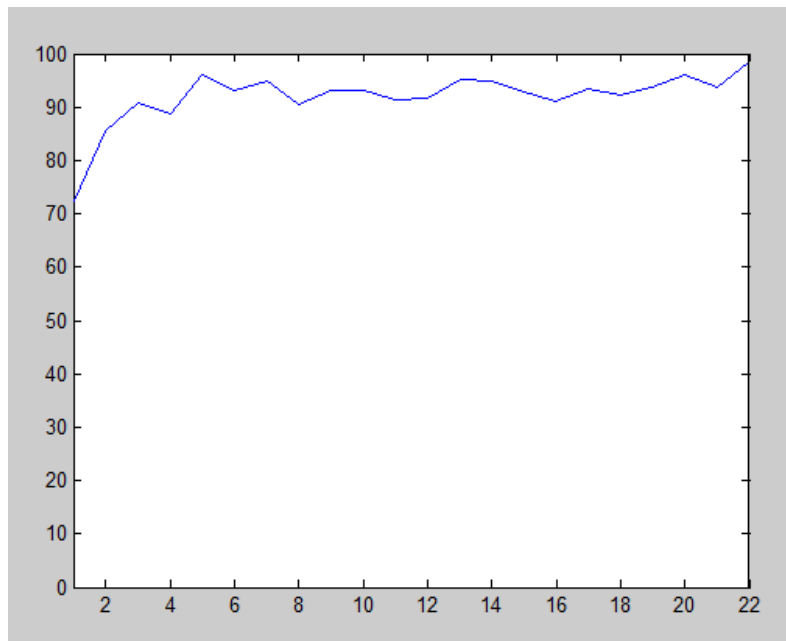
Confusion matice z neurónových sietí pre jedlosť a nálezisko

	Edible	Poisonous
Edible	741	2
Poisonous	12	432

	Trava	Listie	Luky	Cesty	Mesta	Lesy
Trava	334	0	39	31	69	31
Listie	0	18	0	0	3	0
Luky	12	0	19	0	0	0
Cesty	24	0	0	22	0	6
Mesta	0	0	0	0	0	0
Lesy	54	0	0	61	1	463

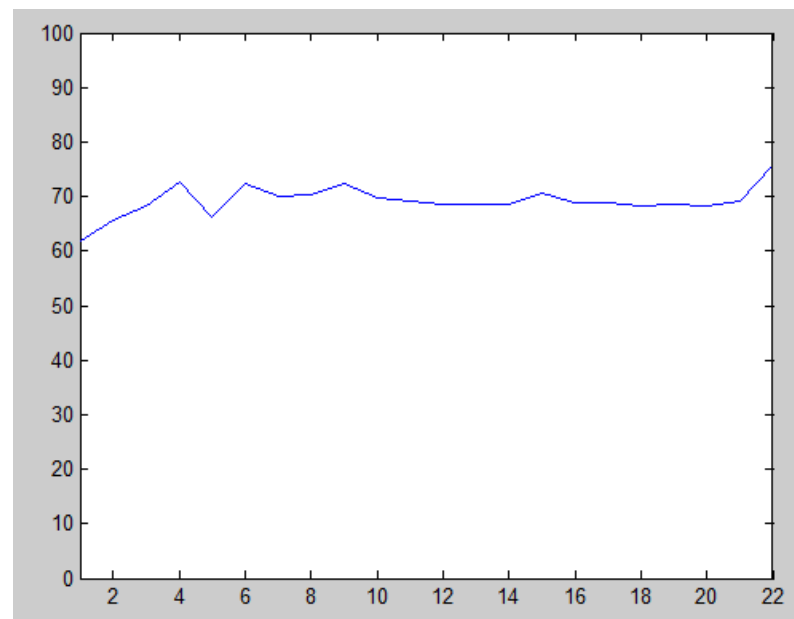
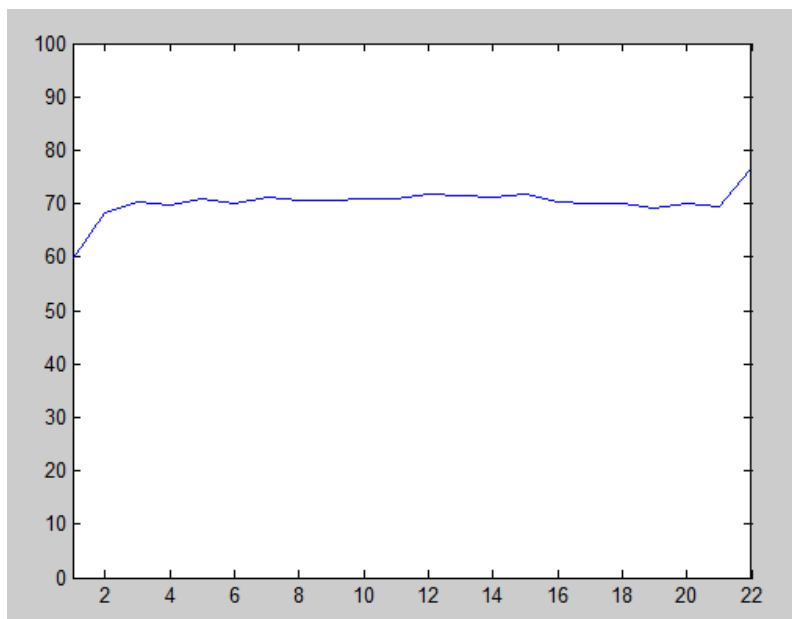
ROC krivky

- SOM pre jedlosť



ROC krivky

- SOM pre nálezisko



Confusion matice

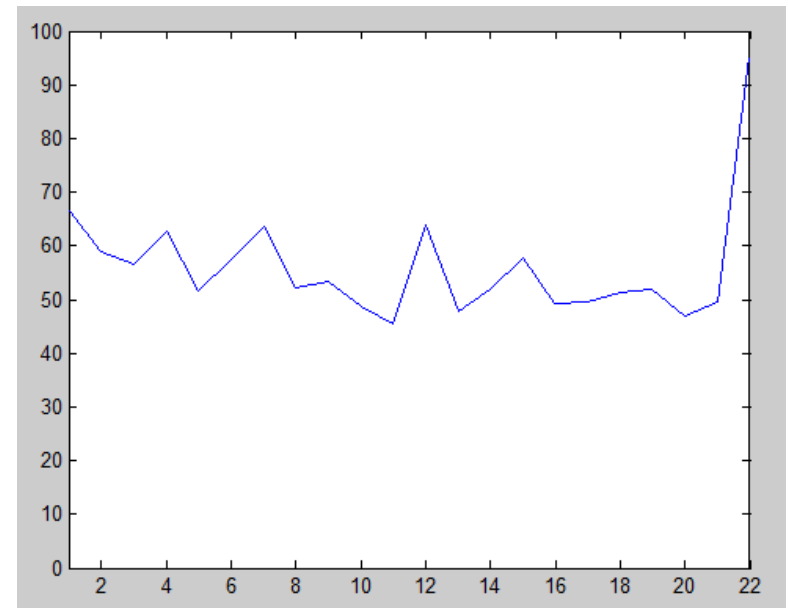
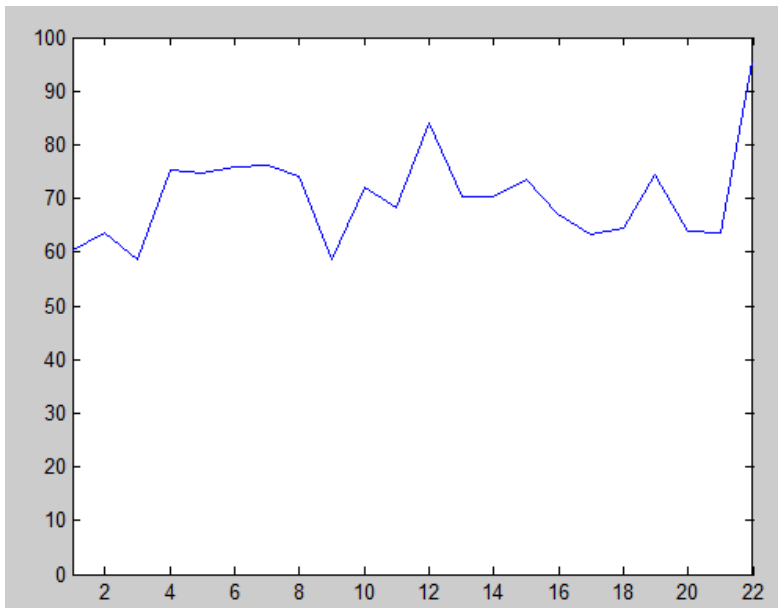
SOM pre jedlosť a nálezisko

	Edible	Poisonous
Edible	735	18
Poisonous	18	415

	Trava	Listie	Luky	Cesty	Mesta	Lesy
Trava	339	2	41	36	38	43
Listie	0	16	0	0	0	0
Luky	5	0	11	0	0	0
Cesty	44	0	0	57	0	29
Mesta	8	0	0	0	35	0
Lesy	27	0	7	20	0	429

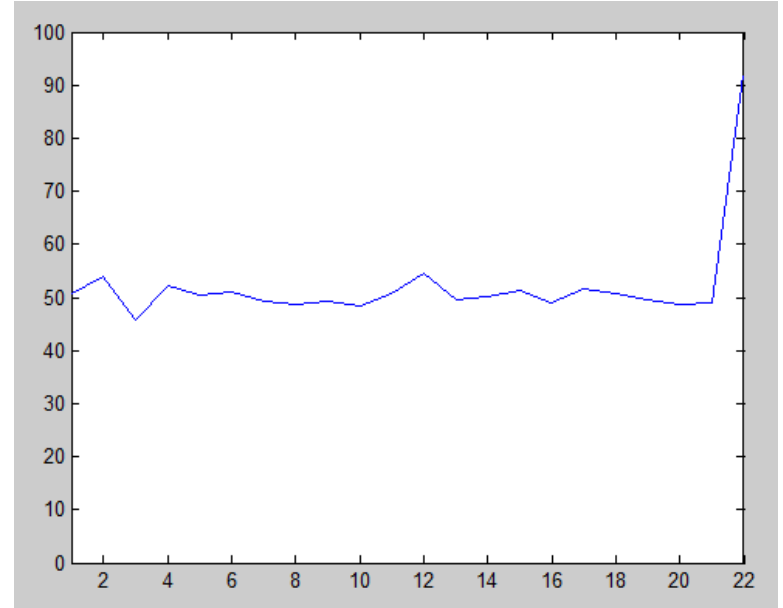
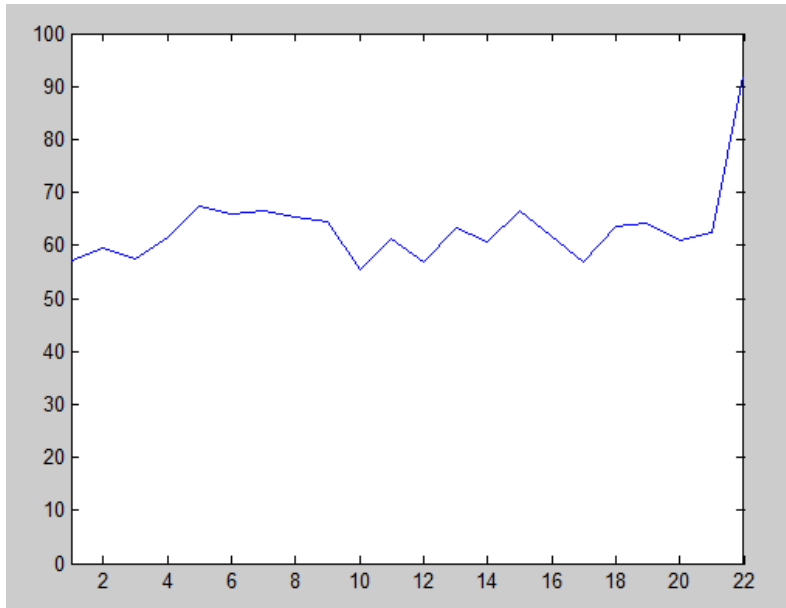
ROC krivky

- Lineárny klasifikátor pre jedlosť



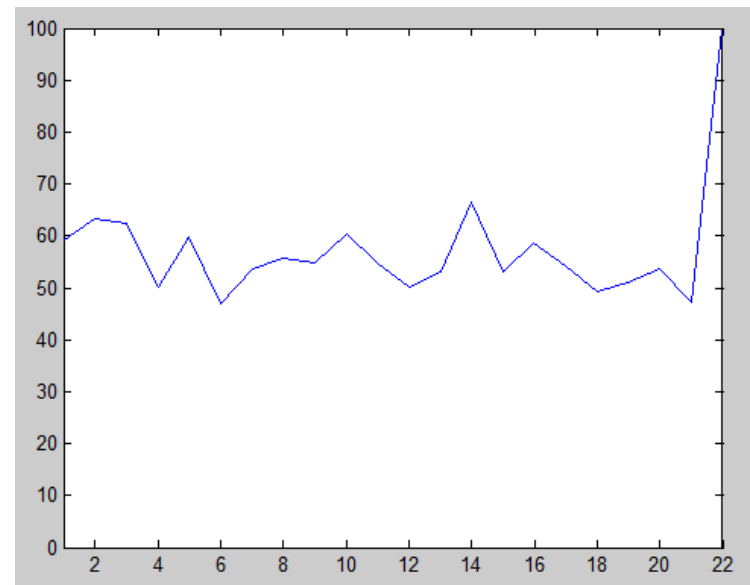
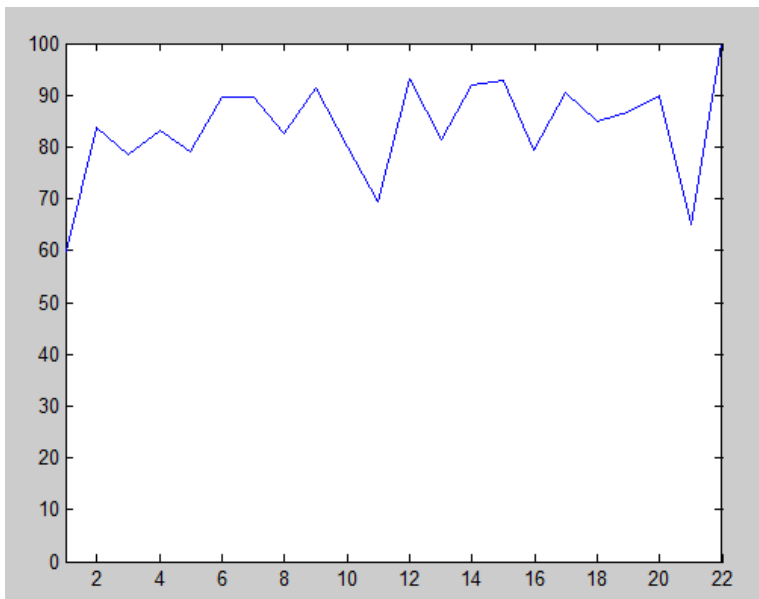
ROC krivky

- Lineárny klasifikátor pre nálezisko

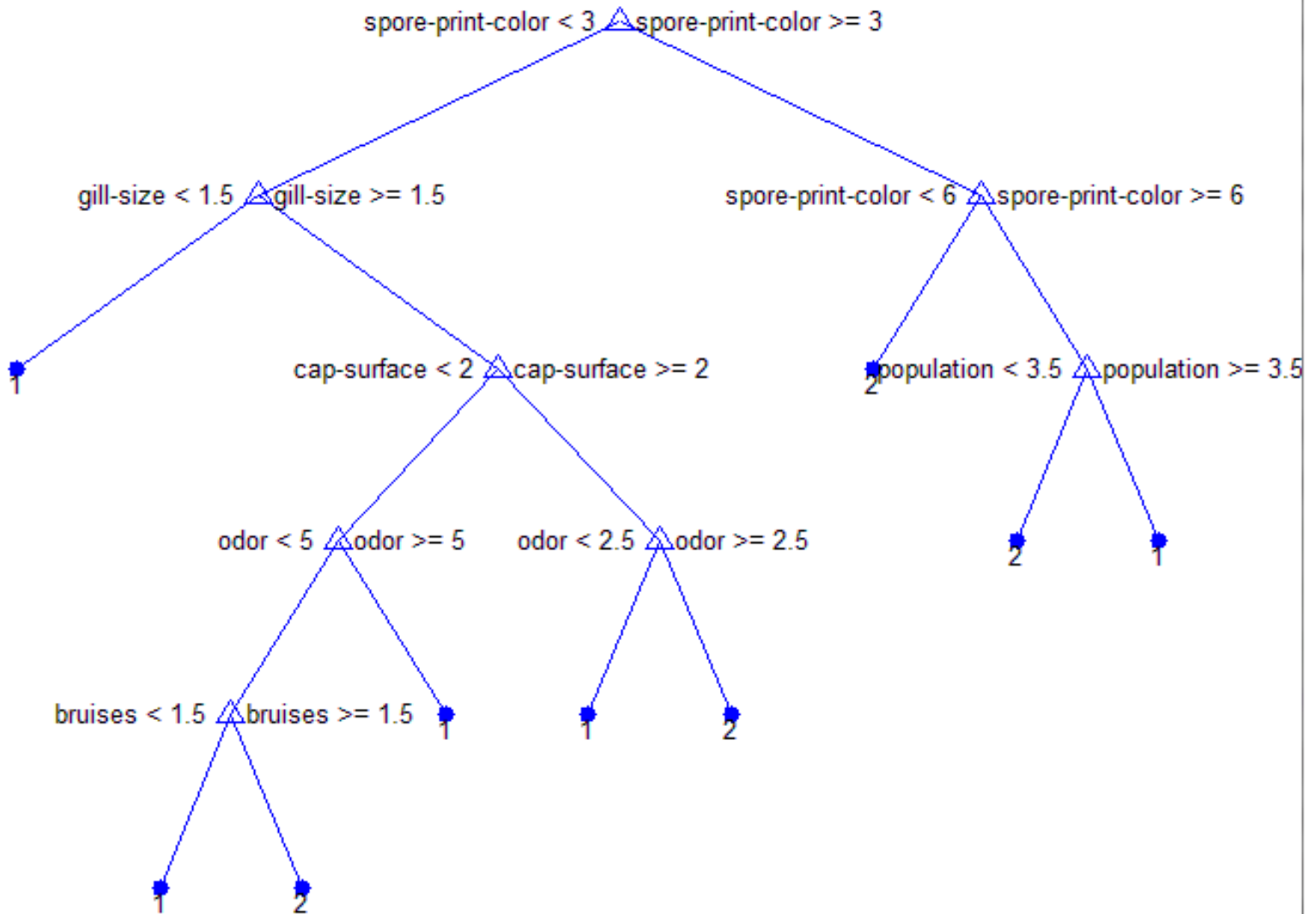


ROC krivky

- Rozhodovací strom pre jedlosť

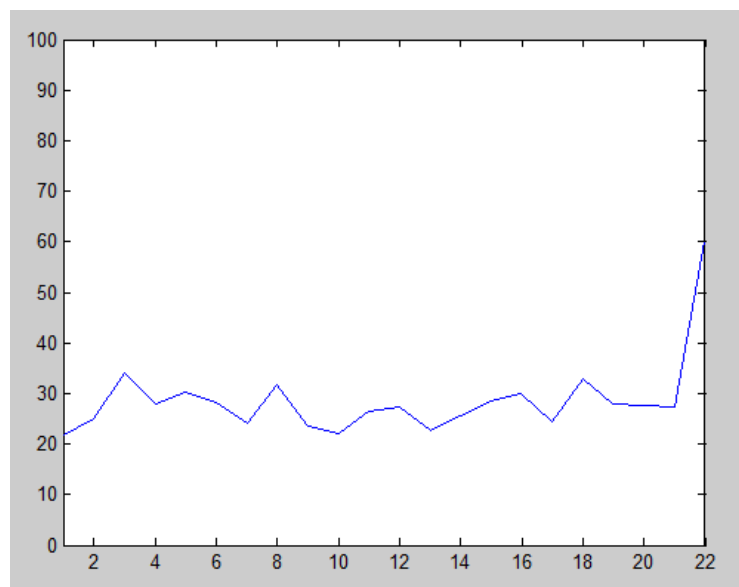
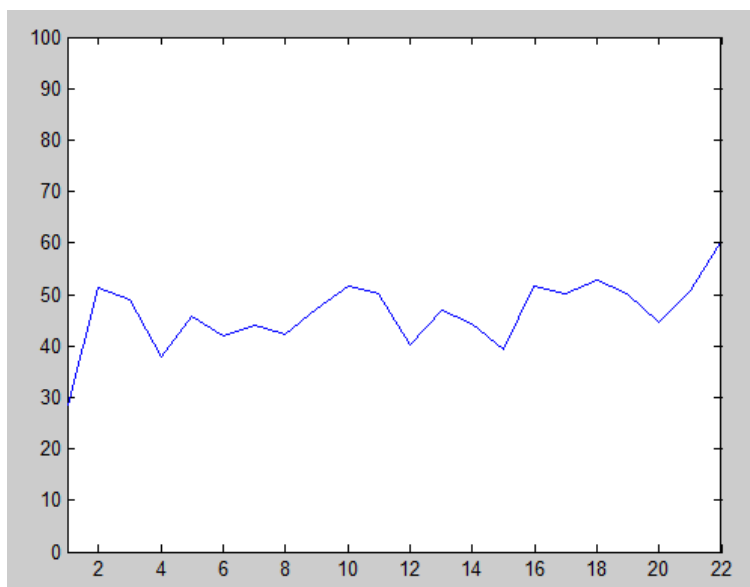


Rozhodovací strom jedlost

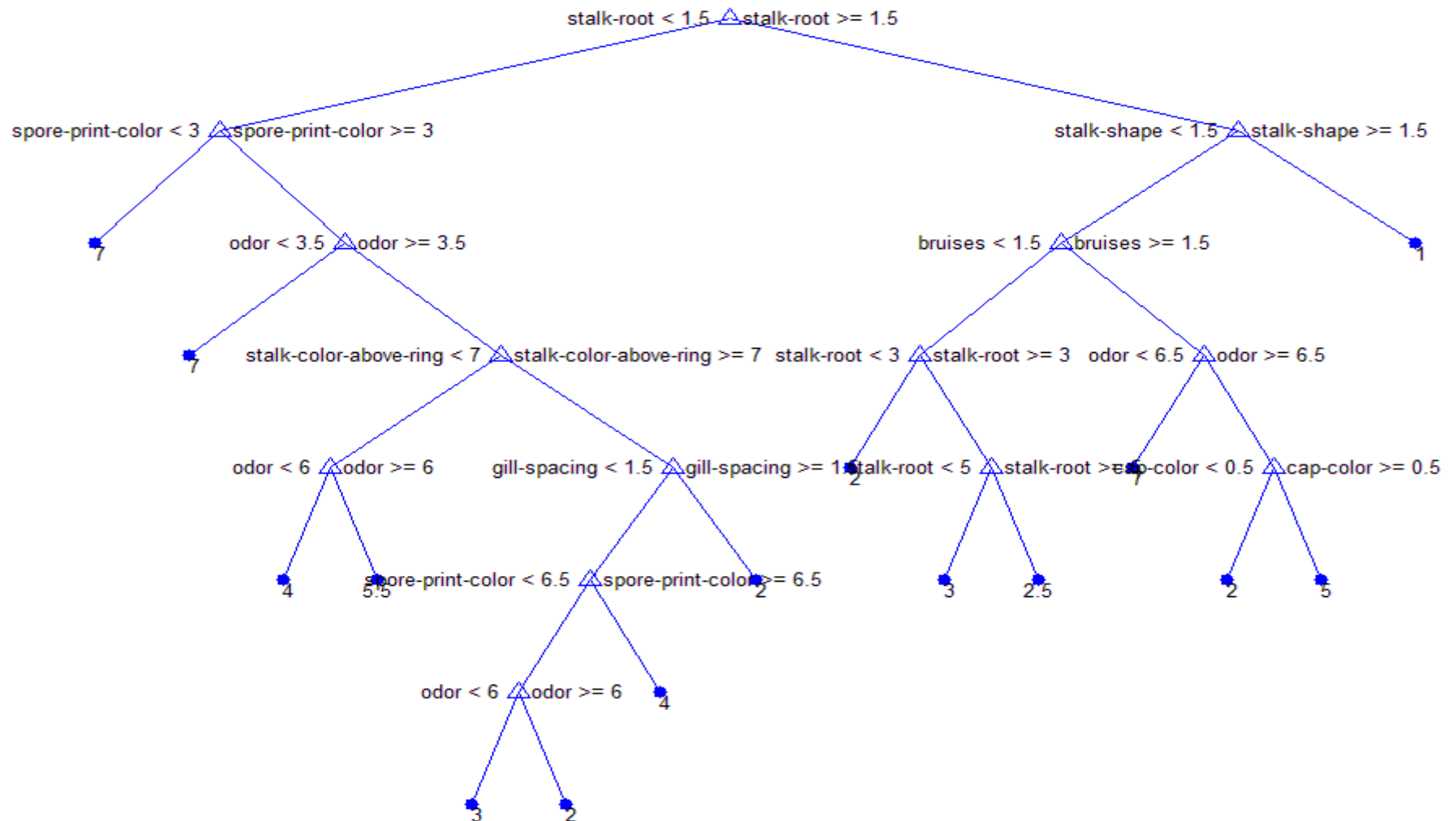


ROC krivky

- Rozhodovací strom pre nálezisko



Rozhodovací strom nálezisko



Vyhodnotenie

- PCA redukcia skoro vždy lepšia
- Najúspešnejší klasifikátor SOM
- Dáta lineárne separovateľné
 - Cez 90% úspešnosť s lineárnym klasifikátorom bez redukcie príznakov