

Rozpoznávanie obrazcov

šk.r. 2017-18

Matematika 3

Zuzana Berger Haladova

Termíny

Zadávanie projektov: 13.3.

Odovzávanie reportov: 13.5. (cca 2 strany, email)

Prezentácie: 15.5. (cca 10 minút)

Každý robí samostatný projekt

Projekt

Kroky:

1. Výber databázy
2. Redukcia príznakov
3. Klasifikácia dát
4. Vyhodnotenie

Databázy

Databazy <https://www.kaggle.com/datasets>

Príklad: Databáza rýb

Príznaky: Typ ryby, váha, dĺžka, šírka, počet plutiev, počet očí, počet šupín, výskyt

Čo by ste v tejto databáze zisťovali a prečo?

Čo by bol zlý výber a prečo?

Databázy

Požiadavky:

Min. 11príznačkov

Min. 500objektov

Čo chcete klasifikáciou zistiť!

Vybranú databázu je potrebné skonzultovať

Môžete si vybrať aj viacero databáz

Výber a redukcia príznakov

Výber príznakov: vyberieme podmnožinu

Dopredný, Spätný

Napr. Výber N príznakov s najvyšším skóre

Redukcia príznakov: transformujeme pôvodnú množinu do menejdimenzionálnej

Redukčné algoritmy

Unsupervised (minimize the information loss)

Latent Semantic Indexing (LSI): truncated SVD

Independent Component Analysis (ICA)

Principal Component Analysis (PCA)

Manifold learning algorithms (a manifold is a topological space which is locally Euclidean) –Nonlinear

Supervised (maximize the class discrimination)

Linear Discriminant Analysis (LDA)

Canonical Correlation Analysis (CCA)

Partial Least Squares (PLS)

Klasifikácia

Podstata

Zaradovanie objektov do kategórií na základe cieľovej premennej

Pre každý objekt množina premenných, jedna je cieľová

Cieľ

Nájsť model, ktorý opisuje cieľovú premennú ako funkciu vstupných premenných (prediktorov)

Trénovanie

Klasifikačného modelu vyžaduje znalosť hodnôt cieľovej premennej a prediktorov

Klasifikačné algoritmy

K-means

K najbližších susedov

Rozhodovacie stromy

Bayesovský klasifikátor

Lineárny klasifikátor

SVM / Nelineárny SVM

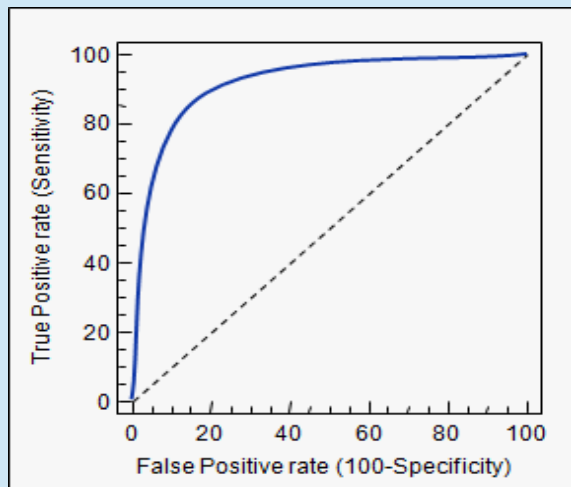
ANN / SOM

HMM

Vyhodnotenie výsledkov

ROC (Receiver Operating Characteristic) krivka

Matica zámen (Confusion matrix)



číslica	klasifikácia										
	1	2	3	4	5	6	7	8	9	0	R
1	87	0	0	0	1	0	0	0	0	0	0
2	0	88	1	0	0	0	0	0	1	1	1
3	0	0	75	1	0	0	0	10	4	0	3
4	0	0	0	79	0	0	0	0	0	0	0
5	0	0	0	0	79	6	0	0	0	4	1
6	0	0	0	0	8	80	1	0	0	2	0
7	0	1	0	0	0	0	83	0	0	0	0
8	0	0	15	0	0	1	0	65	7	0	0
9	0	0	4	0	0	0	0	10	71	0	1
0	0	0	0	1	0	1	0	0	0	90	1

Povinné časti projektu

Redukcia príznakov (min 1 metóda)

PCA, ICA...

Klasifikácia (min 2 metódy, alebo 1+ CNN)

Lineárny klasifikátor, Neurónové siete-ANN, SOM...

Vyhodnotenie

Matica zámen /ROC

Percentuálne vyhodnotenie

Report

Databáza

Ukážka, popis (počet prvkov, príznakov,...)

Použité redukčné algoritmy

Zdôvodnenie výberu

Stručný popis algoritmu(ov), výhody/nevýhody

Report

Použité klasifikátory

Zdôvodnenie výberu

Stručný popis algoritmov ich výhody/nevýhody

Porovnanie na databáze

Vyhodnotenie výsledkov

Porovnanie výsledkov klasifikácie z/bez redukčných algoritmov

Zdôvodnenie

Prezentácia

Databáza

Stručné info

Dôvod výberu

Redukčné algoritmy

Použité algoritmy

Dôvodvýberu

Ukážka účinnosti na vybranej databáze

Klady/zápory

Prezentácia

Klasifikátory

Použité algoritmy

Dôvod výberu

Ukážka účinnosti na vybranej databáze

Klady/zápory

Vyhodnotenie

Hodnotenie

40 bodov

Hodnotí sa

Zvládnutie danej problematiky

Prezentácia výsledkov projektu a samotné výsledky

Zvolené klasifikátoroy

Konzistencia

Redukcia priznakov

Zvolená databáza