

STROJNO UČENJE

4. Domaća Zadaća

Krešimir Špes

0036419866

ak. god. 2011. / 2012.

a) STATLOG odabrane datoteke (xaa, xab, xad, xae, xag, xai)

b)

STATLOG:

Logistička Regresija

accuracy: 79.29%					
	true van	true saab	true bus	true opel	class precision
pred. van	39	2	0	0	95.12%
pred. saab	0	27	0	16	62.79%
pred. bus	1	5	44	2	84.62%
pred. opel	0	9	0	24	72.73%
class recall	97.50%	62.79%	100.00%	57.14%	

SVM sa polinomnom jezgrom

accuracy: 73.37%					
	true van	true saab	true bus	true opel	class precision
pred. van	28	1	4	2	80.00%
pred. saab	5	23	1	5	67.65%
pred. bus	2	4	39	1	84.78%
pred. opel	5	15	0	34	62.96%
class recall	70.00%	53.49%	88.64%	80.95%	

Naivni Bayes

accuracy: 45.56%					
	true van	true saab	true bus	true opel	class precision
pred. van	35	14	22	11	42.68%
pred. saab	2	19	2	17	47.50%
pred. bus	2	1	9	0	75.00%
pred. opel	1	9	11	14	40.00%
class recall	87.50%	44.19%	20.45%	33.33%	

k-NN

accuracy: 68.05%					
	true van	true saab	true bus	true opel	class precision
pred. van	30	2	1	0	90.91%
pred. saab	3	22	0	21	47.83%
pred. bus	3	2	42	0	89.36%
pred. opel	4	17	1	21	48.84%
class recall	75.00%	51.16%	95.45%	50.00%	

Stablo Odluke

accuracy: 63.31%					
	true van	true saab	true bus	true opel	class precision
pred. van	28	4	1	5	73.68%
pred. saab	3	26	0	27	46.43%
pred. bus	6	2	43	0	84.31%
pred. opel	3	11	0	10	41.67%
class recall	70.00%	60.47%	97.73%	23.81%	

SPLICE

SVM

accuracy: 81.85%				
	true EI	true IE	true N	class precision
pred. EI	66	8	18	71.74%
pred. IE	6	63	11	78.75%
pred. N	4	6	110	91.67%
class recall	86.84%	81.82%	79.14%	

Naivni Bayes

accuracy: 89.38%				
	true EI	true IE	true N	class precision
pred. EI	64	3	6	87.67%
pred. IE	5	69	5	87.34%
pred. N	7	5	128	91.43%
class recall	84.21%	89.61%	92.09%	

k-NN

accuracy: 90.07%				
	true EI	true IE	true N	class precision
pred. EI	74	1	9	88.10%
pred. IE	2	74	15	81.32%
pred. N	0	2	115	98.29%
class recall	97.37%	96.10%	82.73%	

Stablo Odluke

accuracy: 89.38%				
	true EI	true IE	true N	class precision
pred. EI	64	3	6	87.67%
pred. IE	5	69	5	87.34%
pred. N	7	5	128	91.43%
class recall	84.21%	89.61%	92.09%	

Logistička Regresija

Nažalost, logistička regresija za splice problem se jako jako dugo izvršavala tako da nisam stigao dočekati rezultate..

c) za izračun vrijednosti napravi sam malu python skriptu, u zad3/tablica.py koja po formulama izračunava vrijednosti

STATLOG

klasifikator	Točnost	Mikro-F1	Makro-Preciznost	Makro-Odziv	Makro-F1
Log. Regresija	79.29%	0.79	0.79	0.79	0.79
SVM	73.37%	0.73	0.73	0.74	0.73
Naivni Bayes	45.56%	0.46	0.43	0.51	0.46
k-NN	68.05%	0.68	0.68	0.69	0.68
Stablo Odluke	63.31%	0.63	0.61	0.61	0.63

SPLICE

klasifikator	Točnost	Mikro-F1	Makro-Preciznost	Makro-Odziv	Makro-F1
Log. Regresija					
SVM	81.85%	0.89	0.81	0.81	0.83
Naivni Bayes	89.38%	0.89	0.89	0.89	0.89
k-NN	90.97%	0.90	0.90	0.89	0.92
Stablo Odluke	89.38%	0.89	0.89	0.89	0.89

d) za statlog problem najbolje rezultate daje logistička regresija a za splice problem svi osim svm-a daju dobre rezultate od kojih je za dlaku najbolji k-NN. Zanimljivo je da naivni bayes u statlog problemu daje dosta loše rezultate u usporedbi s ostalima. Iz toga se daje zaključiti da distribucija primjera znatno odstupa od gaussove, što i objašnjava zašto su rezultati statlog klasifikacije lošiji od splice-a, za koje se može zaključiti da imaju jasniju distribuciju i da su razredi više odvojeniji u prostoru značajki nego u statlog primjeru.

Rekao bih da mikro-f1 najbolje određuje pogrešku klasifikatora.

Postoji razlika između mikro-f1 i makro-f1 iako neznatna jer jedan gleda sve vrijednosti matrice zabune zbrojeno a drugi po klasama.

e) nažalost krivo sam procjenio potrebno vrijeme za izradu ove zadaće i jednostavno mi nije ostalo dovoljno vremena za riješiti ovaj zadatak, tako da ću ga riješiti za vježbu za završni ispit kada dođe vrijeme, a u kontekstu ove zadaće, svjesno prihvaćam manji broj bodova zbog toga.. Inače, zanimljiva informacija, za ovu zadaću trebalo mi je oko 16h sve skupa efektivnog vremena! Ne uključujući e) zadatak.. bome paprena zadaća nema što :)