

2. Domača naloga: Olfaction Prediction Subchallenge 2

Luka Krsnik (63110179)

29. marec 2015

1 Uvod

Ta naloga je razširitev prve domače naloge. Razlika je v tem, da je pri prvi nalogi napovedana le intenziteta vonja, tokrat pa pa smo poleg tega napovedali še prijetnost/neprijetnost molekul in 19 drugih atributov, ki so pridobljeni s strani tesnih subjektov.

2 Metode

LASSO LASSO (Least Absolute Shrinkage and Selection Operator) je regresijska metoda, ki se od ostalih tovrstnih metod razlikuje po tem, da poleg dejanske regresije (učenje funkcije napovedovanja na testnih podatkih) izvede tudi selekcijo parametrov. To konkretno pomeni da iz množice značilk izbere tiste, ki so najbolj uporabni in raznoliki. Zaradi te funkcionalnosti nam ni več potrebno uporabljati metod za izbor značilk (npr. PCA).

Cross validation Cross validation (K-means) je metoda, ki pomaga pri izboru parametrov, ki jih uporabljajo druge funkcije (npr. pri LASSO-ju α). Deluje tako, da so učni podatki razdeljeni na k delov (ponavadi 10), nato pa se metoda (LASSO) izvede k -krat. V vsaki iteraciji je eden izmed k -tih delov izpuščen iz učnih podatkov in uporabljen kot testni podatek, s pomočjo katerega izračunamo napako naših napovedi in povprečje vseh napak.

3 Rezultati

Zaradi omejenega števila poskusov na strežniku (4 na dan), so rezultati razdeljeni na dve tabeli. V prvi tabeli so zapisani rezultati metode Cross validation (k je bil enak 10), za različne α (Tabela: 1). Za najboljšo α se je med testiranimi izkazala 0.8, zato sem se odločil, da na končni oddaji oddam takšno α .

V drugi tabeli (Tabela: 2) so zapisane oddaje na strežnik. Tokrat je zadna oddaja ta, ki se je izkazala za najboljšo po cross validation metodi. To pa zato, da podatki niso preveč overfittani. To je oddaja, z α 0.8.

Tabela 1: Tabela z rezultati metode cross validation.

alpha	ocena - cross validation
100	1.70
50	2.14
10	3.91
5	4.14
2	4.33
1.5	4.34
1	4.25
0.8	4.50
0.6	4.45
0.5	4.27
0.1	3.24
0.01	3.09
0.001	2.92

Tabela 2: Tabela rezultatov iz strežnika. (zadnja oddaja je oddaja z alphi 0.8)

ime metode	alpha	oddaja
regresija, regularizacija, pearson	/	2.64592
LASSO	8	4.06052
LASSO	0.8	3.93837

4 Izjava o izdelavi domače naloge

Domačo nalogo in pripadajoče programe sem izdelal sam.