

Peta domača naloga

Anže Pečar (63060257)

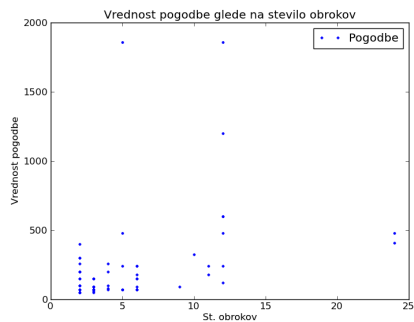
12. april 2012

1 Uvod

Cilj domače naloge je bil seznaniti se z linearno regresijo.

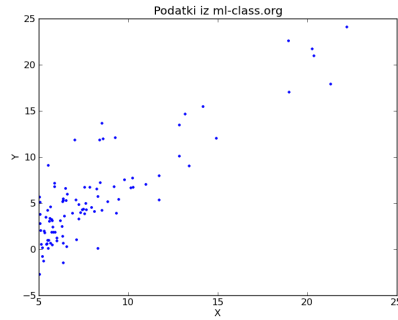
2 Podatki

Za podatke sem želel uporabiti bazo podatkov o pogodbah podjetja pri katerem občasno delam. Pridobil sem si dovoljenje za uporabo teh podatkov, vendar so se na žalost izkazali za neprimerne, saj nobeni atributi niso bili medsebojno odvisno. Še najboljše je kazalo atributu št. obrokov in razredom znesek pogodbe, vendar kot je razvidno iz slike 2 ni nobene lepe povezanosti.



Slika 1: Znesek pogodbe v odvisnosti od št. obrokov

Zato sem na koncu uporabil podatke iz online ml tečaja, prikaz podatkov je na sliki 2.



Slika 2: Podatki iz ml-class.org

3 Rezultati

3.1 Prva točka

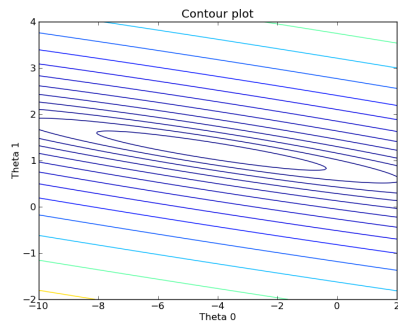
Formula za kriterijsko funkcijo je

$$J(\Theta) = J(\Theta_0, \Theta_1) = 0.5 * n \sum (\Theta_0 + \Theta_1 * x)^2$$

in sem jo v Python kodo prevedel kot

```
0.5/len(x)*sum([(x1*theta1+theta0-y[i])**2 for i,x1 in enumerate(x)])
```

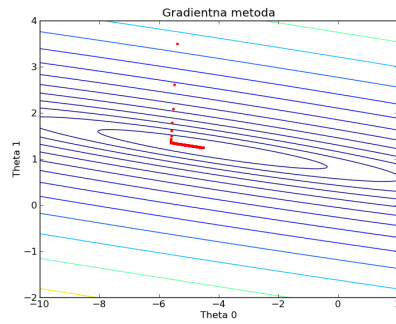
Iz dobljenih rezultatov sem nato izrisal contour plot. Rezultat je viden na sliki 3.1.



Slika 3: Prikaz J funkcije

3.2 Druga točka

Prikaz poteka batch metode je na sliki 3.2.



Slika 4: Prikaz paketne metode

3.3 Tretja točka

4 Izjava o izdelavi domače naloge

Domačo nalogo in pripadajoče programe sem izdelal sam.

Literatura

- [1] Ian H. Witten & Eibe Frank, *Data Mining Practical Machine Learning Tools and Techniques, Second Edition* Morgan Kaufmann Publishers, 2005.