

Peta domača naloga

Anže Pečar (63060257)

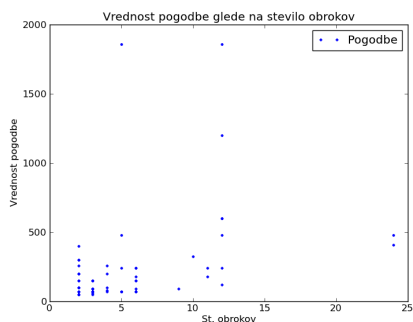
13. april 2012

1 Uvod

Cilj domače naloge je bil seznaniti se z linearno regresijo.

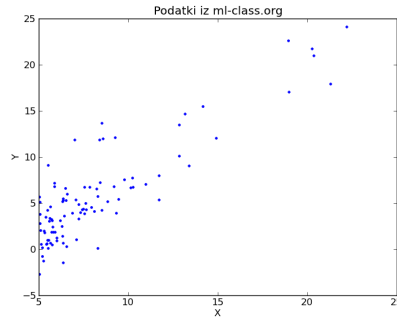
2 Podatki

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

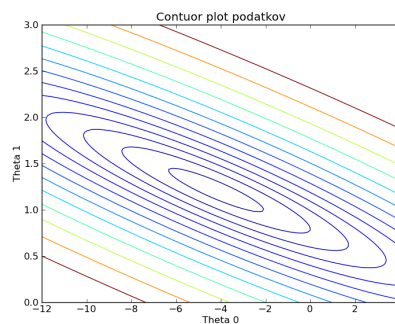
3 Rezultati

3.1 Prva točka

Formula za kriterijsko funkcijo je

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

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



Slika 3: Prikaz J funkcije

3.2 Druga točka

S paketno, stohastično in analitično metodo sem izračunal različne Thete, ki sem jih zaokrožil na 5 mest, saj sem imel nastavljen epsilon na 0.00001.

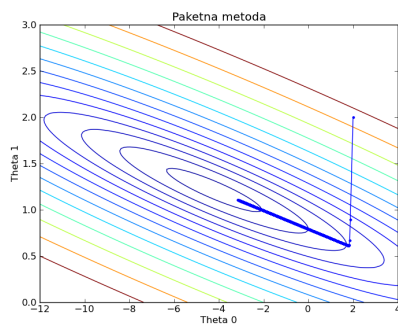
batch [-3.1406, 1.10629]

sto [-3.1175, 1.07008]

ana [-4.2115, 1.21354]

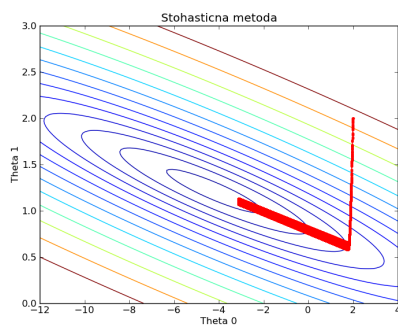
Na slikah 3.2 in 3.2 vidimo, da obe metodi dokaj hitro skonvergirata do določenega nivoja, nato pa se zelo počasi začneta približevati središču. Razlog za to je najbrž v majhnih razlikah okoli

globalnega maksimuma. Prikaz poteka batch metode je na sliki 3.2.



Slika 4: Prikaz paketne metode

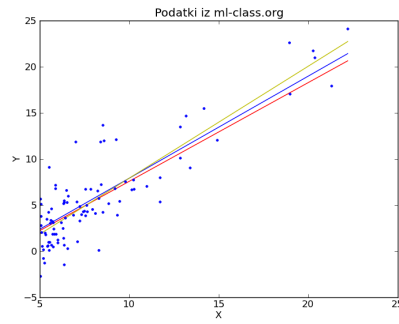
Prikaz poteka stohastične metode je na sliki 3.2.



Slika 5: Prikaz stohastične metode

3.3 Tretja točka

Slika 3.3 končne rezultate vseh uporabljenih funkcij. Rdeča črta prikazuje model zgrajen iz stohastične metode, modra iz paketne metode in rumena iz analitične metode. Kot je lepo vidno iz slike obe iterativni metodi nekoliko odstopata od analitične rešitve. Če bi povečali število iteracij, ki jih imata metodi na voljo, bi se razlike nekoliko zmanjšale, ampak samo do neke mere, saj je področje globalnega maksimuma dokaj ravno, kar nekoliko zmede iterativna algoritma.



Slika 6: Prikaz premic

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.