

# 1. domača naloga: Linearna regresija

Jernej Henigman (63100259)

5. marec 2016

## 1 Uvod

V 1. domači nalogi je bilo potrebno implementirati manjkajoče dele funkcij v povezavi z linearno regresijo. Cilj naloge je bila implementacija gradientnega spusta, s katerim pridobimo parametre linearne regresije ( $\theta$ ). Potrebno je bilo primerjati parametre gradientnega spusta z rezultati, ki jih izračunamo analitično in temi, ki jih vrne optimizacija L-BFGS. Potrebno je bilo izračunati tudi gradient s pomočjo končnih diferenc in ga primerjati z gradientom, ki smo ga izračunali analitično. Na koncu izrišemo točke v ravnini (učni primeri) in jih prikažemo skupaj z rezultati linearne regresije (premica v grafu).

## 2 Podatki

Blood and Breath Alcohol Elimination Rates among Austrian Males and Females. Breath alcohol elimination rates (mg per litre per hour) and Blood alcohol elimination rates (g per litre per hour) measured on a sample of 59 Austrians (32 M, 27 F).

Opomba: podatke sem pridobil iz spletne strani <http://www.stat.ufl.edu/winner/datasets.html>  
Opis podatkov nisem prevajal v slovenščino.

## 3 Metode

- **Gradientni sestop**

Potrebno je bilo minimizirati kriterijsko funkcijo  $J(\theta) = \frac{1}{M} \sum_{i=1}^M (y^{(i)} - (\theta_0 + \theta_1 X))^2$ . Izračunali smo parcialna odvoda kriterijske funkcije in s pomočjo teh odvodov v vsakem obhodu zanke obnovili vrednost parametrov  $\theta_0$  in  $\theta_1$ , ki skonvergirata v lokalni minimum.

- **Optimizacija L-BFGS**

Vgrajena metoda L-BFGS iz knjižnice scipy. Funkcija L-BFGS mnogo hitreje izračuna vrednost parametrov  $\theta_0$  in  $\theta_1$ .

- **Postopek končnih diferenc**

S postopkom končnih diferenc preverimo če je analitično izračunan gradient pravilen in obratno.

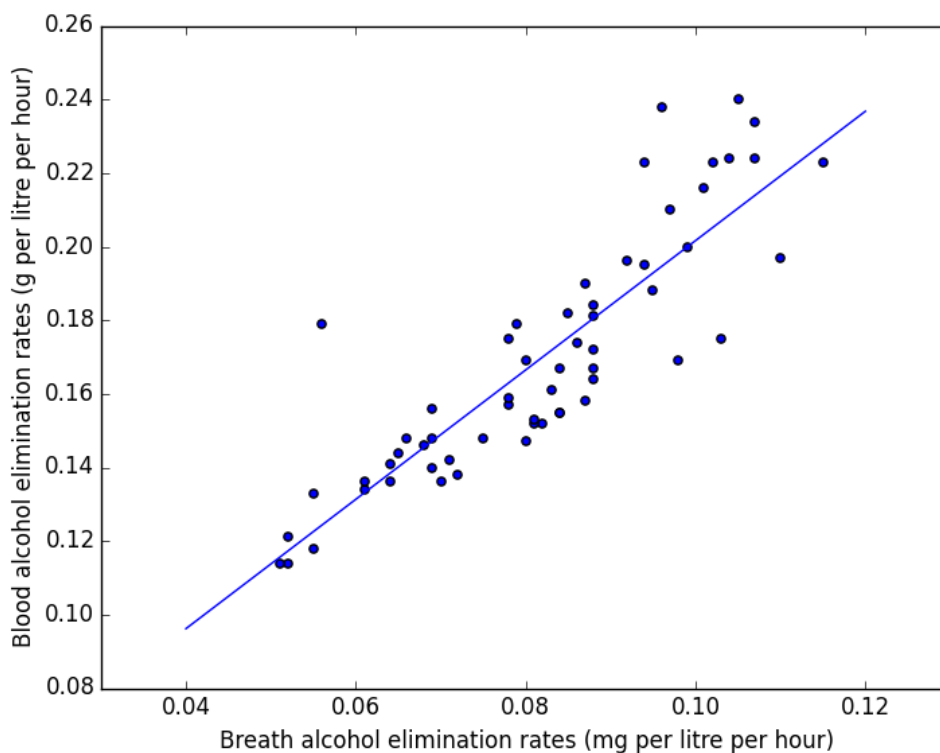
- Ostalo

Dodajane stolpca enic k vektorju  $X$ . Funkcija  $J$ , odvod funkcije  $J$  kot  $dJ$ . Risanje razsevnega grafa. Branje podatkov.

## 4 Rezultati

	$\theta_0$	$\theta_1$	Št. iteracij	Čas računanja
Gradienti spust	0.025872	1.757632	100000	11s.
L-BFGS	0.025871	1.757640	13	0.003s
Analitična rešitev	0.025871	1.757636	/	0.12s

Vidimo, da dobimo praktično enake vrednosti  $\theta_0$  in  $\theta_1$  v treh različnih metodah. Optimizacija L-BFGS je mnogo učinkovitejša od gradientnega sestopa, kar se vidi v času računanja iz zgornje tabele. L-BFGS je tudi najučinkovitejša metoda izmed treh.



Slika 1: Razsevni diagram. Učni podatki kot modre pike in premica v grafu kot rezultat linearne regresije.

Rezultati iz slike lahko rečemo da so intuitivni. Atribut (breath alcohol elimination rate) in razred (blood alcohol elimination rate) sta visoko kolerirana. Višja kot je vrednost atributa, višja bo vrednost razreda. Zaradi narave podatkov, je linearna regresija dober model, če bi želeli napovedovati vrednost razreda na novih podatkih.