

# Leren: Opdracht 2

Semester 1 2014/15

Deadline: 6 November 2014

**MAG IN TEAMS VAN MAXIMAAL 2 GEMAAKT WORDEN**

NB: Als je deze opgave in een team van 2 maakt, lever hem dan 1 keer en maar vermeld duidelijk wie de auteurs zijn.

1. *Multipele Lineaire Regressie met de hand*

Doe 1 iteratie van gradient descent voor multipele regressie met onderstaande data. Gebruik een learning rate van 0.1 en initialiseer  $\theta_0$ ,  $\theta_1$  en  $\theta_2$  allemaal met 0.5. Bereken de cost voor en na deze iteratie.

<b>X1</b>	2	4	4
<b>X2</b>	3	5	3
<b>Y</b>	6	6	10

2. *Terminatie*

Het gradient descent algorithm heeft een terminatie conditie nodig. We kunnen dit opgeven als (a) het aantal iteraties, of als (b) de maximale grootte van de laatste verandering in parameters of (c) van de laatste verandering in cost. Geef voor en nadelen van deze verschillende mogelijkheden.

3. *Logistische Regressie met de hand*

Doe 1 iteratie van gradient descent voor logistische regressie met onderstaande data. Gebruik een learning rate van 0.1 en initialiseer  $\theta_0$ ,  $\theta_1$  en  $\theta_2$  allemaal met 0.5. Bereken de cost voor en na deze iteratie. Plot de boundary ( $h(x1, x2) = 0.5$ ) voor en na de iteratie.

<b>X1</b>	5	5	3	2
<b>X2</b>	3	5	3	4
<b>Y</b>	0	0	1	1

<b>X1</b>	5	5	3	2
<b>X2</b>	3	5	3	4
<b>Y</b>	0	0	1	1

4. *Cost Logistische Regressie*

Twee maten voor de "cost" van een classifier zijn accuracy, ofwel de proportie goede klassificaties (als we het over kost hebben misschien beter: 1- accuracy) en de cost function die Andrew gebruikt. Voor wat voor data maakt het verschil welke functie gebruikt wordt? Kan je een (kleine) dataset bedenken waarin de class boundary verschillend is voor deze twee maten voor de cost?