**Leren homework # 2**

**Date:** November 6, 2014

**Name (Student Number):** M. Pfundstein (10452397), T.M. Meijers (10647023)

# Question 1

Hypothesis function:
$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)}$

Cost function:
$J(\theta) = \frac{1}{2m}\sum\limits_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$

Update function:
$\theta_i' = \theta_i - \alpha\frac{\delta}{\delta\theta_i}J(\theta)$

Cost after 0 iterations:
With $\theta_0 = 0.5, \theta_1 = 0.5, \theta_2 = 0.5$

$J(\theta) = \frac{1}{6}((0.5 + 1 + 1.5 - 6)^2 + (0.5 + 2 + 2.5 - 6)^2 + (0.5 + 2 + 1.5 - 10)^2) =$
$\frac{1}{6}((-3)^2 + (-1)^2 + (-6)^2) =$
$\frac{46}{6} = \frac{23}{3}$

Update parameters (1 iteration):
$\theta_0' = 0.5 - \frac{0.1}{3}(-3 - 1 - 6) =$
$0.5 - \frac{-1}{3} = \frac{5}{6}$

$\theta_1' = 0.5 - \frac{0.1}{3}(2 \cdot -3 + 4 \cdot -1 + 4 \cdot -6) =$
$0.5 - \frac{0.1}{3}(-34) =$
$0.5 + \frac{3.4}{3} = \frac{49}{30}$

$\theta_1' = 0.5 - \frac{0.1}{3}(3 \cdot -3 + 5 \cdot -1 + 3 \cdot -6) =$
$0.5 - \frac{0.1}{3}(-32) =$
$0.5 + \frac{3.2}{3} = \frac{47}{30}$

Cost after 1 iteration:
With $\theta_0 = \frac{5}{6}, \theta_1 = \frac{49}{30}, \theta_2 = \frac{47}{30}$

$J(\theta) = \frac{1}{6}((\frac{5}{6} + \frac{49}{30} \cdot 2 + \frac{47}{30} \cdot 3 - 6)^2 + (\frac{5}{6} + \frac{49}{30} \cdot 4 + \frac{47}{30} \cdot 5 - 6)^2 + (\frac{5}{6} + \frac{49}{30} \cdot 4 + \frac{47}{30} \cdot 3 - 10)^2) \approx$
$16.12518$

## Question 2

Positive and negative aspects per terminating condition (gradient descent algorithm):

(a) Iteration:
+ One can directly control the running time of the algorithm.
+ The algorithm will terminate guaranteed.
- If the number of iterations is too high, the algorithm will do redundant work.
- If the number of iterations is too low, you can obtain parameters far form optimal.

(b) Change in parameters:
+ A very small change can be chosen which guarantees nearly optimal parameters.
+ No need to calculate cost in between updating.
- If chosen learning rate is too high, algorithm will never terminate.
- Change in parameters does not nessecarily mean a change in cost.

(c) Change in cost:
+ A very small change can be chosen which guarantees nearly optimal parameters.
+ Cost is the result one wants to minimize.
- Algorithm also needs to calculate cost, opposed to using latest change in parameters.

## Question 3

Hypothesis function:
$h_\theta(x^{(i)}) = \frac{1}{1+e^{-\theta^T x^{(i)}}}$

Cost function:
$J(\theta) = \frac{1}{m}\sum\limits_{i=1}^{m} \begin{cases} -log(h_\theta(x^{(i)}) & \text{if } y^{(i)} = 1, \\ -log(1 - h_\theta(x^{(i)})) & \text{if } y^{(i)} = 0 \end{cases}$

Simplified cost function:
$J(\theta) = -\frac{1}{m}\sum\limits_{i=1}^{m} y^{(i)} log(h_\theta(x^{(i)})) + (1-y^{(i)}) log(1 - h_\theta(x^{(i)}))$

Update function (using simplified $J(\theta)$):
$\theta'_j = \theta_j - \alpha\frac{1}{m}\sum\limits_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$

Calculate hypothethes:
$h_\theta(x^{(1)}) = \frac{1}{1+e^{-9/2}} \approx 0.9890$
$h_\theta(x^{(2)}) = \frac{1}{1+e^{-11/2}} \approx 0.9959$
$h_\theta(x^{(3)}) = \frac{1}{1+e^{-7/2}} \approx 0.9707$
$h_\theta(x^{(4)}) = \frac{1}{1+e^{-7/2}} \approx 0.9707$

Cost after 0 iterations (Normal cost function):
With $\theta_0 = 0.5, \theta_0 = 0.5, \theta_0 = 0.5$

$J(\theta) = -\frac{1}{4}(log(1-0.9890) + log(1-0.9959) + log(0.9707) + log(0.9707)) \approx 2.5187$

Update parameters (1 iteration):
$\theta'_0 = 0.5 - 0.1\frac{1}{4}\sum\limits_{i=1}^{4}(h_\theta(x^{(i)}) - y^{(i)})x_0^{(i)}$
$= 0.5 - \frac{0.1}{4}[([(0.9890-0)*1 + (0.9959-0)*1 + (0.9707-1)*1 + (0.9707-1)*1])] \approx 0.4518$
$\theta'_1 = 0.5 - 0.1\frac{1}{4}\sum\limits_{i=1}^{4}(h_\theta(x^{(i)}) - y^{(i)})x_1^{(i)}$
$= 0.5 - \frac{0.1}{4}[([(0.9890-0)*5 + (0.9959-0)*5 + (0.9707-1)*3 + (0.9707-1)*2])] \approx 0.2555$
$\theta'_2 = 0.5 - 0.1\frac{1}{4}\sum\limits_{i=1}^{4}(h_\theta(x^{(i)}) - y^{(i)})x_2^{(i)}$
$= 0.5 - \frac{0.1}{4}[([(0.9890-0)*3 + (0.9959-0)*5 + (0.9707-1)*3 + (0.9707-1)*4])] \approx 0.3065$
Cost after 1 iteration (Normal cost function):
With $\theta_0^{'} = 0.4518, \theta_0^{'} = 0.2555, \theta_0^{'} = 0.3065$

Calculate hypothethes:
$h_{\theta'}(x^{(1)}) = \frac{1}{1+e^{-9/2}} \approx 0.9339$
$h_{\theta'}(x^{(2)}) = \frac{1}{1+e^{-11/2}} \approx 0.9631$
$h_{\theta'}(x^{(3)}) = \frac{1}{1+e^{-7/2}} \approx 0.8945$
$h_{\theta'}(x^{(4)}) = \frac{1}{1+e^{-7/2}} \approx 0.8992$

$J(\theta') = -\frac{1}{4}(log(1-0.9339)+log(1-0.9631)+log(0.8945)+log(0.8992)) \approx 1.5586$

## Plotting

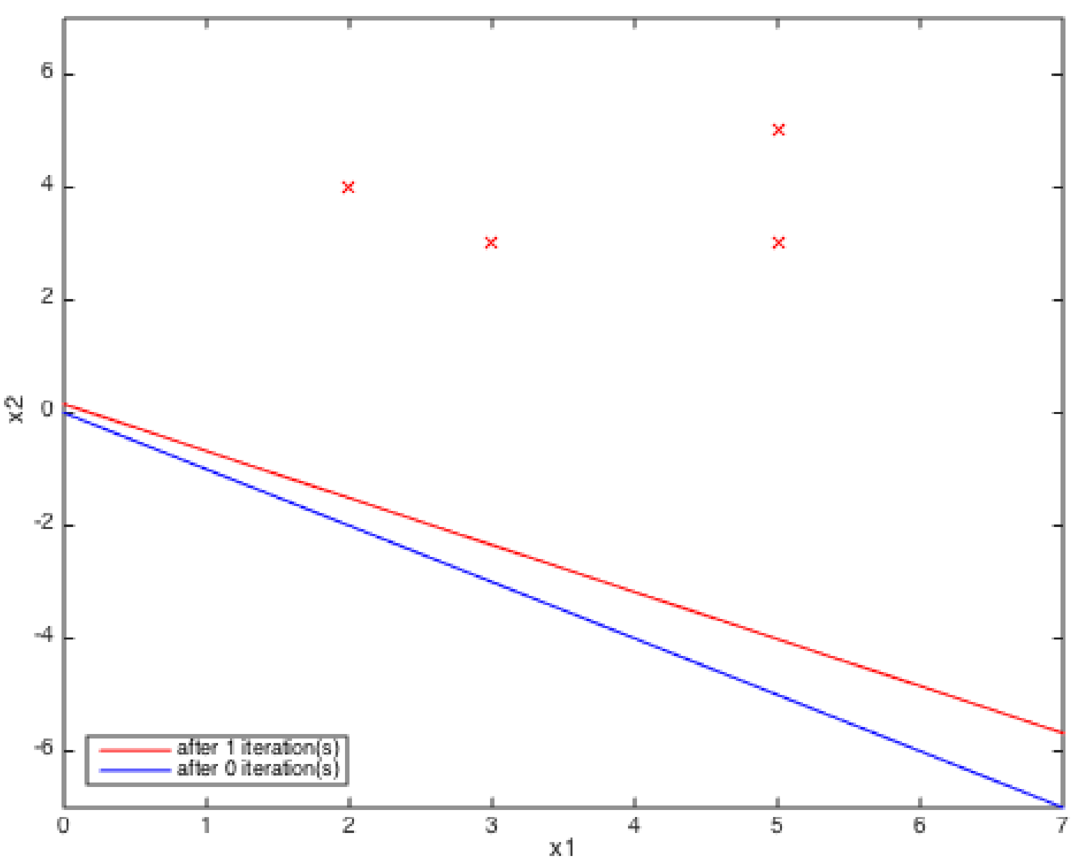
$\theta_0 = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}$

$h_{\theta_0}(x_1, x_2) = 0.5 = 0.5 + 0.5x_1 + 0.5x_2$
$x_2 = -x_1$

$\theta_1 = \begin{bmatrix} 0.4518 \\ 0.2555 \\ 0.3065 \end{bmatrix}$

$h_{\theta_1}(x_1, x_2) = 0.5 \rightarrow 0.5 = 0.4518 + 0.2555x_1 + 0.3065x_2$
$\rightarrow x_2 = 0.157259 - 0.833605x_1$

Figure 1: Class boundaries plotted for both sets of Theta's



## Question 4

The cost method (accuracy or Andrew's method) used makes a different when one has a dataset where all the data points are very apart, i.e. they're all far away from the class boundary. When one has such a dataset the accuracymethod will report 100% accuracy whereas Andrew's method will return a very large cost.

Figure 2: A small dataset with significant difference between cost methods