Leren homework # 3

Date: November 11, 2014

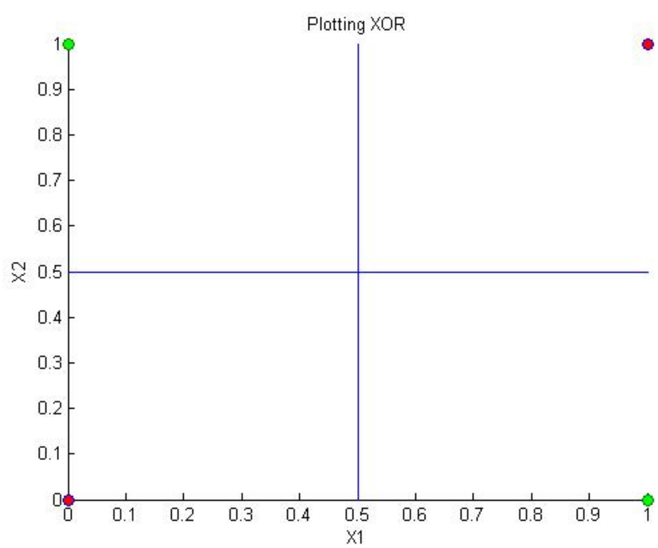
Name (Student Number): T.M. Meijers (10647023)

Question 1

(a) When one uses regularization the parameters (θ's) will always decrease since regularization will penalize through the value of theta's (with factor λ).

(b) When one has a very large dataset, overfitting is nearly impossible, this means that the optimal λ will be a lot lower which means that the optimal θ's can be relatively higher.

Question 2



As can be seen above, the green points are the ones that are class 1, where one of the variables ($X_1 or X_2$) is 1 while the other is 0. When both are 0 or 1 you obtain a 0, represented by the two red points. (Maybe this explanation is redundant, but I included it for completeness).
When we want an accurate boundary, we would want to have two quadrants (the ones which contain the green points), where evyerthing inside this quadrant is a 1. The remaining two quadrants (with the red points inside) should be a 0. So for this one would need a complex function. This cannot be done by logistic regression with a class boundray, this would create a single line (probably one of the two plotted lines).

Question 3

$$\begin{tikzcd} X_1 \arrow[r, "\theta_1^{(1)} = 0.2"'] & Z_1 \arrow[r, "\theta_1^{(2)} = 0.1"'] & Y_1 \end{tikzcd}$$

(a) $Y_1 = g(\theta_1^{(2)} g(\theta_1^{(1)} X_1)) = g(0.1 \cdot g(0.2 \cdot -5)) \approx 0.5067$
Where $g(z)$ is the sigmoid function.

(b) Again, using the sigmoid function ($g(z)$) to explain my answer:
As the activation of Y_1 that was found was lower then it was supposed to be (0.8 as opposed to 1.0). The sigmoid function's outcome will be lower, as z will be higher. This due to dividing by the natural number to the power of z. Thus if our answer should be higher, our parameters should decrease which in turn makes z lower.

(c) For calculation of Y (0.5067) see 3a.
$\delta^{(3)} = 0.5067 - 1 = -0.4933$
$\delta^{(2)} = 0.1 \cdot -0.4933 \cdot (\frac{1}{1+e^{-(0.2 \cdot -5)}} \cdot (1 - \frac{1}{1+e^{-(0.2 \cdot -5)}})) = -0.0097$
$\theta_1^{(2)'} = \theta_1^{(2)} - \alpha \cdot \delta^{(3)} = 0.1 - 0.1 \cdot -0.4933 = 0.1493$
$\theta_1^{(1)'} = \theta_1^{(1)} - \alpha \cdot \delta^{(2)} = 0.2 - 0.1 \cdot -0.0097 = 0.2010$