Q2

Let N be the number of observations and $d+1$ be the number of features in each observation (including the bias). Then the cost-function for simple logistic regression is given by

$$J(w) = \frac{1}{N}\sum_{i=1}^{N}\left(-y_i \log(h(x_i)) - (1-y_i)\cdot\log(1-h(x_i))\right)$$

where $h(x_i)$ is the sigmoid function

Adding the regularization term to this cost

$$J(w) = \frac{1}{N}\sum_{i=1}^{N}\left(-y_i \log(h(x_i)) - (1-y_i)\cdot\log(1-h(x_i))\right) + \frac{\lambda}{2N}\sum_{j=1}^{d} w_j^2$$

Differentiating this cost w.r.t w_j we get for $j>0$

$$\frac{\partial J(w)}{\partial w_j} = \frac{1}{N}\sum_{i=1}^{N}(h(x_i) - y_i)\cdot x_{ji} + \frac{\partial}{\partial w_j}\left(\frac{\lambda}{2N}\sum_{k=1}^{d} w_k^2\right)$$

$$= \frac{1}{N}\left(\sum_{i=1}^{N}(h(x_i)-y_i)\cdot x_{ji}\right) + \frac{\lambda w_j}{N}$$

for w_0 the gradient would be the same as simple logistic regression ie

$$\frac{\partial J(w)}{\partial w_0} = \frac{1}{N} \sum_{i=1}^{N} \left(h(x_i) - y_i \right) x_{0i}$$

So weight update equation would be

$$w_j = w_j - \eta \frac{\partial J(w)}{\partial w_j}$$

Q4 Given

$w_0 = -8, \ w_1 = 0.05, \ w_2 = 1$

a) Probability that a ~~standard~~ student who studies for 5h & has GPA 7.5 gets an A

$$= \frac{1}{1 + e^{-(-8 + 0.05 \times 5 + 7.5)}}$$

$$= \frac{1}{1 + e^{-(-0.25)}}$$

$$= \frac{1}{1 + e^{0.25}}$$

$$= 0.4378$$

Scanned by CamScanner

b) Let

$$\frac{1}{1+e^{-z}} = 0.6$$

$$\Rightarrow \frac{2}{3} = e^{-z}$$

$$\Rightarrow z = \log(2/3)$$

$$\Rightarrow z = 0.4055$$

Let x be the number of hours the student should study, then

$$-8 + 0.05x + 7.5 \times 1 = 0.4055$$

$$\Rightarrow x = 19.1$$

So the number of hour the student should study is 19.1.