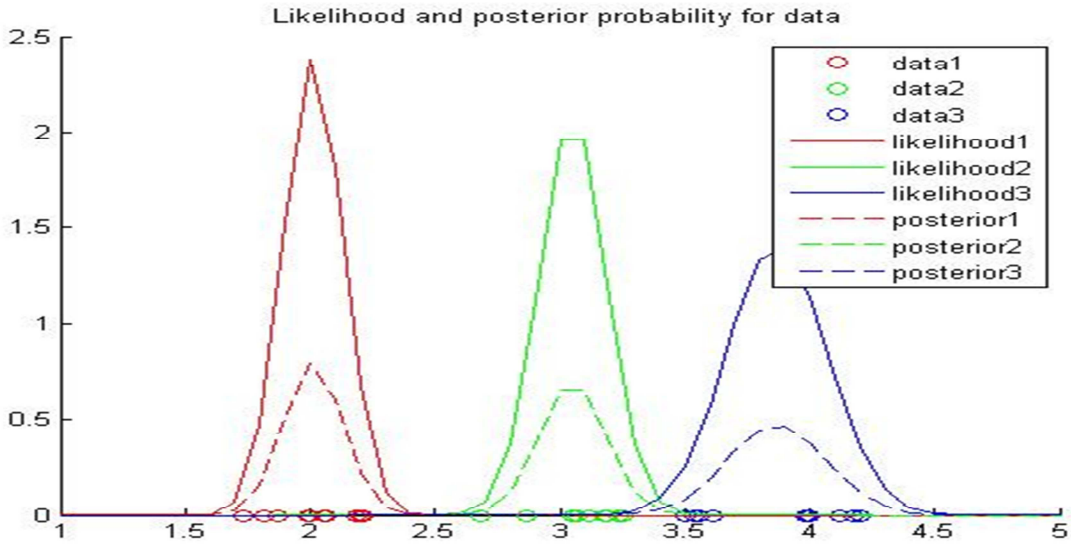
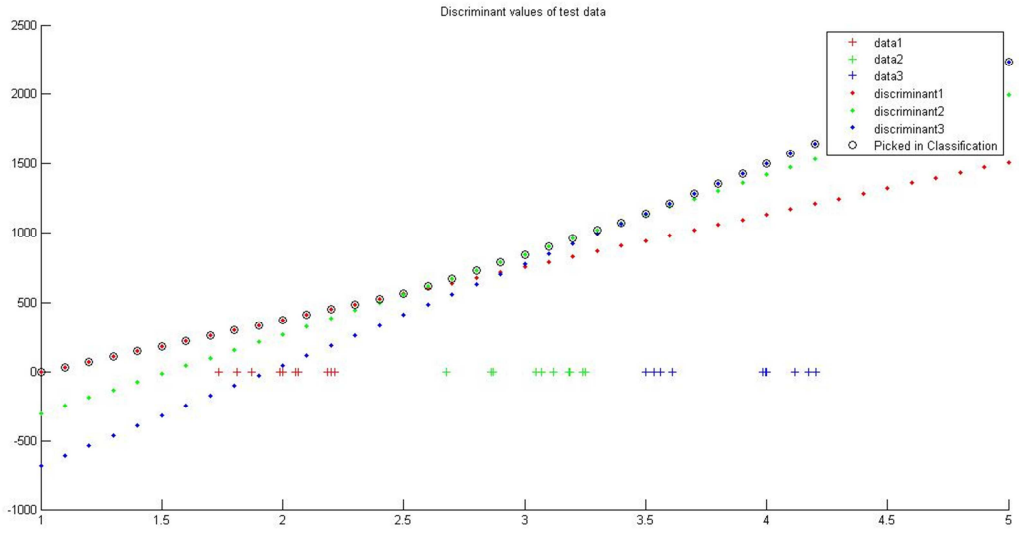
Report 2011CS1011

CSL 407 Machine Learning Homework 2

Question 1 :

The plots found in the question are :





Question 2 :

2. The cost function without regularization is :

$$J(w) = \frac{1}{N} \sum_{n=1}^{N} (-y_n \log(g(x_n)) - (1-y_n) \log(1-g(x_n))) \quad \text{--(i)}$$

where w is weight vector, N is number of x, $g(x_n) = \frac{1}{1+e^{-w^T x}}$

We know that:

$$\frac{\partial J(w)}{\partial w} = \frac{1}{N} \sum_{n=1}^{N} (g(x_n) - y_n) x_{d_n} \quad \text{--(i)}$$

After adding regularisation, the cost function becomes :

$$J(w) = \frac{1}{N} \sum_{n=1}^{N} (-y_n \log(g(x_n)) - (1-y_n) \log(1-g(x_n))) + \frac{\lambda}{2N} \|w\|^2 \quad \left\{ \|w\|^2 = \sum_{j=1}^{m} w_j^2 \right\}$$

So when we differentiate :-

for $w_j = 0$ $\frac{\partial J(w)}{\partial w_j}$ is same as (i)

But for $j \neq 0$; we add differentiation of regularisation term :-

$$\frac{\partial J(w)}{\partial w_j} = \frac{1}{N} \sum_{n=1}^{N} (g(x_n) - y_n) x_{d_n} + \frac{\partial}{\partial w_j} \left(\frac{\lambda}{2N} \sum_{j=1}^{m} w_j^2 \right)$$

$$= \frac{1}{N} \sum_{n=1}^{N} (g(x_n) - y_n) x_{d_n} + \left(\frac{\lambda}{2N} \cdot 2w = \frac{\lambda w_j}{N} \right)$$

Therefore the weight update equation :

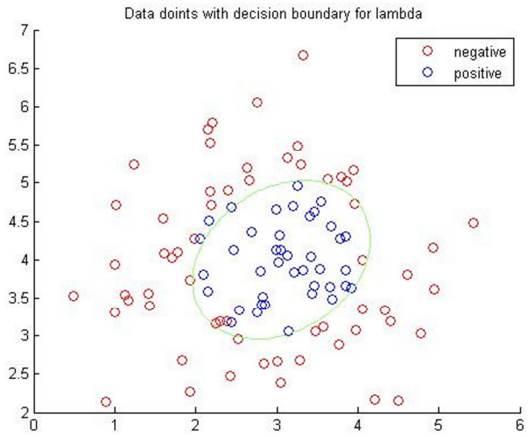
$$w_j = \begin{cases} w_j - \alpha \frac{1}{N} \left(\sum_{n=1}^{N} (g(x_n) - y_n) x_{d_n} \right) & \text{for } j = 0 \\ = w_j - \alpha \frac{1}{N} \left(\sum_{n=1}^{N} (g(x_n) - y_n) x_{d_n} + \lambda w_j \right) & \text{for } 1 \leq j \leq m \end{cases}$$

Question 3 :

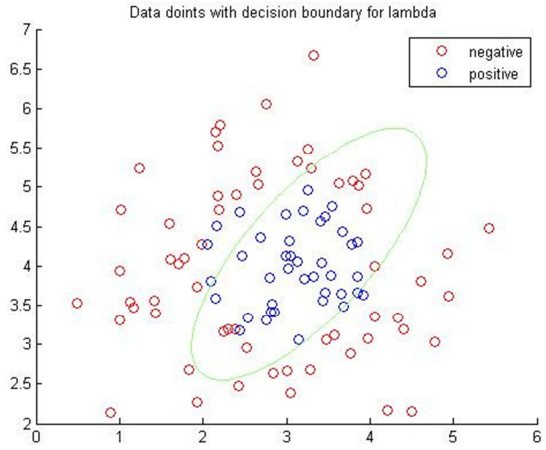
Observation : Graph for varying degree and lambda .

Varying lambda for degree = 2

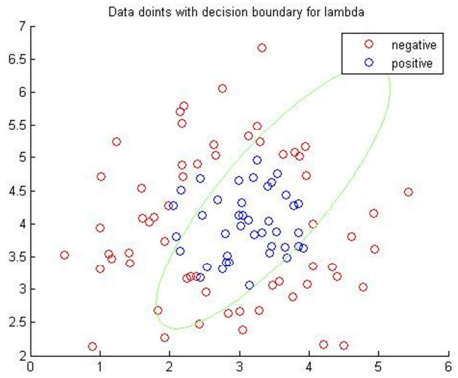
Lambda = 0.0



Lambda = 0.5

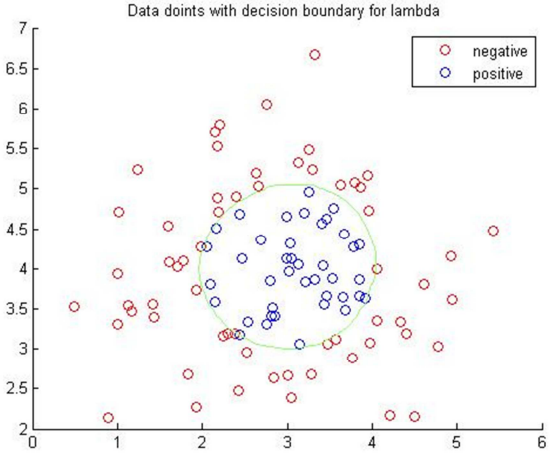


Lambda = 1.0

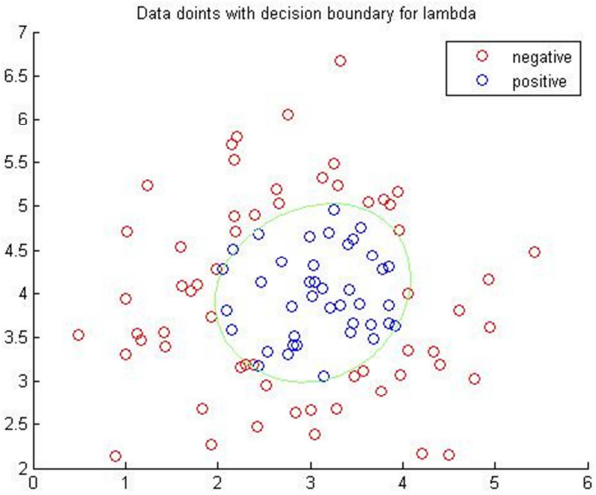


For Degree = 3

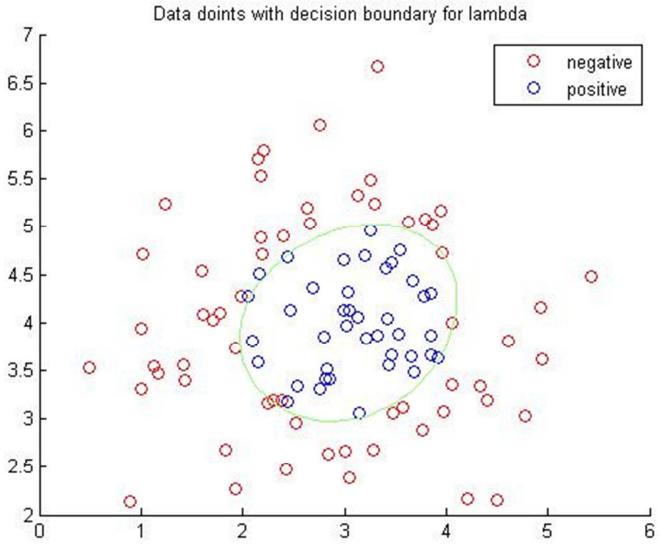
Lambda = 0.0



Lambda = 0.5

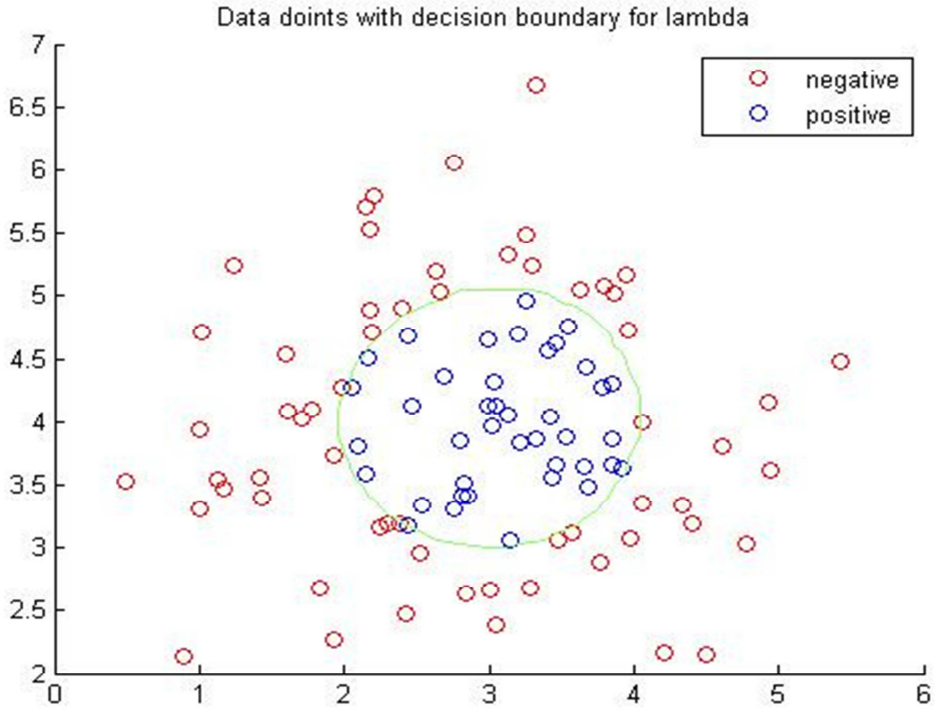


Lambda = 1.0



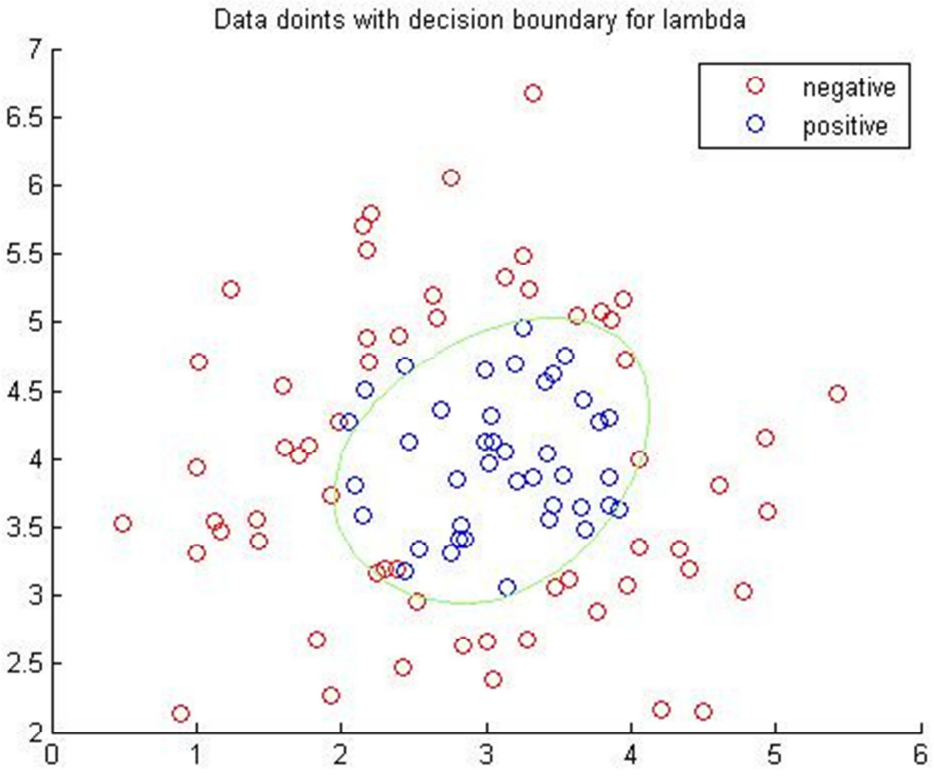
(ii) Overfitting happens in the case where value of lambda is low .Example given Lambda = 0.0 Degree 4

We can notice that all the points are separated by decision boundary.



Under fitting happens when there is extra regularization . Example given Lambda = 100.0 Degree = 4.0

We can notice that some significant amount of points are missed by decision boundary.



Question 4 :

Solution : x_1 = hours studied, x_2 = undergrad GPA, and Y = receive an A

Estimated coefficient , $w_0 = -8$, $w_1 = 0.05$, $w_2 = 1$

(a) $x_1 = 5$ $x = 7.5$

$w^T x = -8 + (0.05*5) + (1*7.5) = -0.25$

$g(w^T x) = g(-0.25) = 0.4378.$

(b) Probability of student getting A in class = 0.6

Therefore $g(w^T x) = 0.6$

So, $w^T x = 0.4054$

$-8 + (0.05*x_1) + (1*7.5) = 0.4054$

$x_1 = 18.108$

Student need to study for 18.108 hours .