Mohit Garg

2012CSB1020

CSL407 Machine Learning

Homework 1

Q 1. 1-dimensional data that facilitates intuitive visualization of linear discriminant analysis.

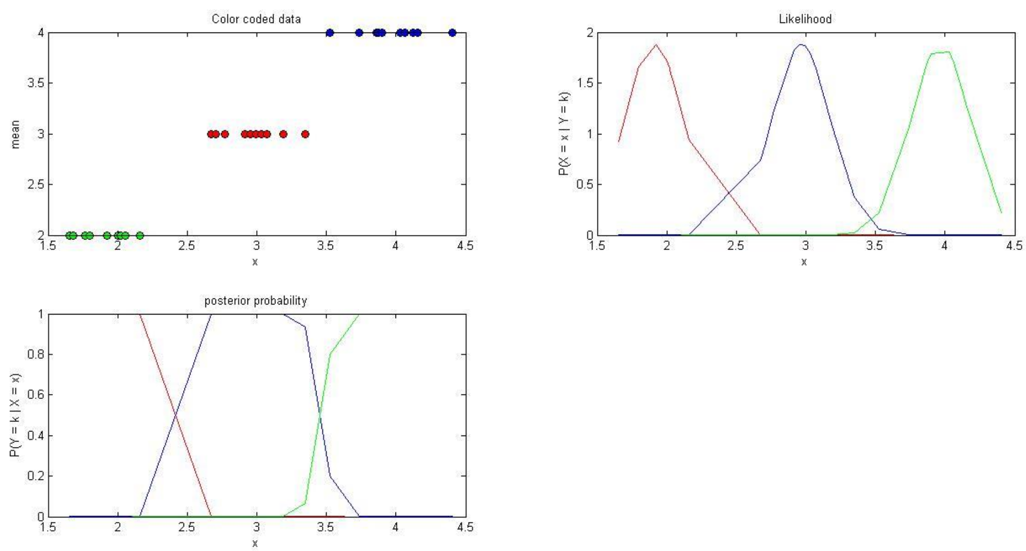
Figure 1 :

Observation :

Color coded data is dense near the mean value .

Likelihood of the the x values are maximum at the mean of the data , and is nearly zero for far away points, it decreases slowly from one to zero.

Posterior probability is also maximum at mean of the data , and is nearly zero for far away points , but it decreases very fast from one to zero.



First Subplot correspond to the plotting of the color coded data .

Second Subplot Corresponds to the Likelihood of the training data.

Third Subplot Corresponds to the posterior probability of the training data.

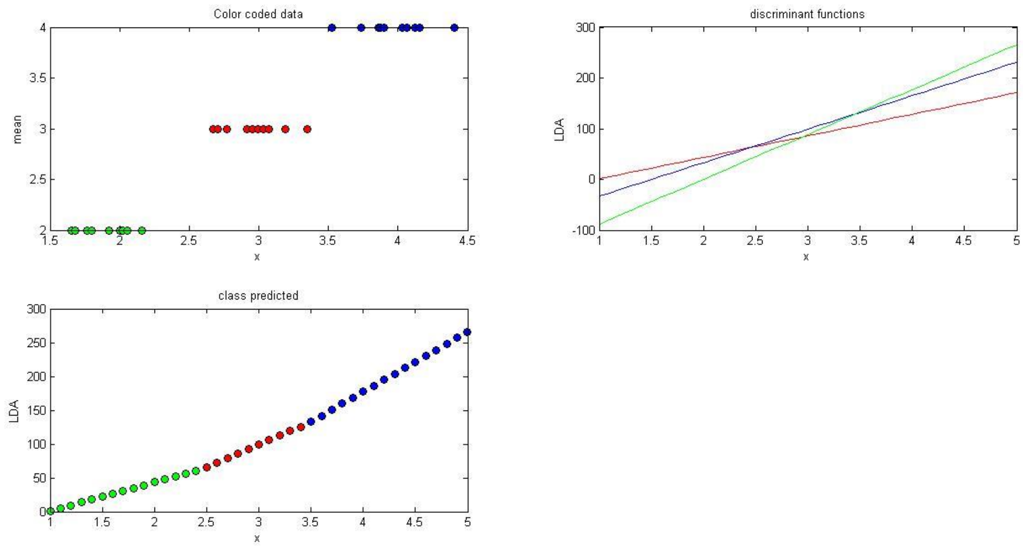
Figure 2:

Observations :

Color coded data is dense near the mean value

Discriminant function is monotonically increasing for all the classes.

Discriminant functions for a point x is maximum for the class that it belongs to.



First Subplot Corresponds to the plotting of the color coded data.

Second Subplot Corresponds to the plotting of the three discriminant functions for the test data

Third Subplot Corresponds to the class predicted by the linear discriminants for the test dataset.

Q2

Since, Logistic Regression is also prone to overfitting, hence a regularization term is added.

Modified cost function,

$$J(\omega) = \frac{1}{2N} \sum_{n=1}^{N} \left(g(x_n) - y_n \right)^2 + \frac{\lambda}{2N} \|\omega\|^2$$

where $g(x) = \dfrac{1}{1 + e^{-\omega^T x}}$

$$\frac{\partial J(\omega)}{\partial \omega_i} = \frac{1}{2N} \sum_{n=1}^{N} \left(g(x_n) - y_n \right) . 2 . x_{in} + \frac{\lambda}{2N} . \frac{\partial}{\partial \omega_i} \left(\omega_1^2 + \omega_2^2 + \dots + \omega_i^2 + \dots + \omega_n^2 \right)$$

$$= \frac{2}{2N} \sum_{n=1}^{N} \left(g(x_n) - y_n \right) . x_{in} + \frac{\lambda}{2N} . 2\omega_i$$

$$= \frac{1}{N} \sum_{n=1}^{N} \left(g(x_n) - y_n \right) . x_{in} + \frac{\lambda}{N} . \omega_i$$

Now,

$$\omega_i = \omega_i - \alpha \left(\frac{\partial J(\omega)}{\partial \omega_i} \right)$$

$$= \omega_i - \alpha \left(\frac{1}{N} \sum_{n=1}^{N} \left(g(x_n) - y_n \right) . x_{in} + \frac{\lambda}{N} . \omega_i \right)$$

$$\omega_i = \omega_i \left(1 - \frac{\alpha . \lambda}{N} \right) - \frac{\alpha}{N} . \sum_{n=1}^{N} \left(g(x_n) - y_n \right) . x_{in}$$

Q3.

Featuretransform(X, degree):

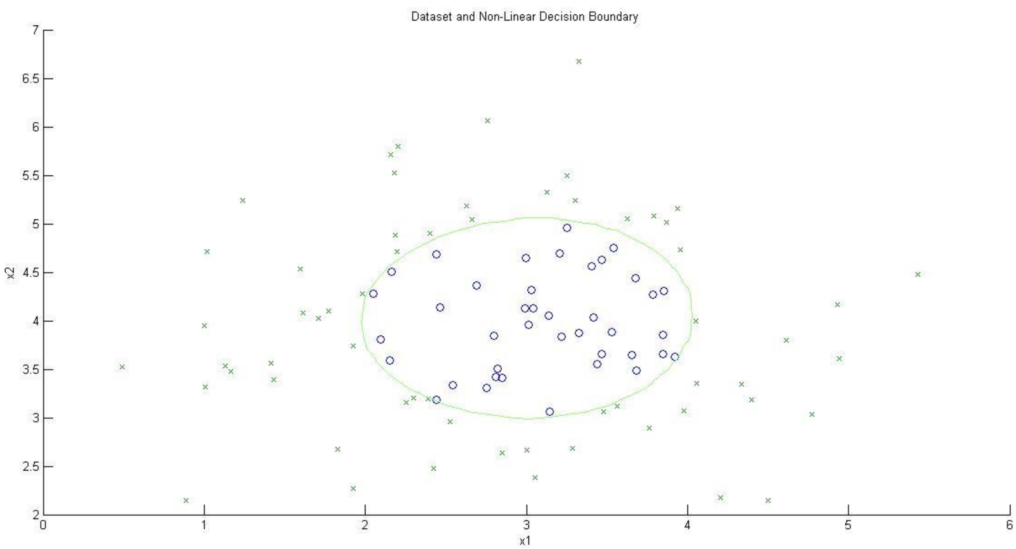
Since the data is not linearly separable, therefore it is transformed and more features are created.

Q. Plotdecisionboundary(w, X, Y) that plots the non-linear decision boundary that separates the two classes as learnt by the classifier.

Contour Plot for the curve that corresponds to regression value 0.5.

Observation :

The points are not linearly separable, so the features were transformed. The logistic regression creates a non-linear boundary separating the two classes, the classes are separated very accurately.

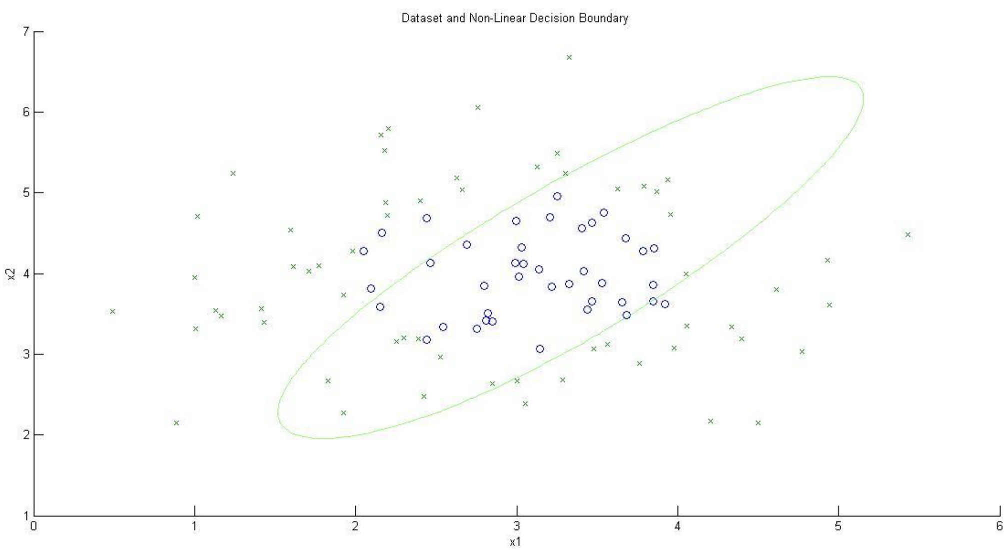


Q. Vary the value of the regularization parameter λ, and observe the changes in the decision boundary. Include in the report one figure each depicting under fitting and over fitting along with the value of λ.

Under fitting : λ = 10

Observation:

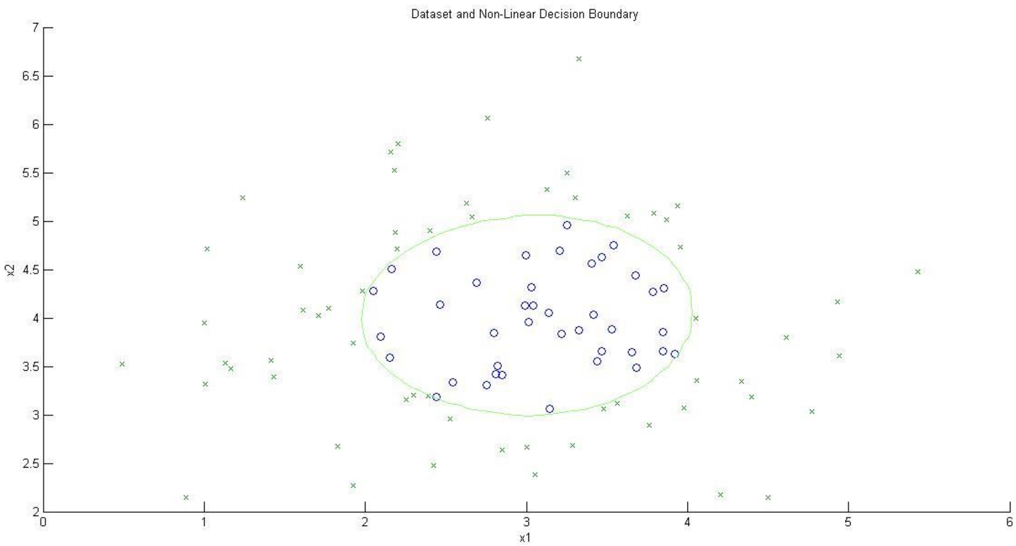
When the value of lambda is 10, the non-linear boundary of the curve is underfitting the given data.



Over Fitting $\lambda = 0$

Observation:

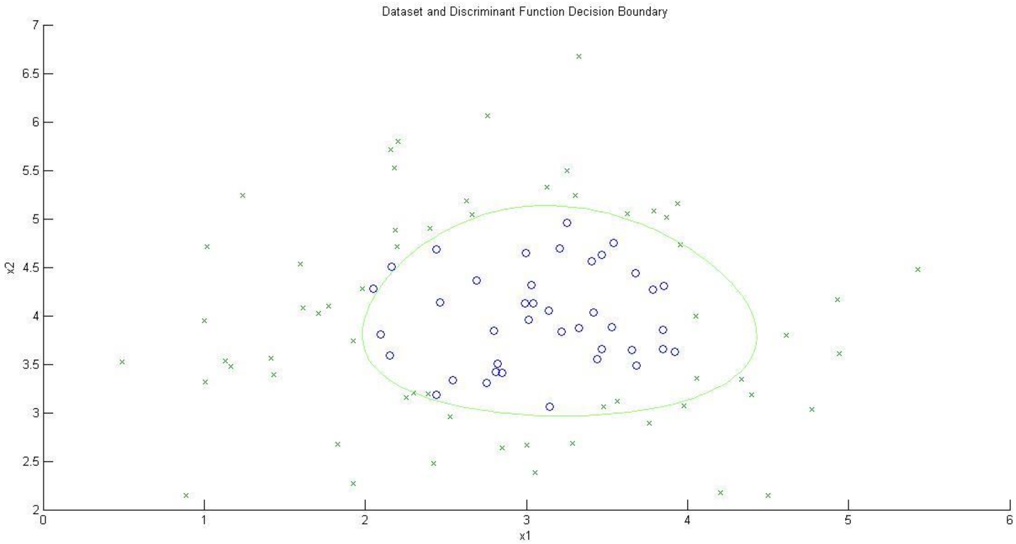
For the value of lambda=0 the non-linear curve is over-fitting the data.



Q. lindiscriminant(X, Y) computes the discriminants for two classes given the input data and target variables. The function should also compute the decision boundary using the discriminant functions and overlays it on the plot of the training data

Observation :

The decision boundary predicted by the discriminant function is fitting the data very well .



Q4.(a)

4.

x_1 = Hours Studied

x_2 = Undergrad GPA

y = Receive an A.

LOGISTIC REGRESSION :-

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

$$w = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} -8 \\ 0.05 \\ 1 \end{bmatrix}$$

$x_1 = 5$ $\quad x_2 = 7.5$ $\quad \therefore x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 7.5 \end{bmatrix}$

$$p = \frac{1}{1+e^{-w^T x}} \qquad w^T x = \begin{bmatrix} -8 & +0.05 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \\ 7.5 \end{bmatrix}$$

$$= -8 + 0.25 + 7.5 = -0.25$$

$$\therefore p = \frac{1}{1+e^{-(-0.25)}} = \frac{1}{1+e^{0.25}} = 0.4378$$

$\therefore$ Probability = 43.7%.

Q4.(b)

(b) Let Number of Hours = H.

$\therefore \quad p = 0.6 = \dfrac{1}{1 + e^{-w^T x}}$

$$w^T x = \begin{bmatrix} -8 & 0.05 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ H \\ 7.5 \end{bmatrix}$$

$$= -8 + 0.05H + 7.5$$

$$= 0.05H - 0.5$$

$$0.6 = \frac{1}{1 + e^{0.5 - 0.05H}}$$

$$e^{0.5 - 0.05H} = \frac{1}{0.6} - 1 = \frac{10}{6} - 1 = \frac{4}{6} = \frac{2}{3}$$

$$0.5 - 0.05H = \log_e (2/3) = -0.4055$$

$$0.05H = 0.5 + 0.4055 = 0.9055$$

$\therefore H = 18.11$ Hrs.