ASSIGNMENT-1 REPORT

Pushpam Anand

(2016CS10347)

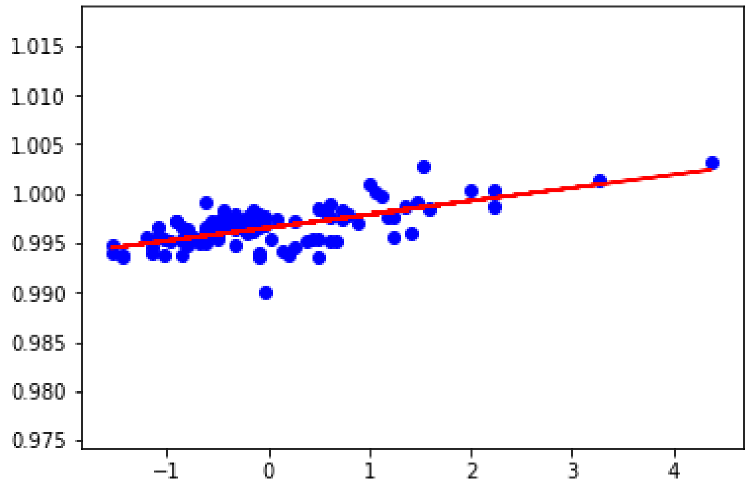
Date : 12-02-2019

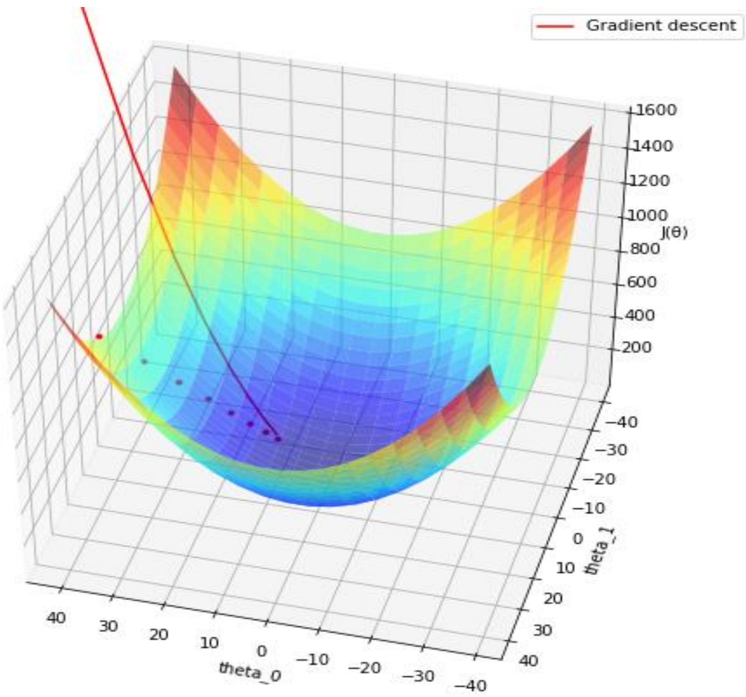
1. QUESTION 1

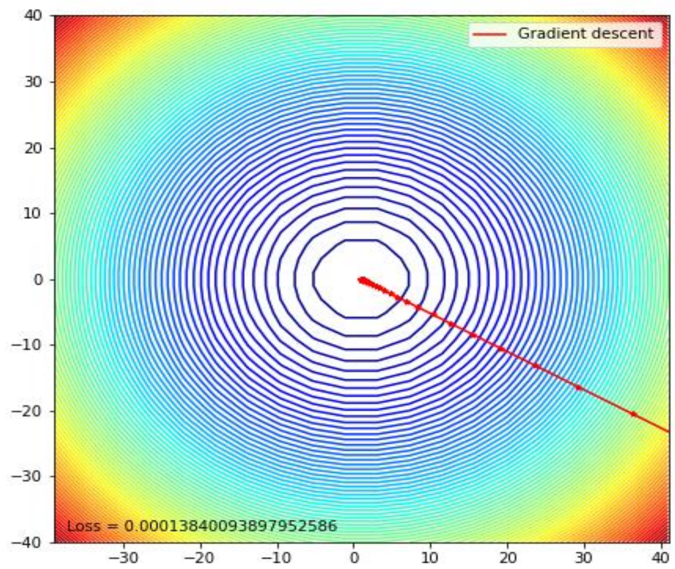
Learning rate = 0.2
Stopping criteria:
Threshold value = 0.00001

If the loss function in does not change more than threshold value in one update then the gradient descent converges at that θ .

Parameters are:
$\theta_0 =$ 0.0010478
$\theta_1 =$ 0.997124







As we increase the value of the η the learning becomes faster and faster, but if we increase more than a certain limit, the minima is overshoot and we never converge.

η = 0.1,0.5,0.9 (gradient descent converges quickly)
η = 1.3,1.7,2.1 (ocillates around minima)
η = 2.5 (overshoots the minima , does not converge)

2. QUESTION 2

(a)

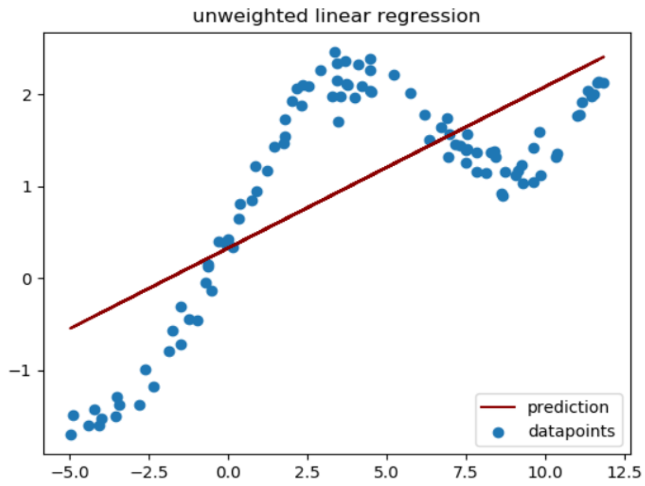
Unweighted Linear regression

The normal equation:

$\boldsymbol{\theta} = (\boldsymbol{X^T X})^{-1}(\boldsymbol{X^T Y})$

The diagonal matrix W is such that:

$W_{ii} = w_i$



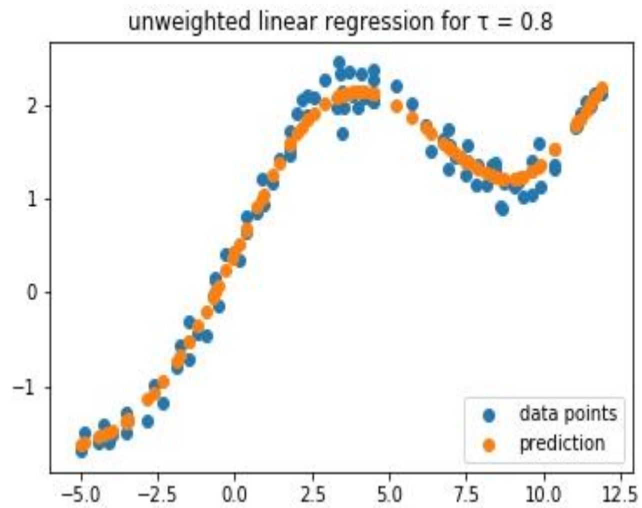
(b, c)

Locally weighted linear regression:

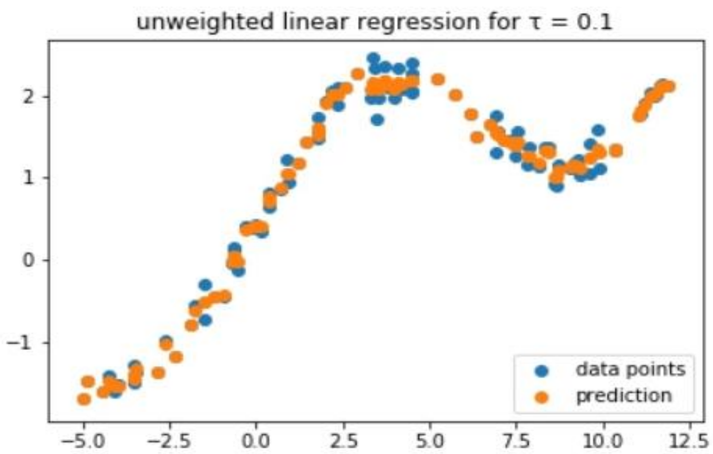
The normal equation:

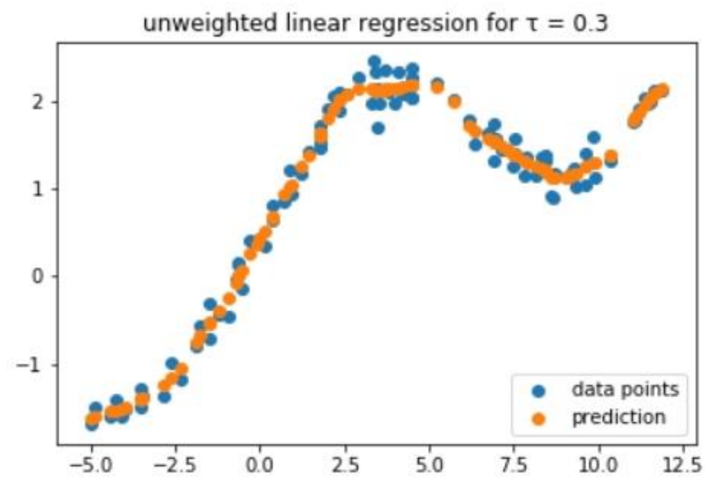
$\boldsymbol{\theta} = (\boldsymbol{X^T W X})^{-1}(\boldsymbol{X^T W Y})$

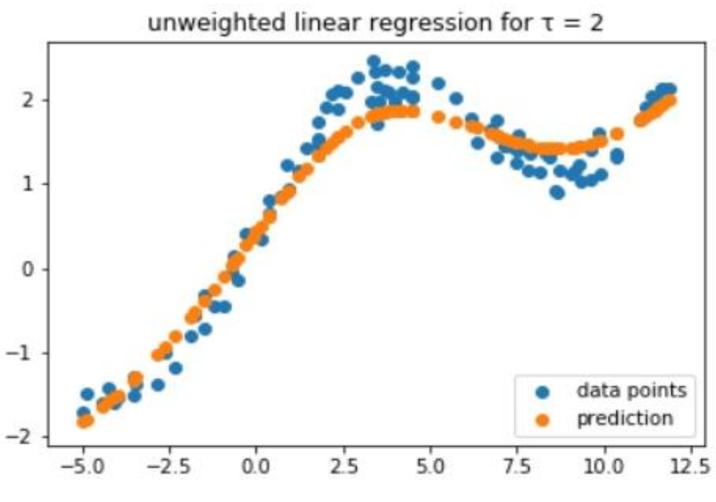
τ = 0.8 - 1.2 works best!

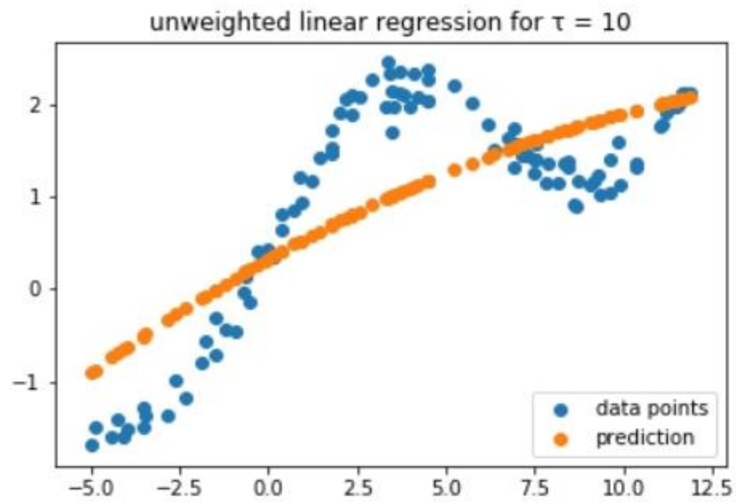


When the value of **τ** is very less then it results in overfitting of the curve to the data.
And when it is too large then the it results in poor fitting (overgeneralization).
τ =0.1,0.3 result in overfitting
τ =10 results in underfitting

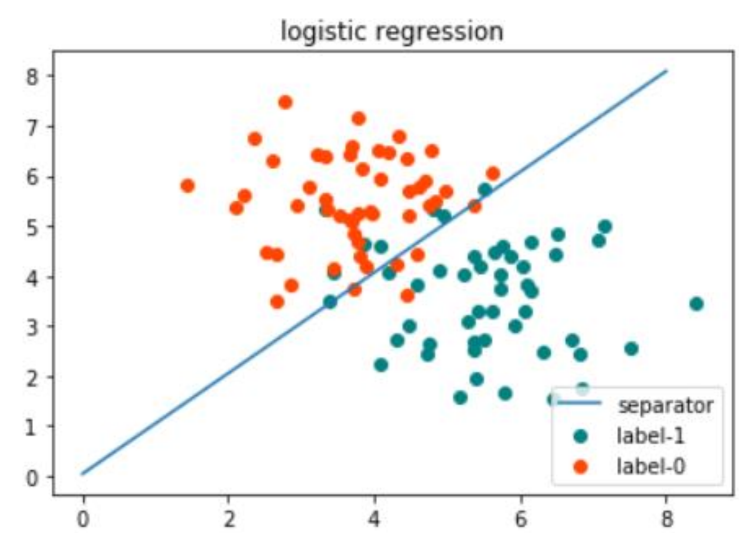








3. QUESTION 3



The value of the parameters are:

θ* = *[0.166 1.858 -1.856]

stopping parameters:

With a threshold value of convergence of θ to be 0.001 (i.e., the update converges when the difference of values of each of the parameters θ_j between new and old value is less than the threshold value.)

4. QUESTION 4

(a)

The values of parameters are:

μ_0 = [137.46 366.62]
μ_1 = [98.38 429.66]

Σ = [[287.482 -26.748]
[-26.748 1123.25]]

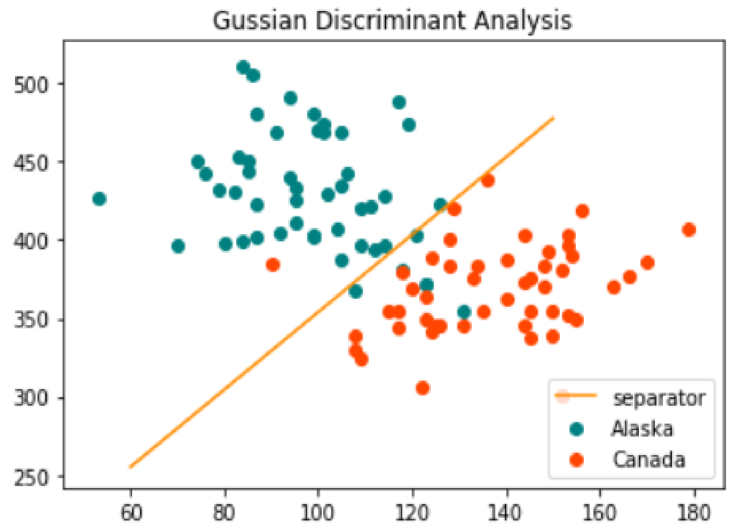
(b)

The equation of linear separator is:

$$\boxed{x^T(\Sigma^{-1}\mu_1^T - \Sigma^{-1}\mu_0^T) = \log\left(\frac{1-\phi}{\phi}\right) + \mu_0\Sigma^{-1}\mu_0 - \mu_1\Sigma^{-1}\mu_1}$$

The values of parameter θ obtained is:
[-5.6542, -0.1310, 0.05300]

The equation of straight line thus becomes:
$$(\mathbf{0.053})x_2 - (\mathbf{0.131})x_1 - \mathbf{5.654} = \mathbf{0}$$



(c)

The obtained equation of the separator is:

$$\boxed{x^T(\Sigma_1^{-1} - \Sigma_0^{-1})x + x^T(\Sigma_1^{-1}\mu_1^T - \Sigma_0^{-1}\mu_0^T) = \log\left(\frac{1-\phi}{\phi}\right) + \mu_0\Sigma^{-1}\mu_0 - \mu_1\Sigma^{-1}\mu_1}$$

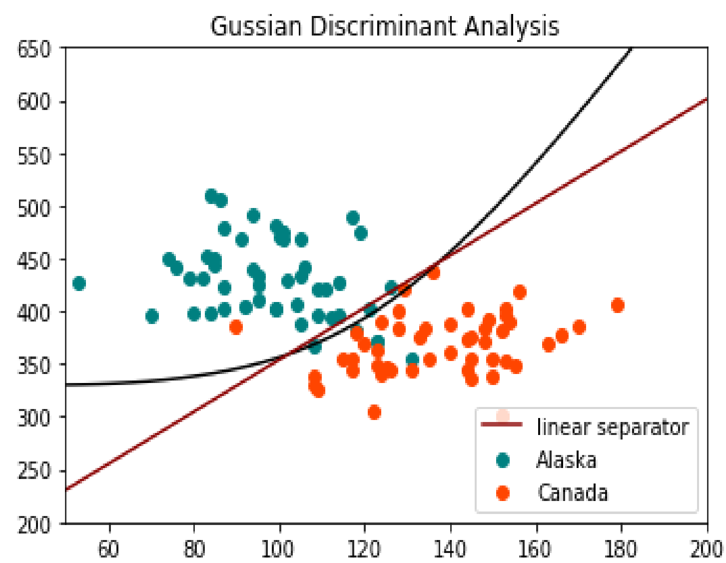
And the values of parameters are:

μ_0 = [137.46 366.62]
μ_1 = [98.38 429.66]

Σ_0 = [[255.3956 -184.3308]
[-184.3308 1371.1044]]

Σ_1 = [[319.5684 130.8348]
[130.8348 875.3956]]

ϕ = 0.5



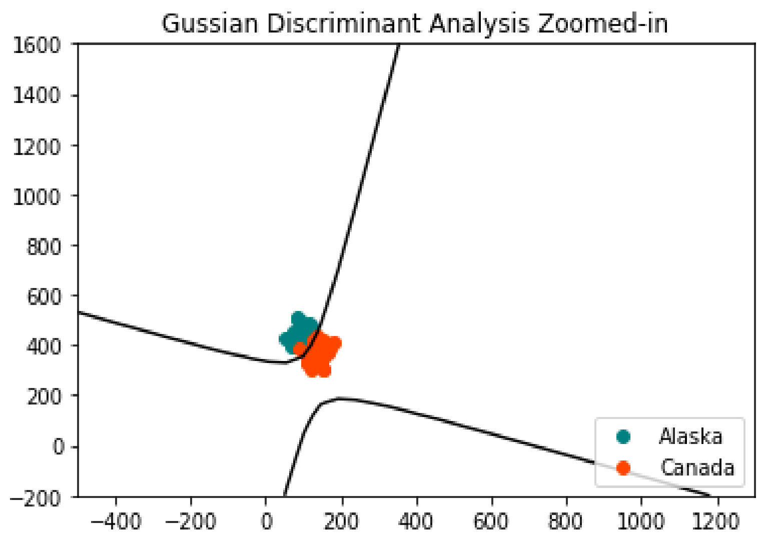
The equation of the separator thus becomes:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

Where:

a = 0.0005015
b = 0.001081
c = -0.0002045
d = -0.40153
e = -0.02677
f = 32.1140

The quadratic boundary is thus a Hyperbola.



The quadratic boundary better separates the data better than the linear separator, as it is evident from the plot in Q.4(b). This will become more evident when the dataset used is bigger and there are more points for training.
the quadratic boundary works every time the linear boundary suits the data, but reverse is not true. If the data is such that the regions are penetrated into each other and a clear linear boundary is not possible then quadratic boundary will work and linear separator fails, which is not so evident in this dataset.

References:

1. For drawing contour animations and surface plot animations in question 1.
   https://xavierbourretsicotte.github.io/animation_ridge.html
   (referred the code of “funcanimation” only, and not the gradient descent part)
2. Matplotlib and NumPy documentations for other functions.