NAME: DUONG DUY CHIEN

Logistic Regression

I. Theory:

Hypothesis:

$$h_\theta(x) = g(\theta^T x) = g(\theta_0 x_0 + \theta_1 x_1 + ... + \theta_n x_n)$$

$$= \mathrm{g(z)} = \frac{1}{1+e^{-z}}$$

Cost function:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}[-y^{(i)}\log(h_\theta(x^{(i)})) - (1-y^{(i)})\log(1-h_\theta(x^{(i)})]$$

Gradient descent algorithm:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$$

with m= number of samples

j= 1....n : number of feartures

$NOTE : the$ fomula of gradient is the same with linear regression, but the hypothesis is different

- The way to implement by vectorization is quite similar to the homework 1 so that I won't mention it again in here.

Calculate Gradient of Euclidean loss function of Logistic Regression:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})\frac{\partial h(\theta)}{\partial \theta_j} \qquad \mathrm{h}(\theta) = \frac{1}{1+e^{-\theta^T x}}$$

with

$$\frac{\partial h(\theta)}{\partial \theta_j} = \frac{-(1+e^{-\theta^T x})'}{(1+e^{-\theta^T x})^2} = \frac{e^{-\theta^T x} * x_j}{(1+e^{-\theta^T x})^2} = \frac{(1+e^{-\theta^T x} - 1) * x_j}{(1+e^{-\theta^T x})^2} = \left[\frac{1}{1+e^{-\theta^T x}} - \frac{1}{(1+e^{-\theta^T x})^2}\right] * x_j = (h - h^2) * x_j = h(1-h) * x_j$$

$$=> \frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)}) * h(1-h) * x_j$$

II. Results:

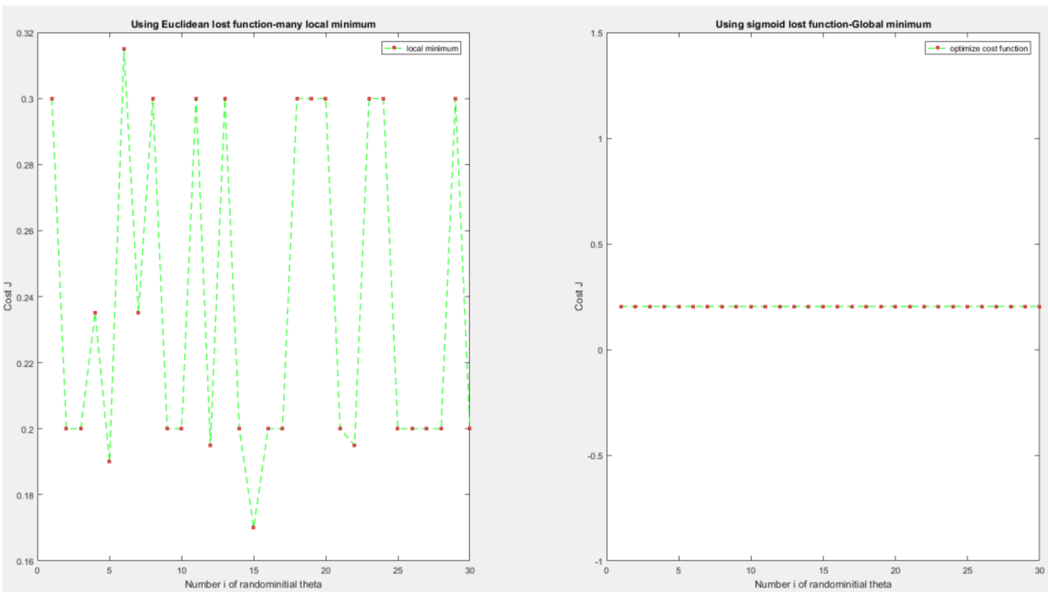
In logistic regression, we cannot use the same cost function which we use in Linear Regression. Because, if we use the same cost function, we will get stuck at local minimum (the Euclidean cost function of Logistic regression is not a bowl shape as Linear regression) (as shown in figure 1a, with different theta we can get stuck at different local minimum) then we may not have a good decision boundary as in figure 2.

Thus, we need to use another cost function (log function) as I write in the section so that the cost function now has the bowl shape, and we can find the global minimum. (as shown in figure 1b) then we can have a good decision boundary as in figure 3.

Minimize cost: 0.203511

Theta at converge: -24.861160 ;0.20385; 0.199014

Note: Because we use different equation to calculate cost function in figure 1. So that the values of cost function smaller don't mean that it is better.



(a) (b)

Figure 1 Cost function with different Initial theta. Figure(a) the Euclidian cost function get stuck at local minimum. Figure(b) uses log cost function get the global minimum

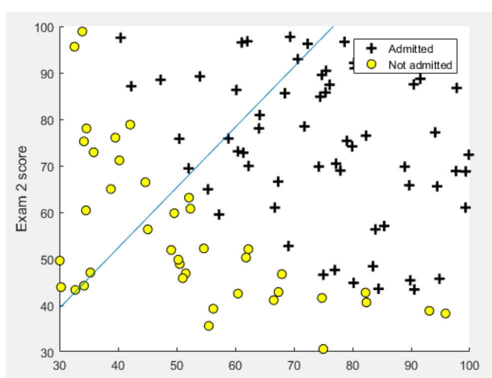


Figure 2 Decision boundary with euclidian lost function get stuck at local minimum.

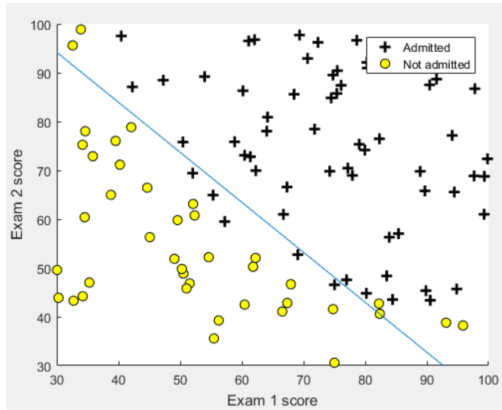


Figure 3 Decision boundary with log lost function at global minimum