Assignment-1(Linear Regression)

Artificial Intelligence (CSE-241N)
IIT (BHU) Varanasi

January 15, 2020

In this document we'll describe the equations for forward and backward passes in a succinct format.

1 Forward Pass

The forward pass for the linear regression looks like:

$$Y_{[n\times 1]} = X_{[n\times f]}\theta_{[f\times 1]} \tag{1}$$

where Y is the predicted values for the input data X and θ is the parameter/weight vector. The dimensions of the elements are in terms of n: the number of data points and f: length of feature for each data point. We'll use $\hat{Y}_{[n\times 1]}$ to represent the actual values corresponding to the data points x.

The loss function which we optimize is:

$$J = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

2 Backward Pass

The gradient descent update equation is given by:

$$\theta := \theta - \eta\frac{\partial J}{\partial \theta} \tag{3}$$

which is equivalent to

$$\begin{bmatrix}\theta_1\\ \vdots\\ \theta_f\end{bmatrix} = \begin{bmatrix}\theta_1\\ \vdots\\ \theta_f\end{bmatrix} - \eta\begin{bmatrix}\frac{\partial J}{\partial \theta_1}\\ \vdots\\ \frac{\partial J}{\partial \theta_f}\end{bmatrix}$$

Now, we know the fact that if $f(x, y)$ is a function. Also $x(W)$ and $y(W)$ are functions of W, then

$$\frac{\partial f}{\partial W} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial W} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial W} \tag{4}$$

In a similar way,

$$\begin{aligned}\frac{\partial J}{\partial \theta_j} &= \sum_{i=1}^{n} \frac{\partial J}{\partial y_i} \frac{\partial y_i}{\partial \theta_j} \\ &= \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i) \frac{\partial y_i}{\partial \theta_j}\end{aligned} \tag{5}$$

From the Equation(1), we have $y_i = \sum_{j=1}^{f} x_{ij}\theta_j$.

Using this in Equation(5), we get :

$$\begin{aligned}\frac{\partial J}{\partial \theta_j} &= \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i) x_{ij} \\ &= \frac{1}{n} \begin{bmatrix} x_{1j} & x_{2j} & \cdots & x_{nj} \end{bmatrix} \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix}\end{aligned} \tag{6}$$

Hence,

$$\begin{bmatrix} \frac{\partial J}{\partial \theta_1} \\ \frac{\partial J}{\partial \theta_2} \\ \vdots \\ \frac{\partial J}{\partial \theta_n} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1f} \\ x_{21} & x_{22} & \cdots & x_{2f} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nf} \end{bmatrix}^T \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix} \tag{7}$$

which is equivalent to

$$\frac{\partial J}{\partial \theta} = \frac{1}{n}.X^T.(Y - \hat{Y}) \tag{8}$$

Hence Equation(3) becomes:

$$\theta := \theta - \eta \frac{1}{n}.X^T.(Y - \hat{Y}) \tag{9}$$

This ends our discussion on backward and forward pass.