HINT FILE: Assignment 2 (Linear Regression on 2D data)
Artificial Intelligence (CSE-241N)
IIT (BHU), Varanasi

February 3, 2018

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

The forward pass for the linear regression looks like:

$$y_{[n\times 1]} = x_{[n\times f]} W_{[f\times 1]}, \tag{1}$$

where y is the predicted values for the input data x and W is the 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.

Now we'll give the loss equation:

$$L = \frac{1}{2n} \sum_{i=1}^{n} [y_i - \hat{y}_i]^2 \tag{2}$$

The gradient descent update equation is given by:

$$W = W - \eta \frac{\partial L}{\partial W} \tag{3}$$

which is equivalent to

$$\begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_f \end{bmatrix} = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_f \end{bmatrix} \text{ - } \eta \begin{bmatrix} \frac{\partial L}{\partial W_1} \\ \frac{\partial L}{\partial W_2} \\ \vdots \\ \frac{\partial L}{\partial W_f} \end{bmatrix} \tag{4}$$

Now, we know the fact that if $f = f(x, y)$ is a function, then

$$\frac{\partial f}{\partial W} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial W} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial W}. \tag{5}$$

In a similar way, if we view the loss function as $L = L(y_1, \ldots, y_n)$, then

$$\frac{\partial L}{\partial W_j} = \sum_{i=1}^{n} \frac{\partial L}{\partial y_i} \frac{\partial y_i}{\partial W_j} \tag{6}$$

$$= \sum_{i=1}^{n} \frac{1}{n} \left(y_i - \hat{y}_i\right) \frac{\partial y_i}{\partial W_j}. \tag{7}$$

From the Eq. 1, we have $y_i = \sum_{j=1}^{f} x_{ij} W_j$; using this in the above equation, we get:

$$\frac{\partial L}{\partial W_j} = \sum_{i=1}^{n} \frac{1}{n} \left(y_i - \hat{y}_i\right) x_{ij} \tag{8}$$

$$= \frac{1}{n} \begin{bmatrix} x_{1j} & x_{2j} & \dots & x_{nj} \end{bmatrix} \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix}. \tag{9}$$

Thus,

$$\begin{bmatrix} \frac{\partial L}{\partial W_1} \\ \frac{\partial L}{\partial W_2} \\ \vdots \\ \frac{\partial L}{\partial W_f} \end{bmatrix} = \frac{1}{n} \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1f} & x_{2f} & \dots & x_{nf} \end{bmatrix} \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix} \tag{10}$$

which is equivalent to

$$\frac{\partial L}{\partial W} = \frac{1}{n} \cdot x^{\top} \left(y - \hat{y}\right). \tag{11}$$

So finally, Eq. 3 becomes:

$$W = W - \eta \cdot \frac{1}{n} \cdot x^{\top} \left(y - \hat{y}\right). \tag{12}$$

This concludes our exposition on forward and backward passes of linear regression.