Week 1

Arjun Sarao

January 25, 2021

1 Hypothesis Function

The hypothesis function is the equation we use to predict values in linear regression problems. The later formulas are what we would use to determine the values of θ_0 and θ_1. The equation is:

$$h_\theta(x) = \theta_0 + \theta_1 x \tag{1}$$

2 Cost Function

To evaluate the accuracy of θ_0 and θ_1 in predicting data, we use a cost function. m is the number of datapoints and y^i is the real data-point. The form of the cost function is:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2 \tag{2}$$

Where m is the number of data-points and y^i is the index of the training data.

3 Gradient Decent

To find the most optimal values of θ_0 and θ_1 or the local minima we use a gradient decent. This formula is to be solved repeatedly until the optimal values for θ_0 and θ_1 have been found. The equation of the gradient decent algorithm is:

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \tag{3}$$

Where α is the learning rate and θ_j is simultaneously θ_0 and θ_1

4 Linear Algebra Review

The best way of representing sets of numbers is using matrices. A vector is a $n \times 1$ sized matrix and a regular matrix is a $n \times m$ sized one, where n is the number of rows and m is the number of columns.

$$A_{m \times n} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \tag{4}$$

Adding and subtracting matrices is only possible when the two matrices are the same size.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} (a+e) & (b+f) \\ (c+g) & (d+h) \end{bmatrix} \tag{5}$$

We can also multiply matrices by numbers called Scalars. To do this we can simply multiply each element of the matrix by the scalar.

$$x \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a(x) & b(x) \\ c(x) & d(x) \end{bmatrix} \tag{6}$$

Additionally we can multiply matrices by other matrices. To do so we multiply the the row of one matrix and the column of the other:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} (ae+bg) & (af+bh) \\ (ce+dg) & (cf+dh) \end{bmatrix} \tag{7}$$

Matrices are not commutable, which means that multiplication can only work one way. Let A and B be matrices. Then in general, $A \times B \neq B \times A$ There are however, exceptions to this rule. One of which are a special type of matrix know as Identity Matrices. They are denoted by $I_{n \times n}$. They can be of any size.

$$I_{3 \times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

We can also take the inverse of a matrix. This is done by: switching the first and last elements (from left to right), turning the other elements negative, and finally multiplying by the determinant.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{9}$$

then:

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{10}$$

Finally Matrices have a special operation they can undergo called a Matrix Transpose. This is when you "flip" the matrix by across its diagonal going from its top left to bottom right.

$$A = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} \tag{11}$$

then:

$$A^T = \begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix} \tag{12}$$