**Machine Learning by Andrew Ng**

**Week 2 short notes**

Gradient Descent can be slow and inefficient due to multiple reasons, in order to make it run faster and more efficiently machine learning programmers adopt certain techniques.

**Feature Scaling** involves dividing the input values of the features by the range of the input variables so as to obtain the results in a range of just 1.

**Mean Normalization** involves the subtracting the mean value of the input variables from each input variable so as to obtain the new mean as zero after this operation.

The speed and correctness of a gradient descent depends heavily on its learning rate.

**Learning Rate *α* (alpha)**

The value of the learning rate should be so as to decrease the value of the cost function on every iteration. Hence it is necessary that it should be set to an optimal value because a very high value may lead to non-convergence and very small value may lead to very slow convergence.

The hypothesis function we obtain using our algorithms need not always be linear but rather can be polynomial in nature. This simply means the terms of function may be combinations of multiple features or raised to different rational number powers or a combination of both.

**Normal Equation** is an alternative method used instead of Gradient Descent for obtaining optimization, in this method the features and outcomes are taken as matrix (M X N+1) and M dimensional array respectively and the optimized array of parameters is given by:

**Θ = (XT X)-1 XT Y**