Multivariate linear regression:

Here the hypothesis function hX) = XX1X2nXn I.e. with n features. This same function can be vectorized to hX) = X .where and X are column vectors containing the respective vectors.

Gradient descent for multivariate linear functions:

Gradient descent calculates the parameters n

jjhXiyiXij

It can be noticed that the equation is mostly similar to linear regression in single variables except here there n feature thus n parameters. Same as before the parameters are updated simultaneously.

Analytical calculations:

Instead of iterating through the given parameters it is sometimes better to calculate them through the use of linear equations. One of the way is to use Normal equations.

Normal equations:

= (XTX)−1XTy

One of the major drawbacks of using normal equations is that if there are large numbers of features the computation cost of computing the inverse of XTX is very high.

Classification Problems:

The aforementioned methods are used for giving a continuous output. Classifications problems involve methods to classify the data into a finite number of classes.

Hypothesis function:

Here the hypothesis function can be represented as a sigmoid function (logistic function)  
1/1+*e*−*z*



As the graph represents the function takes real value and converts it into values varying from 0 to 1. Here the function represents the probability of the output being either one or zero.