**Logistic Regression**

Logistic regression studies the probability of the sample to belong to a particular class.

It uses Bernoulli distribution, a member of the Exponential Family.

φ is the probability of an event.

Here we use the sigmoid function to output the probability of the event.

While training,

If ‘x’ is a vector of features. Let ‘w’ be the vector of coefficients of the features and ‘b’ be a constant.

Therefore,

Where h(x) = g(w.x+b).

The above expression is for one sample. The probability for all the samples in the data will be the product of the probabilities of individual samples. Therefore,

() is the i th sample.

This is also called as the likelihood of the parameters ‘w’ and ‘b’. If we take the log of the above expression and choose w and b to maximize the log of likelihood using gradient ascent, we will get the update rule to be,

α is the learning rate, a hyperparameter.

The cost function used here is the negative log of likelihood. i.e,