When the dependent variable is categorical, then logistic regression is one of the most appropriate regression model. Logistic regression is a binary classifier, but multiple classification is possible using one vs. all model. If a data set D = (x,y) : x ϵ X , y ϵ Y where X is the independent variables and Y is the set of category. The hypothesis function h θ(x) is given below which is a sigmoid function,

$$\text{h}_{\theta}(\text{x}) = \frac{1}{1+e^{-\theta^T X}}$$

Using this hypothesis function a curved line is drawn which is used to classify the data. To derive the coefficients a cost function J(θ) is used. The value of J(θ) is shown in following equation,

$$\text{J}(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y_{\text{i}} \log h_{\theta}(\text{x}_{\text{i}}) + (1\text{-}\text{y}_{\text{i}})\log(1 - h_{\theta}(\text{x}_{\text{i}}))]$$

Figure 1: Basic Logistic Regression

After minimizing the cost function, the values of the coefficients are derived. The following equation is used to update the values of θ,

$$\theta_{\text{j}} = \theta_{\text{j}}\text{-}\ \frac{1}{m}\ \sum_{i=1}^{m}(h_{\text{j}}(x_{\text{i}}) - y_{\text{i}})x_{\text{ij}}$$

1 Logistic Regression