Logistic Regression

The question is how should we model the relationship between

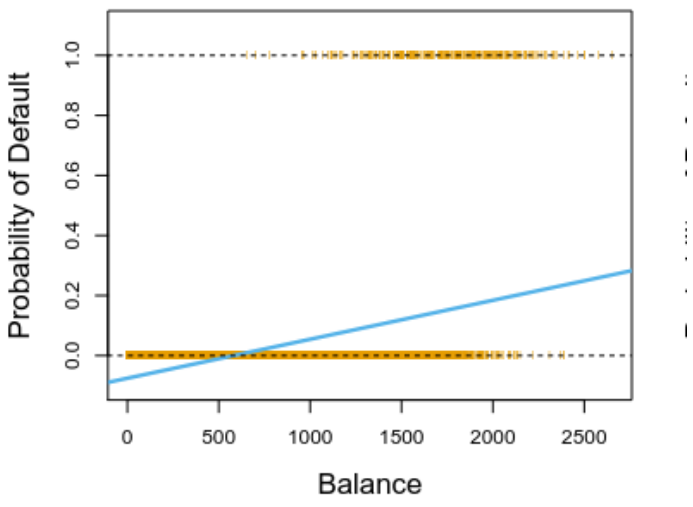
$$P(X) = \Pr(Y = 1|X)$$

- Using 1 and 0 for the Response

Using Linear Regression model to represent these probabilities

$$p(X) = \beta_0 + \beta_1 X$$

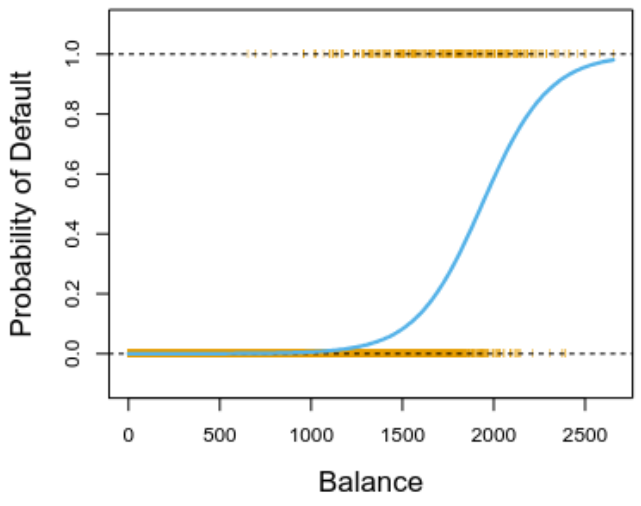
- If we fit the line to predict the **Probability**



- Notice that the `Balance` Lower than 500 our prediction for the probability is **negative**

To avoid this problem we model $p(X)$ to only fall between 1 and 0 for all values X **Logistic Function** which is a Sigmoid Function

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}}$$



- Any output of the **Logistic Function** Falls between 1 or 0

$$\frac{p(X)}{1 - p(X)} = e^{\beta_0+\beta_1 X}$$

- the **quantity** $p(X)/[1 - p(X)]$ is called the $odds$ can can only take values between 0 to ∞
- Values close to 0 indicates low probability
- Values close to ∞ indicate higher probability