Welcome to

Logistic Regression

Day 2

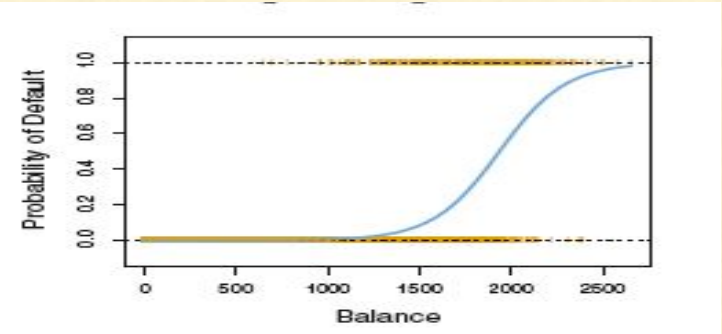
LOGISTIC REGRESSION - THEORY

- Consider the data where the target variable is binary, i.e., outcome you want to model either Yes or No, 0/1 etc.
- Rather than modeling this response Y directly, logistic regression models the probability that Y belongs to a particular category.
- But given the normal regression approach, there is no limitations on the values that the dependent variable can take!
- Since, the range of probability is between 0 and 1, we need to use a function, that can give an output between 0 and 1 for all the values of X (independent variables).

LOGISTIC REGRESSION - THEORY

Case Study: Loan Default Prediction using Deposit Balance

$$Estimate\ of\ P(y=1 \mid x_1, \ldots x_p) = 1/(1+e^{-(a+\sum_k b_k x_k)})$$



- Consider the case where you want to predict the probability of default on loan, given the saving account balance of the customer. You want the model to provide a default probability score between 0 and 1 given the account balance.
- The above function will always produce a S-shaped curve given above and so, regardless of values of X, you will get sensible values of probability.

LOGISTIC REGRESSION - Hands-on

- We will explore mtcars dataset
- Go to Equiskill lab http://lab1.equiskill.com
- Check mtcars dataset
- ?mtcars
- Open file logistic_regression.R

LOGISTIC REGRESSION - Hands-on

Problem

Estimate the probability of a vehicle being fitted with a manual transmission if it has a 120hp engine and weights 2800 lbs.

Steps

1. Fit Logistic regression model
2. Predict the probability

LOGISTIC REGRESSION - Hands-on

Answer

For an automobile with 120hp engine and 2800 lbs weight, the probability of it being fitted with a manual transmission is about 0.64.