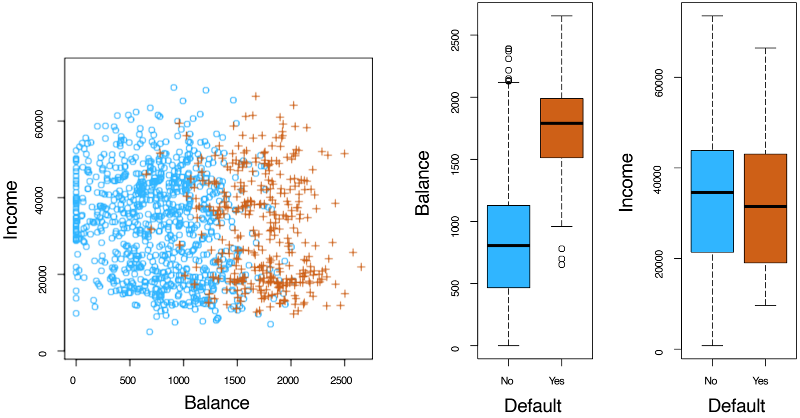
Classification

- Categorical variables take values in an unordered set, such as:
  `eye color` $\in$ {`brown`, `blue`, `green`}
  `email` $\in$ {`spam`, `ham`}.
- Given a feature X and a categorical target Y, the classification task is to build a model that takes as input the feature X and predicts its value for Y;
- Often we are more interested in estimating the *probabilities* that X belongs to each category in Y.

Example: Credit Card Default



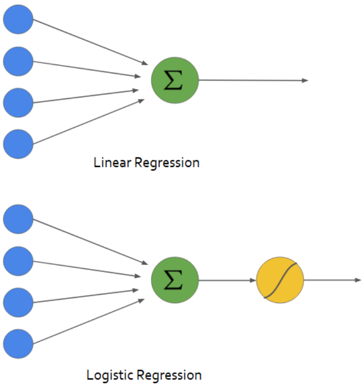
Can we use Linear Regression?

Suppose for the Default classification task that we code

$$Y = \begin{cases} 0 \; if \; No \\ 1 \; if \; Yes \end{cases}$$

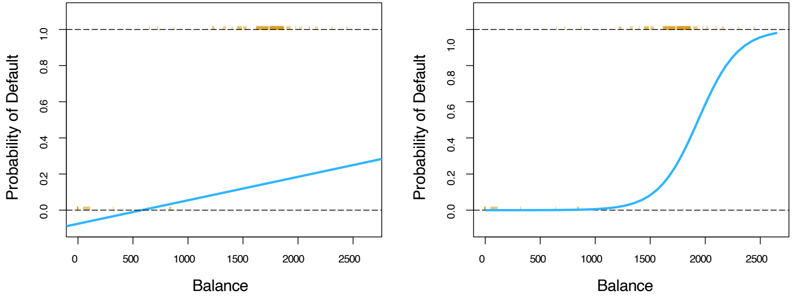
Can we simply perform a linear regression of Y on X and classify as Yes if $\hat{Y} > 0.5$?

Linear Regression vs. Logistic Regression



- In Linear Regression, the output is the weighted sum of inputs.
- Logistic Regression doesn't output the weighted sum of inputs directly, but passes it through an activation function that can map any real value between 0 and 1.

Linear versus Logistic Regression



The orange marks indicate the response Y, either 0 or 1. Linear regression does not estimate $\Pr(Y = 1|X)$ well. Logistic regression seems well suited to the task.

Logistic Regression

Let's write $p(X) = \Pr(Y = 1|X)$ for short and consider using `balance` to predict `default`. Logistic regression uses the form

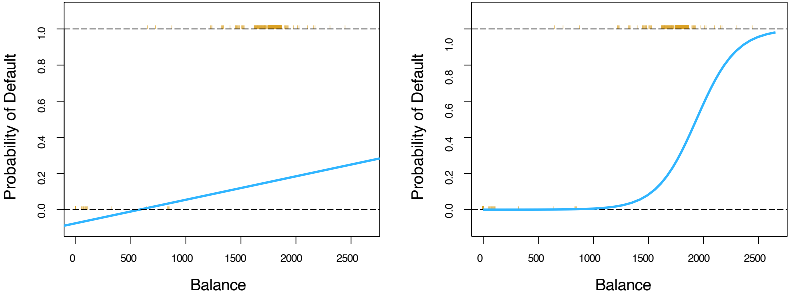
$$P(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

($e \approx 2.71828$ is a mathematical constant [Euler's number.])

It is easy to see that no matter what values β_0, β_1 or X take, $p(X)$ will have values between 0 and 1.

https://machinelearningmastery.com/a-gentle-introduction-to-sigmoid-function

Linear versus Logistic Regression



Logistic regression ensures that our estimate for $p(X)$ lies between 0 and 1.

Making Predictions

What is our estimated probability of `default` for someone with a balance of \$1000?

$$\hat{p}(X) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X}} = \frac{e^{-10.6513 + 0.0055 \times 1000}}{1 + e^{-10.6513 + 0.0055 \times 1000}} = 0.006$$

Making Predictions

What is our estimated probability of `default` for someone with a balance of \$1000?

$$\hat{p}(X) = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 X}}{1 + e^{\hat{\beta}_0+\hat{\beta}_1 X}} = \frac{e^{-10.6513+0.0055\times 1000}}{1 + e^{-10.6513+0.0055\times 1000}} = 0.006$$

With a balance of \$2000?

$$\hat{p}(X) = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 X}}{1 + e^{\hat{\beta}_0+\hat{\beta}_1 X}} = \frac{e^{-10.6513+0.0055\times 2000}}{1 + e^{-10.6513+0.0055\times 2000}} = 0.586$$