EE4708 - Data Analytics Laboratory 2020

Week 6

Logistic Regression

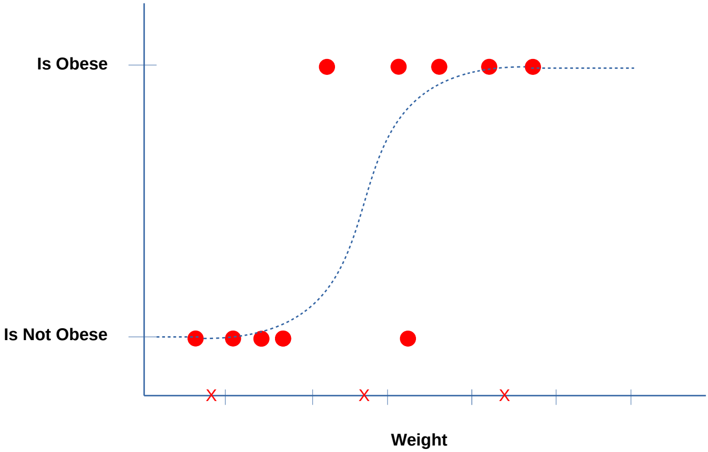

Figure: Example: Whether a person is obese or not?

Logistic Regression

Logistic regression is a technique that can be used to handle binary classification problems.

1. It is a special case of linear regression where the target variable is categorical in nature and therefore can be used for classification
2. It takes the features as input and predicts the probability of occurrence of a binary event using a logitic function
3. The probability is calculated by taking the logistic of a linear regression function as shown in the following slide

Logistic Regression (cont.)

- The linear regression function as defined earlier is given by

$$z = \theta_0 + \theta_1 x_1 + \ldots + \theta_d x_d$$

- The logistic function is given by:

$$y_{logistic}(z) = \frac{1}{1 + e^{-z}}$$

$$y_{logistic}(z) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x_1 + \ldots + \theta_d x_d)}}$$

- The above logistic function models the probability that sample $\boldsymbol{x}$ belongs to class 1 in a binary classification task.

Logistic Regression (cont.)

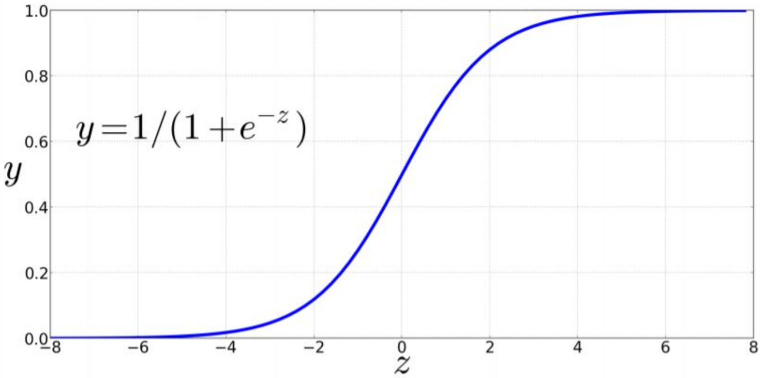

Figure: The logistic function $y = \frac{1}{1+e^{-z}}$ takes a real value and maps it to the range [0,1].

Note: y is nearly linear around 0 but outlier values get squashed toward 0 or 1.

Logistic Regression (cont.)

Working Principle:

If we apply the logistic to the sum of the weighted features, we get a number between 0 and 1. To make it a probability, we just need to make sure that the two cases, P(y = 1) and P(y = 0), sum to 1.

$$P(y=1) = \frac{1}{1+e^{-(\theta.x+\theta_0)}} \quad P(y=0) = \frac{e^{-(\theta.x+\theta_0)}}{1+e^{-(\theta.x+\theta_0)}}$$

For a test instance x, we say yes (class 1) if the probability $P(y=1|x)$ is greater than 0.5, and no (class 0) otherwise. We say 0.5 the decision boundary.

$$\hat{y} = \begin{cases} 1 & \text{if } P(y=1|x) > 0.5 \\ 0 & otherwise \end{cases}$$