Logistic Regression

Logistic regression (Classification Algorithm)

- It is a predictive analysis algorithm and based on the concept of probability.
- Measures the relationship between the dependent variable and the one or more independent variables (features), by estimating probabilities using logistic function.
- It used to assign observations to a discrete set of classes.
- E.g.
  - Email spam or not spam,
  - Online transactions Fraud or not Fraud,
  - Tumor Malignant or Benign.
- Logistic regression transforms its output using **sigmoid (logistic)** function to return a probability value.
- Logistic regression limits the hypothesis function between 0 to 1 i.e. h(x) / $y' \epsilon$ [0,1]

$$0 \leq h_\theta(x) \leq 1$$

Types of classification using Logistic Regression

- Binary classification(e.g. Tumor Malignant or Benign)
- Multi-Class classification(eg. Cats, dogs or Sheep's)

Assumptions for Logistic Regression:

- The dependent variable must be categorical in nature.
- The independent variable should not have multi-collinearity i.e. independent variables must be independent of each other.

Sigmoid Function

- Maps the predicted values to probabilities.
- Sigmoid Function maps any real value into another value between 0 and 1.
- Hence it forms a curve like the "S" form. The S-form curve is called the Sigmoid function or logistic function.

$$\sigma(z) = p = \frac{1}{1+e^{-z}} = \frac{e^{z}}{e^{z}+1}.$$

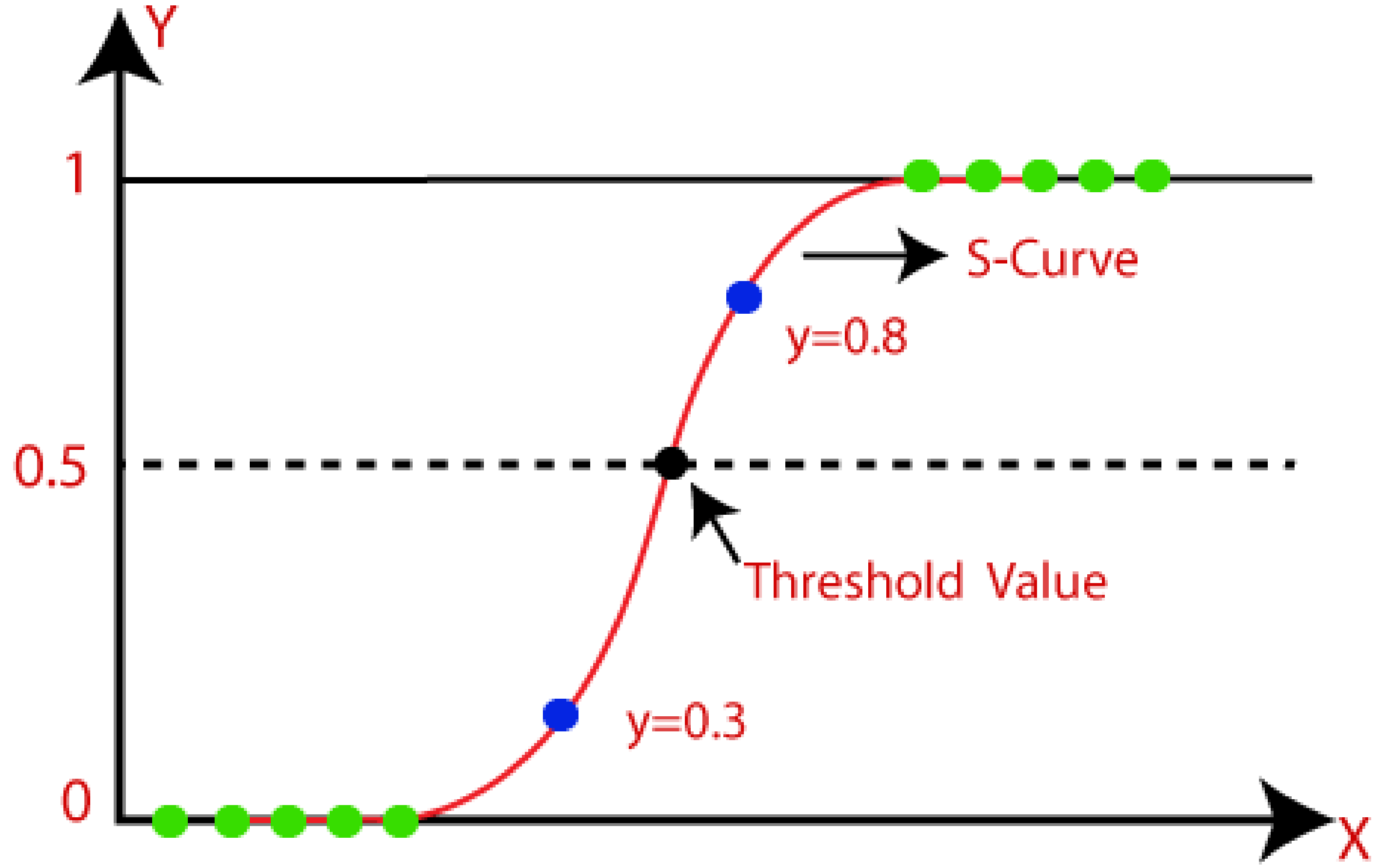
- $z = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots\ldots + \theta_n x_n$ [Here, x_0 (bias) is always 1]
  - x_i is independent variable $i = 0,1,2, \ldots, n$

Decision boundary

Threshold classifier output y' *or* $h_\theta(x)$ at 0.5:

If $\boldsymbol{h_\theta(x) \geq 0.5}$, predict "**y = 1**"

If $\boldsymbol{h_\theta(x) < 0.5}$, predict "**y = 0**"



Linear Decision boundary

- Predict 'y = 0', if $3 + x_1 \ + x_2$ >=0
- $z = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots\ldots + \theta_n x_n$

Non-Linear Decision boundary

- Predict 'y = 0', if $3 + x_1^2 + x_2^3$ >=0
- $z = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_1 x_1^2 + \theta_2 x_2^2$

Cost / Error/ Loss function

- Cost Function $J(\Theta) = -(\frac{1}{m}\sum_{i=1}^{m}(Y^{(i)}\, log\, {y'}^{(i)} + (1 - Y^{(i)})\, log(1 - {y'}^{(i)})))$

$Y^{(i)}$ - Ground truths or Actual output

${y'}^{(i)}$ - Prediction output

m - No. of data points or samples

Gradient Descent

- The objective of training a machine learning model is to minimize the loss or error between ground truths and predictions by changing the trainable parameters.
- Gradient is the extension of derivative in multi-dimensional space, tells the direction along which the loss or error is optimally minimized.
- Gradient is defined as the maximum rate of change.

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

- θ_j-Training parameter α – Learning rate $J(\theta)$ – Error / Cost function

Practical Steps in Logistic Regression:

To implement the Logistic Regression using Python,

- Data Pre-processing step
- Fitting Logistic Regression to the Training set
- Predicting the test result
- Test accuracy of the result(Creation of Confusion matrix)
- Visualizing the test set result.

References

- Richard Szeliski, Computer Vision: Algorithms and Applications, Springer 2010
- Artificial Intelligence and Machine Learning, Chandra S.S. & H.S. Anand, PHI Publications
- Machine Learning, Rajiv Chopra, Khanna Publishing House