# SNS COLLEGE OF TECHNOLOGY

## Coimbatore-37.

## An Autonomous Institution

**COURSE NAME : 19CSE301-INTRODUCTION TO DATA SCIENCE**

**UNIT – IV CLASSIFICATION**

**Topic: Loss function for the Logistic Model**

Ms.B.Sumathi
Assistant Professor
Department of Computer Science and Engineering

# LOGISTIC MODEL

- Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set.

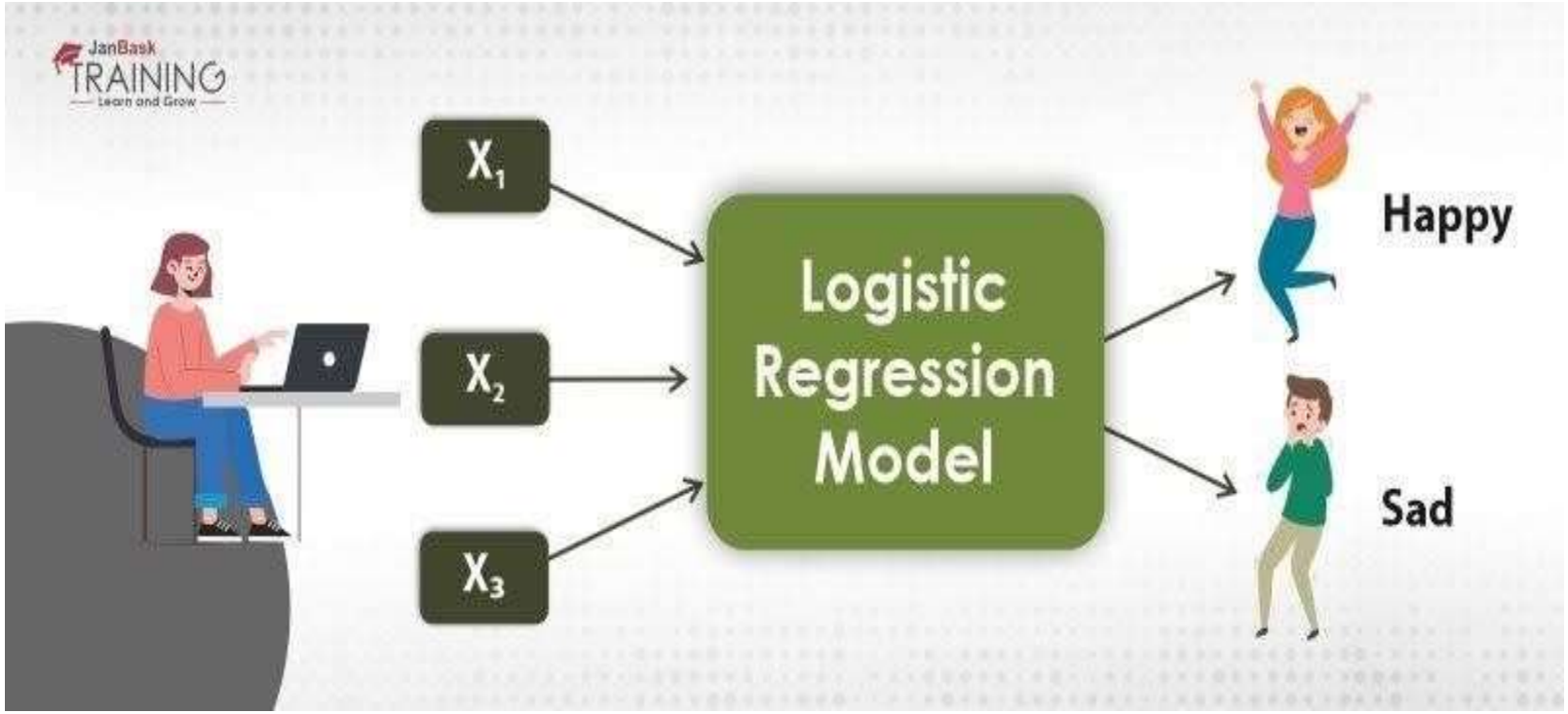
- A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables.

# LOGISTIC MODEL

For example :

- A logistic regression could be used to predict whether a political candidate will win or lose an election or whether a high school student will be admitted or not to a particular college. These binary outcomes allow straightforward decisions between two alternatives.

# LOGISTIC MODEL



# LOSS FUNCTION

- The loss function is the function that computes the distance between the current output of the algorithm and the expected output. It's a method to evaluate how your algorithm models the data.

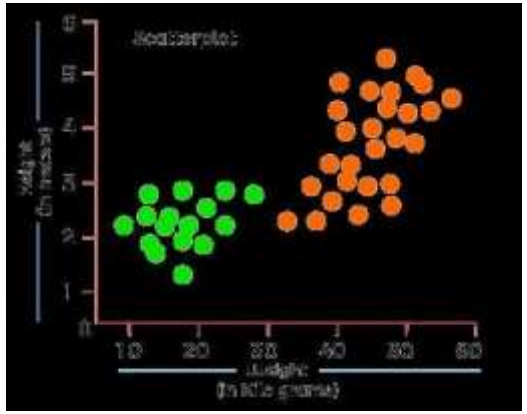
- It can be categorized into two groups. One for classification (discrete values, 0,1,2…) and the other for regression (continuous values).

- The loss associated with a decision should be the difference between the consequences of the best decision that could have been made had the underlying circumstances been known and the decision that was in fact taken before they were known.

# LOSS FUNCTION

- A common example

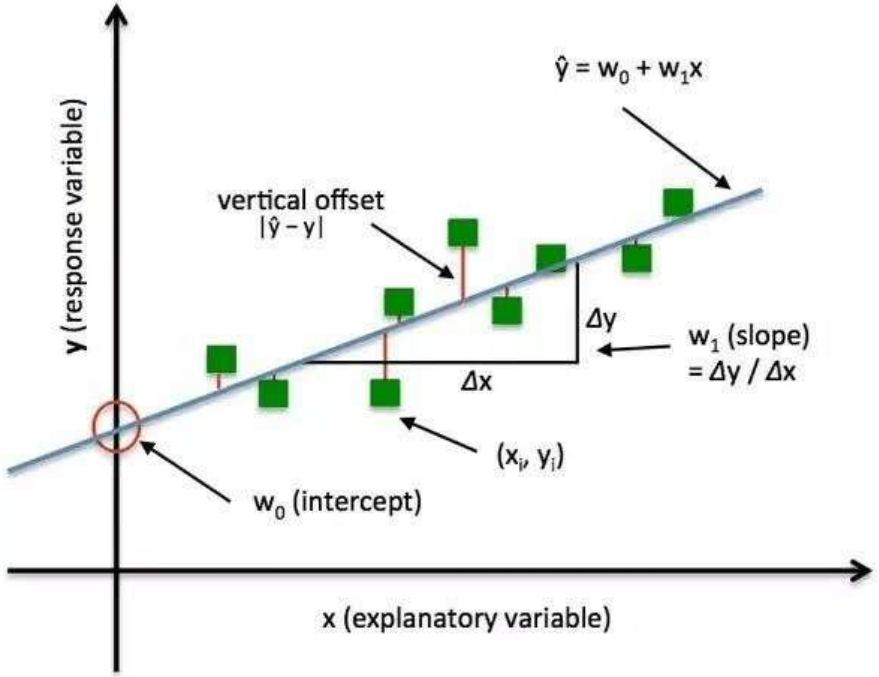
Involves estimating “location”. Under typical statistical assumptions, the mean or average is the statistic for estimating location that minimizes the expected loss experienced under the squared-error loss function, while the median is the estimator that minimizes expected loss experienced under the absolute-difference loss function. Still different estimators would be optimal under other, less common circumstances



# COST FUNCTION

- Cost function refers to the functional relationship between cost and output. It studies the behaviour of cost at different levels of output when technology is assumed to be constant.
- It can be expressed as below: C= f(Q) (Here, C= Cost of production; and Q= Quantum of output).
- The graph of cost function when y=1 and y=0

# COST FUNCTION



# FITTING

- Model fitting is a measure of how well a machine learning model generalizes to similar data to that on which it was trained.

- A model that is well-fitted produces more accurate outcomes.A model that is overfitted matches the data too closely. A model that is underfitted doesn't match closely enough.

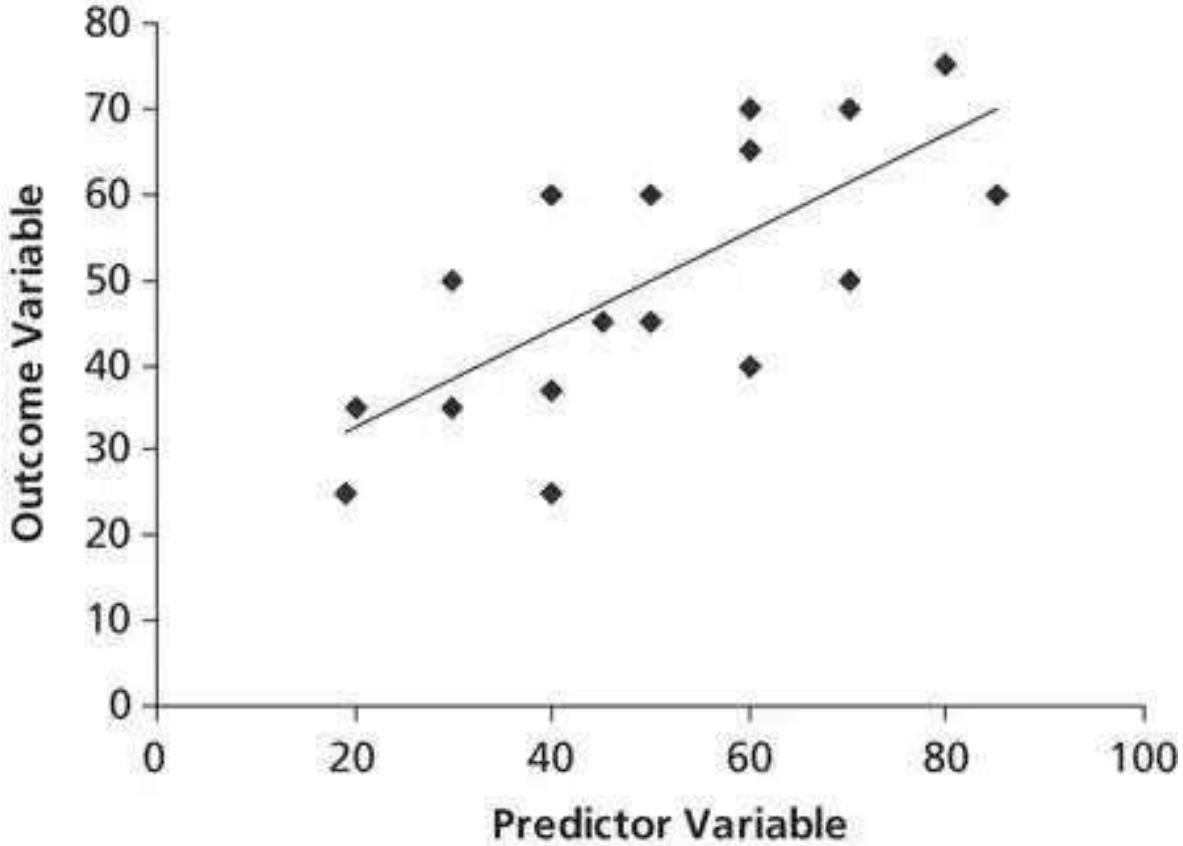
- One involves a gradient method such as the Levenberg-Marquardt method which uses the gradient and a second derivative matrix to quickly find the local minimum nearest to the starting point.

# FITTING

- **Underfitting**: A statistical model or a machine learning algorithm is said to have underfitting when it cannot capture the underlying trend of the data, i.e., it only performs well on training data but performs poorly on testing data. (It's just like trying to fit undersized pants!) Underfitting destroys the accuracy of our machine learning model.

- **Overfitting**: A statistical model is said to be overfitted when the model does not make accurate predictions on testing data. When a model gets trained with so much data, it starts learning from the noise and inaccurate data entries in our data set.

# FITTING


Outcome Variable
Predictor Variable
0
10
20
30
40
50
60
70
80
0
20
40
60
80
100

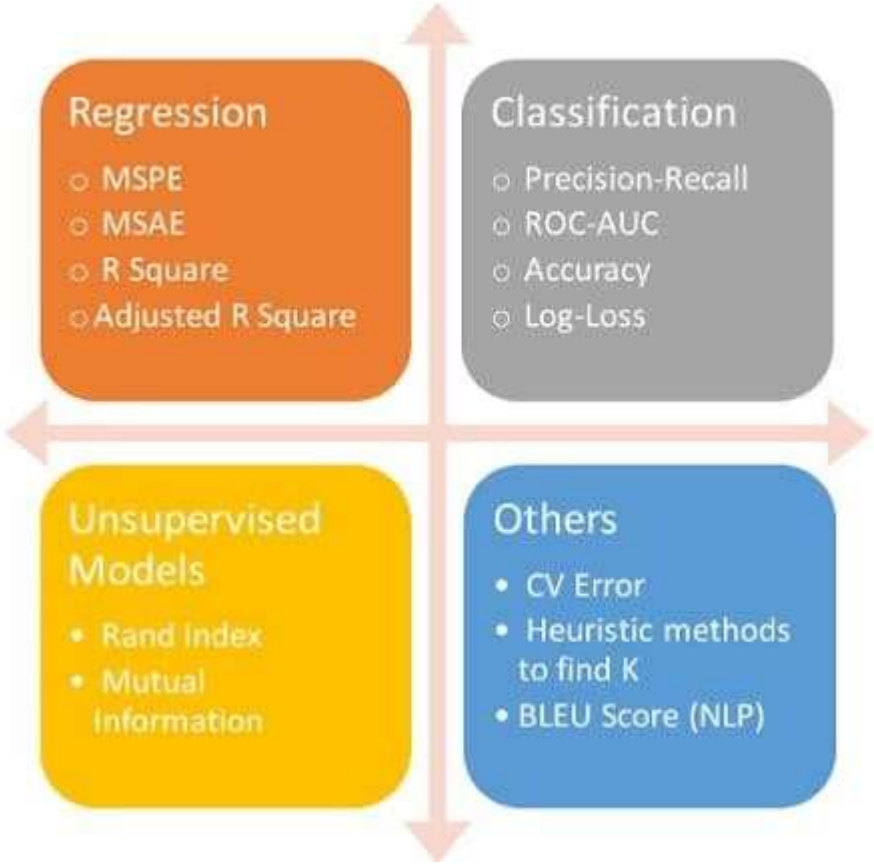
# EVALUATING

- Evaluating a model is a core part of building an effective machine learning modelThere are several evaluation metrics in machine learning, like confusion matrix, cross-validation, AUC-ROC curve, etc.

- Different evaluation metrics for logistic regression are used for different kinds of problems. Root Mean Squared Error (RMSE) RMSE is the most popular evaluation metric used in regression problems.

- Classification Metrics Walkthrough: Logistic Regression with Accuracy, Precision, Recall, and ROC. In this article, I will be going through 4 common classification metrics: Accuracy, Precision, Recall, and ROC in relation to Logistic Regression

# EVALUATING


Regression
MSPE
MSAE
R Square
Adjusted R Square
Classification
Precision-Recall
ROC-AUC
Accuracy
Log-Loss
Unsupervised Models
Rand Index
Mutual Information
Others
CV Error
Heuristic methods to find K
BLEU Score (NLP)

# Reference


- Tom M. Mitchell, “Machine Learning”, McGraw-Hill Education (India) Private Limited, 2013.
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, “An Introduction to Statistical Learning: with Applications in R”, Springer; First Edition 2013.
- P. Flach, —Machine Learning: The art and science of algorithms that make sense of data, Cambridge University Press, 2012.

# THANK YOU