SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

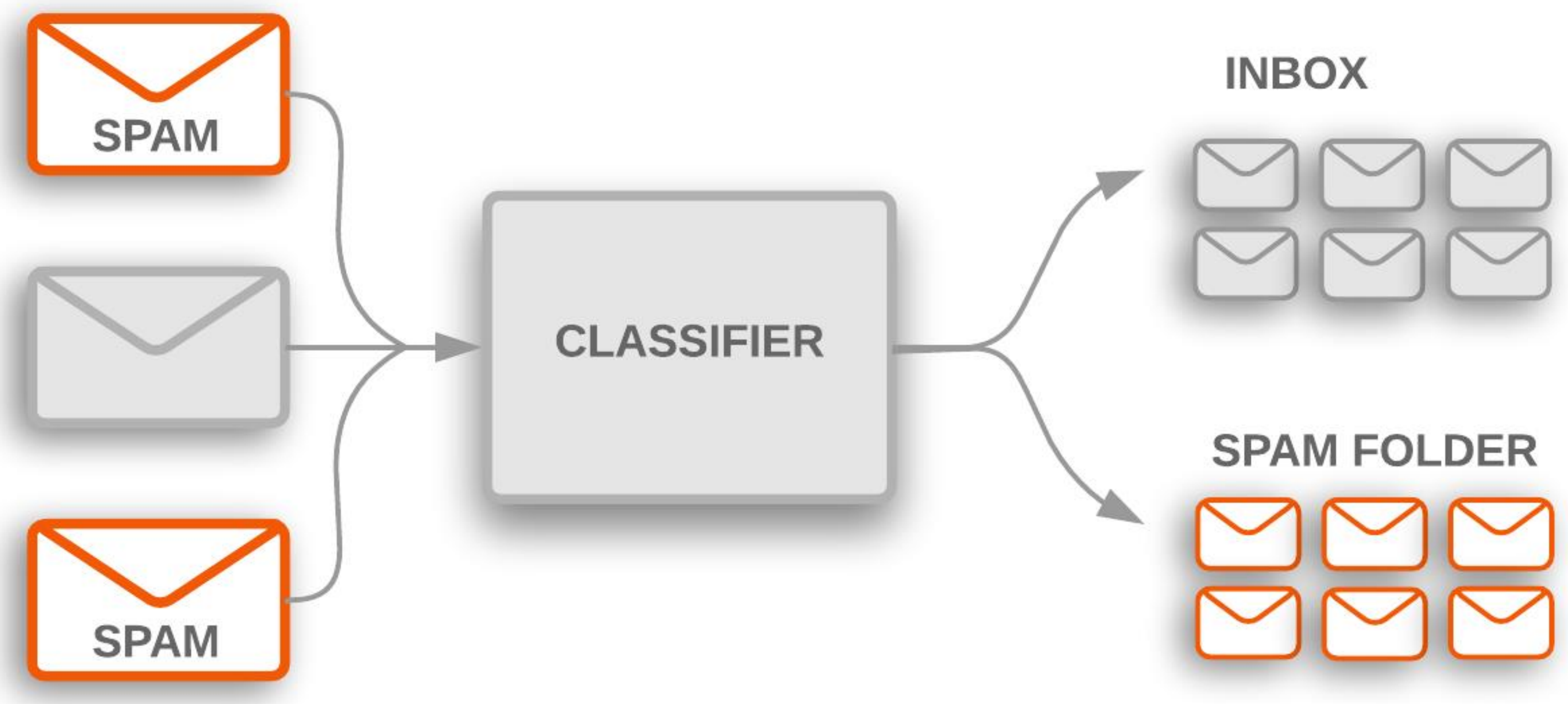
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE NAME : 19CS501 Introduction to Machine Learning

III YEAR /V SEMESTER

Unit 2- SUPERVISED LEARNING

Topic : Linear Models for Classification


SPAM
SPAM
CLASSIFIER
INBOX
SPAM FOLDER

What is Classification?

- We use the training dataset to get better boundary conditions which could be used to determine each target class. Once the boundary conditions are determined, the next task is to predict the target class. The whole process is known as classification.

- Target class examples:

  - Analysis of the customer data to predict whether he will buy computer accessories (Target class: Yes or No)
  - Classifying fruits from features like color, taste, size, weight (Target classes: Apple, Orange, Cherry, Banana)
  - Gender classification from hair length (Target classes: Male or Female)

Classification Algorithms vs Clustering Algorithms

- In clustering, the idea is not to predict the target class as in classification, it's more ever trying to group the similar kind of things by considering the most satisfied condition, all the items in the same group should be similar and no two different group items should not be similar.

- Group items Examples:

- While grouping similar language type documents (Same language documents are one group.)
- While categorizing the news articles (Same news category(Sport) articles are one group)

Basic Terminology in Classification Algorithms

- **Classifier**: An algorithm that maps the input data to a specific category.
- **Classification model**: A classification model tries to draw some conclusion from the input values given for training. It will predict the class labels/categories for the new data.
- **Feature**: A feature is an individual measurable property of a phenomenon being observed.
- **Binary Classification**: Classification task with two possible outcomes. Eg: Gender classification (Male / Female)
- **Multi-class classification**: Classification with more than two classes. In multi-class classification, each sample is assigned to one and only one target label. Eg: An animal can be a cat or dog but not both at the same time.
- **Multi-label classification**: Classification task where each sample is mapped to a set of target labels (more than one class). Eg: A news article can be about sports, a person, and location at the same time.

Applications of Classification Algorithms

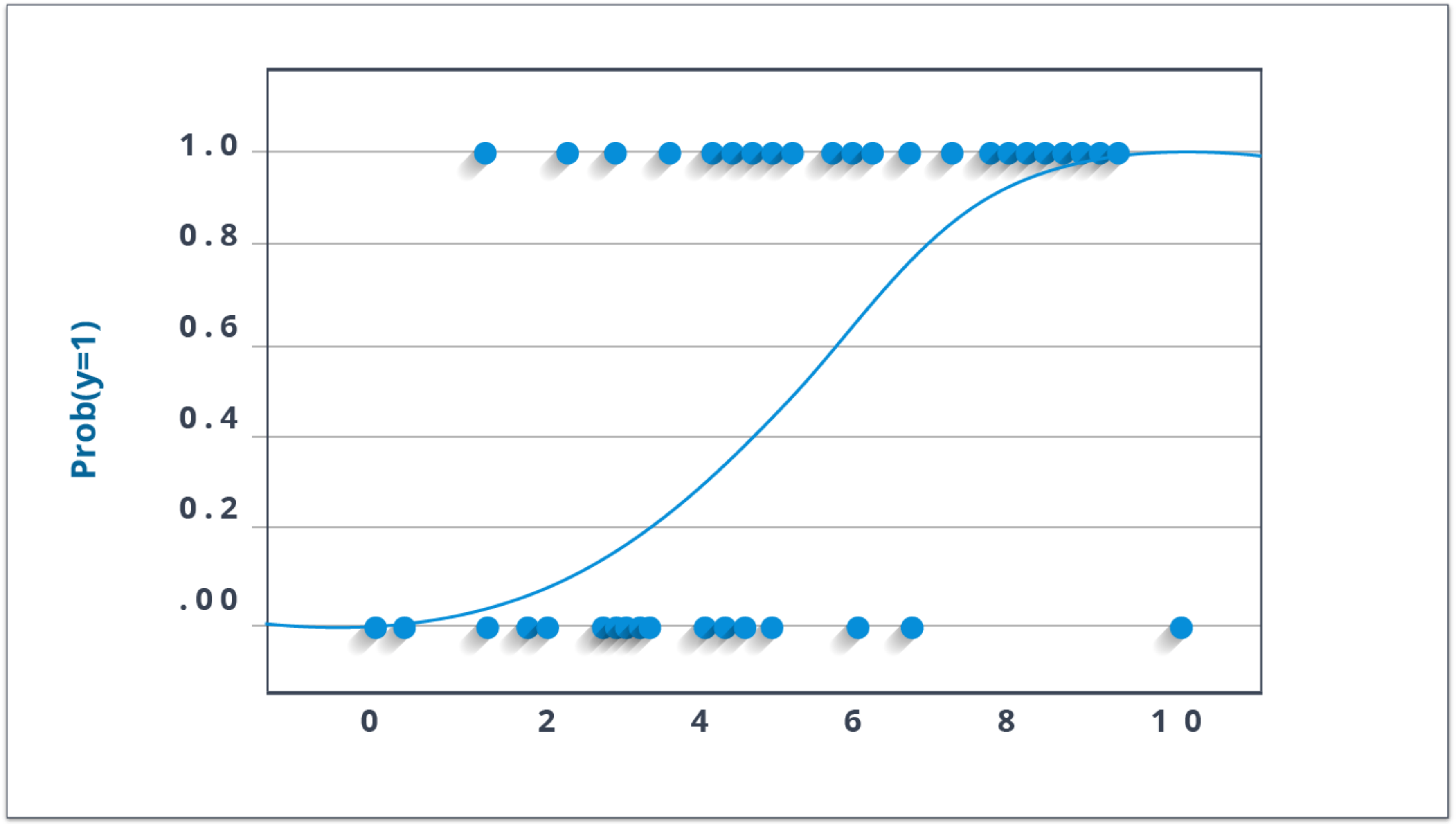
- Email spam classification
- Bank customers loan pay willingness prediction.
- Cancer tumor cells identification.
- Sentiment analysis
- Drugs classification
- Facial key points detection
- Pedestrians detection in an automotive car driving.

Types of Classification Algorithms

- ***Linear Classifiers***
  - Logistic regression
  - Naive Bayes classifier
  - Fisher’s linear discriminant
- ***Support vector machines***
  - Least squares support vector machines
- ***Quadratic classifiers***
- ***Kernel estimation***
  - k-nearest neighbor
- ***Decision trees***
  - Random forests
- ***Neural networks***
- ***Learning vector quantization***

Logistic Regression

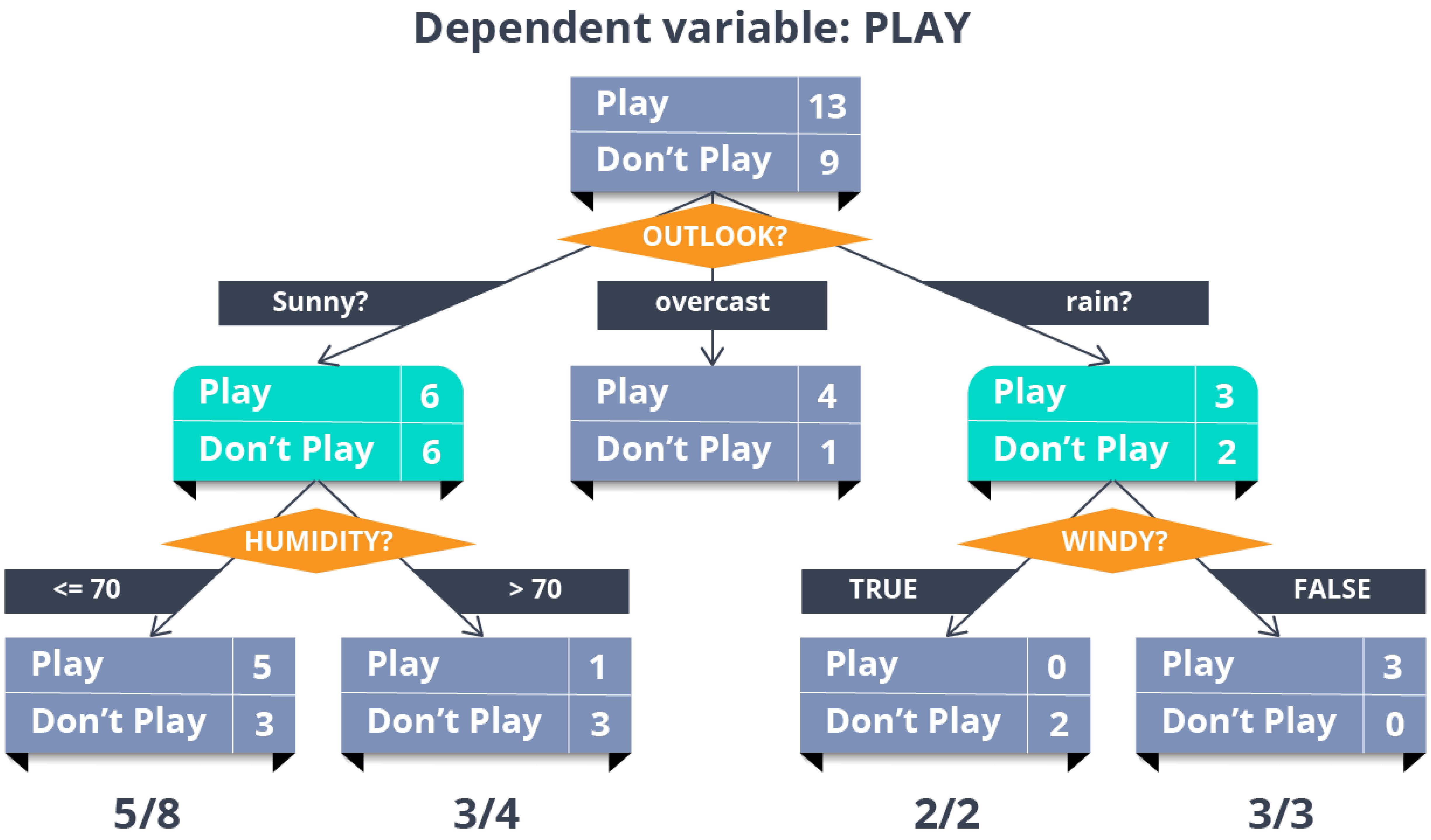
- As confusing as the name might be, you can rest assured. Logistic Regression is a classification and not a regression algorithm. It estimates discrete values (Binary values like 0/1, yes/no, true/false) based on a given set of independent variable(s).
- Simply put, it basically, predicts the probability of occurrence of an event by fitting data to a logit function. Hence, it is also known as logit regression. The values obtained would always lie within 0 and 1 since it predicts the probability.



Decision Trees

- Now, the decision tree is by far, one of my favorite algorithms. With versatile features helping actualize both categorical and continuous dependent variables, it is a type of supervised learning algorithm mostly used for classification problems.

- What this algorithm does is, it splits the population into two or more homogeneous sets based on the most significant attributes making the groups as distinct as possible.



Naive Bayes Classifier

- This is a classification technique based on an assumption of independence between predictors or what's known as Bayes' theorem. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

- For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, a Naive Bayes Classifier would consider all of these properties to independently contribute to the probability that this fruit is an apple

Likelihood

Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)\,P(c)}{P(x)}$$

Posterior Probability

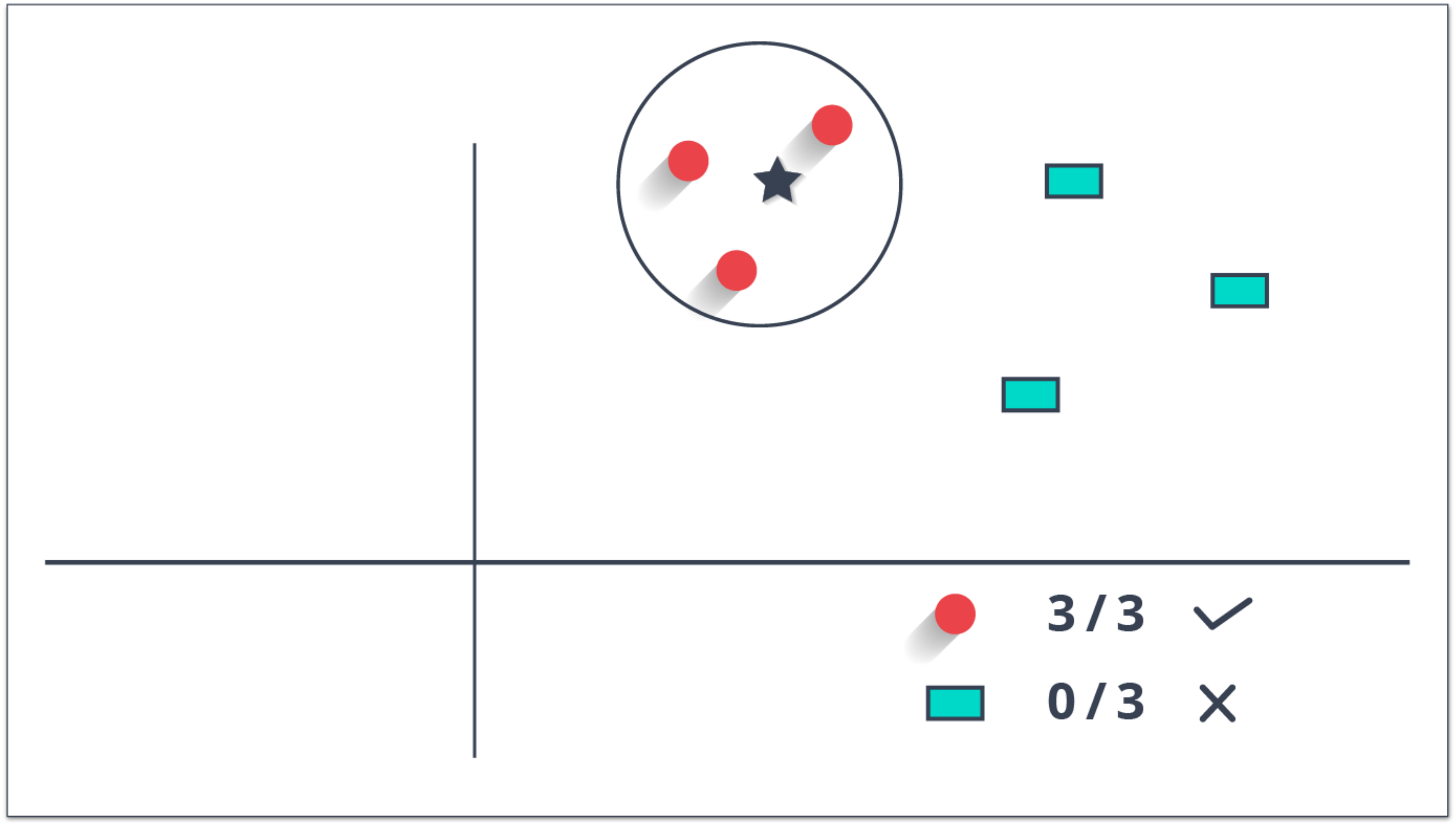
Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

KNN (k- Nearest Neighbors)

K nearest neighbors is a simple algorithm used for both classification and regression problems. It basically stores all available cases to classify the new cases by a majority vote of its k neighbors. The case assigned to the class is most common amongst its K nearest neighbors measured by a distance function (Euclidean, Manhattan, Minkowski, and Hamming).

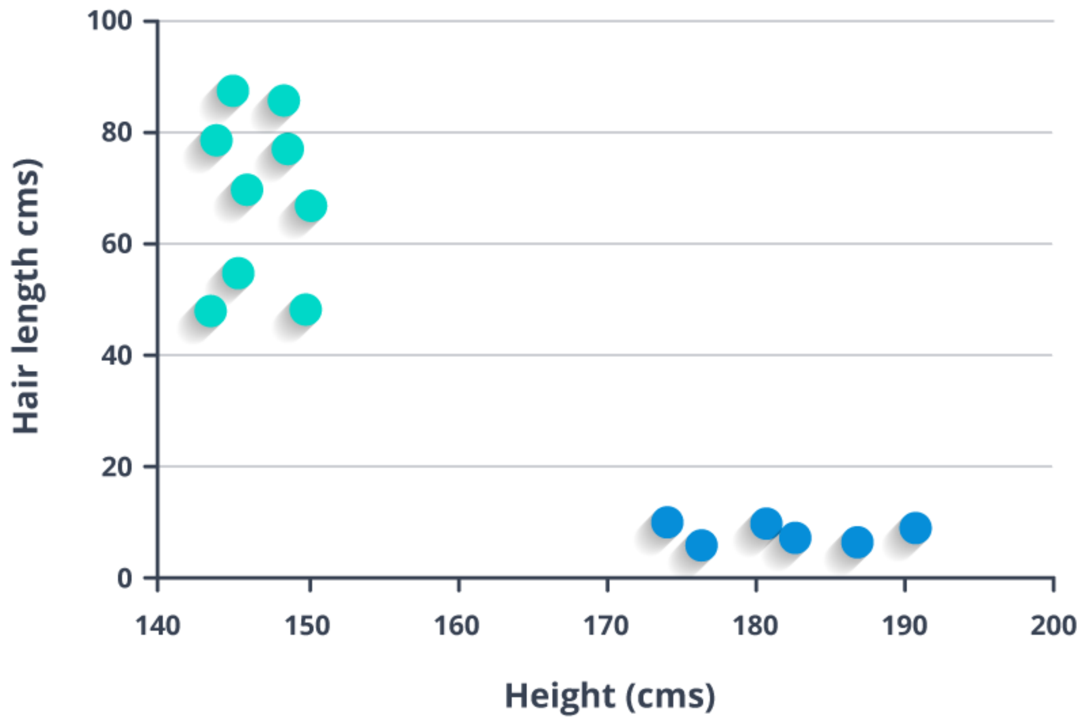
While the three former distance functions are used for continuous variables, Hamming distance function is used for categorical variables. If K = 1, then the case is simply assigned to the class of its nearest neighbor. At times, choosing K turns out to be a challenge while performing kNN modeling.

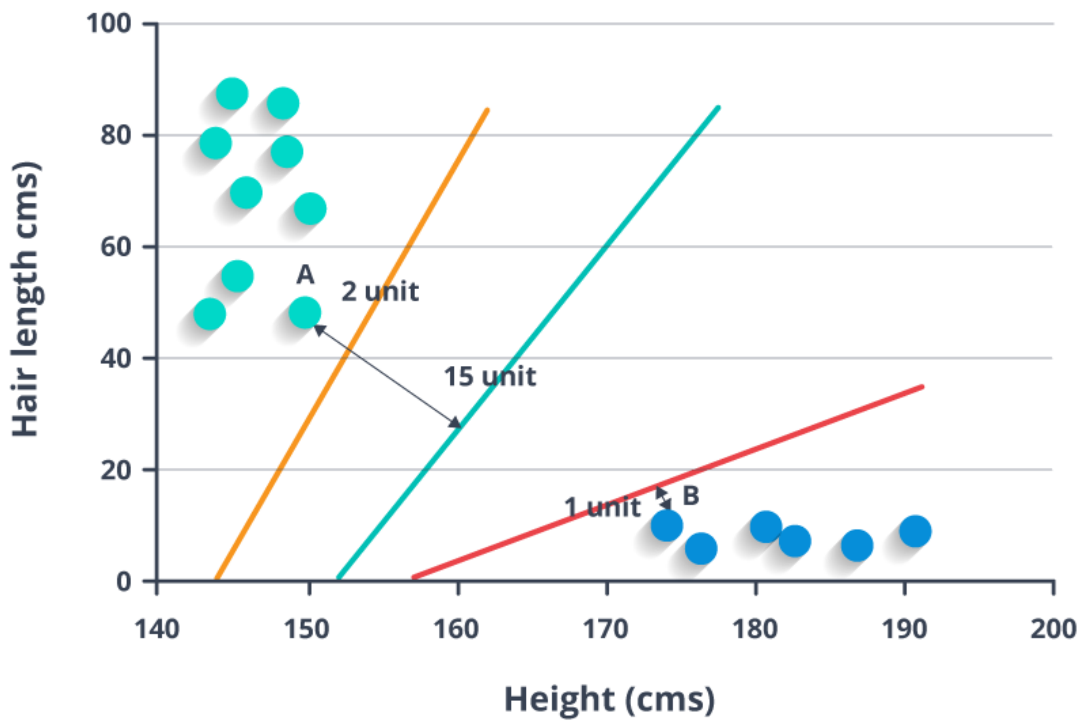


SVM(Support Vector Machine)

In this algorithm, we plot each data item as a point in n-dimensional space (where n is a number of features you have) with the value of each feature being the value of a particular coordinate.

For example, if we only had two features like Height and Hair length of an individual, we'd first plot these two variables in two-dimensional space where each point has two coordinates (these coordinates are known as Support Vectors)





4 Types of Classification Tasks in Machine Learning

- Classification Predictive Modeling
- Binary Classification
- Multi-Class Classification
- Multi-Label Classification
- Imbalanced Classification

Classification Predictive Modeling

- In machine learning, classification refers to a predictive modeling problem where a class label is predicted for a given example of input data.

- Examples of classification problems include:

  - ✓Given an example, classify if it is spam or not.
  - ✓Given a handwritten character, classify it as one of the known characters.
  - ✓Given recent user behavior, classify as churn or not.

A model will use the training dataset and will calculate how to best map examples of input data to specific class labels. As such, the training dataset must be sufficiently representative of the problem and have many examples of each class label.

Class labels are often string values, e.g. "spam," "not spam," and must be mapped to numeric values before being provided to an algorithm for modeling. This is often referred to as label encoding, where a unique integer is assigned to each class label, e.g. "spam" = 0, "no spam" = 1.

Binary Classification

- Binary classification refers to those classification tasks that have two class labels.

- Examples include:

  - ✓Email spam detection (spam or not).
  - ✓Churn prediction (churn or not).
  - ✓Conversion prediction (buy or not).
  - ✓Typically, binary classification tasks involve one class that is the normal state and another class that is the abnormal state.

- For example “not spam” is the normal state and “spam” is the abnormal state. Another example is “cancer not detected” is the normal state of a task that involves a medical test and “cancer detected” is the abnormal state.

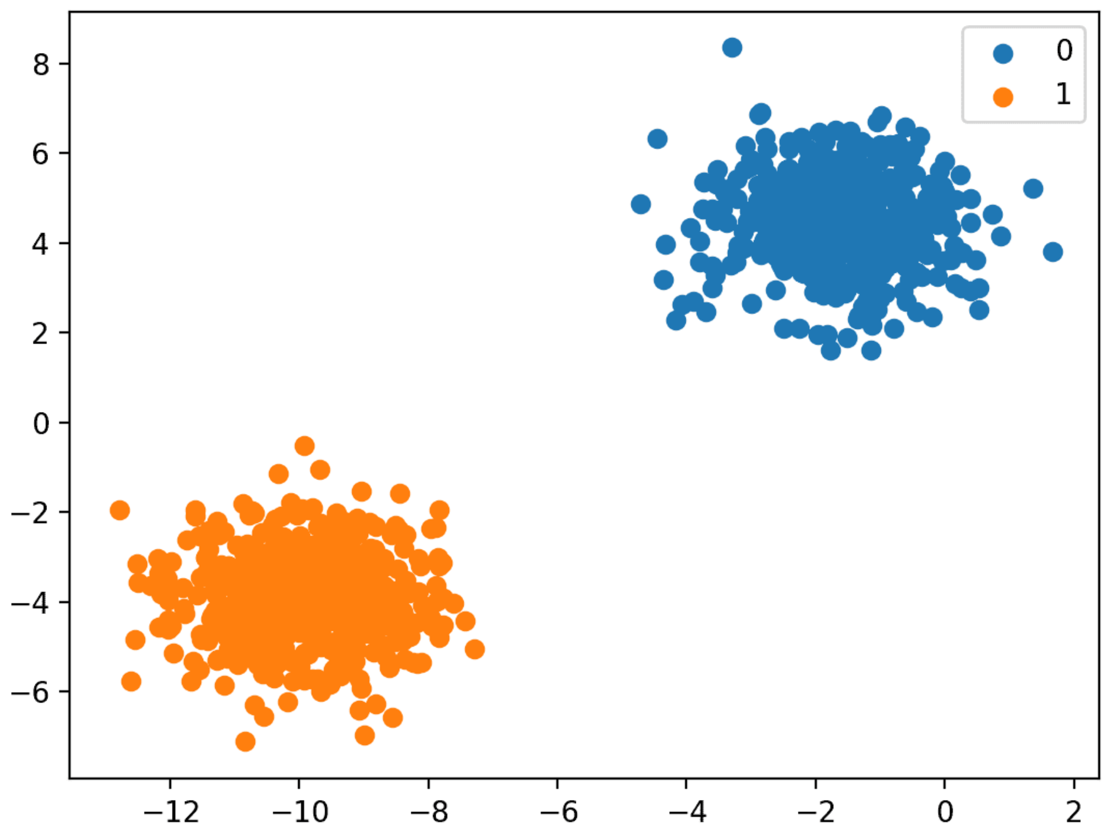
- The class for the normal state is assigned the class label 0 and the class with the abnormal state is assigned the class label 1.

Binary Classification

- ➢ It is common to model a binary classification task with a model that predicts a Bernoulli probability distribution for each example.

- ➢ The Bernoulli distribution is a discrete probability distribution that covers a case where an event will have a binary outcome as either a 0 or 1. For classification, this means that the model predicts a probability of an example belonging to class 1, or the abnormal state.

- ➢ Popular algorithms that can be used for binary classification include:
  - ✓Logistic Regression
  - ✓k-Nearest Neighbors
  - ✓Decision Trees
  - ✓Support Vector Machine
  - ✓Naive Bayes



Multi-Class Classification

- Multi-class classification refers to those classification tasks that have more than two class labels.

- Examples include:
  - ✓Face classification.
  - ✓Plant species classification.
  - ✓Optical character recognition.
- Unlike binary classification, multi-class classification does not have the notion of normal and abnormal outcomes. Instead, examples are classified as belonging to one among a range of known classes.

- The number of class labels may be very large on some problems. For example, a model may predict a photo as belonging to one among thousands or tens of thousands of faces in a face recognition system.

Multi-Class Classification

- It is common to model a multi-class classification task with a model that predicts a Multinoulli probability distribution for each example.
- The Multinoulli distribution is a discrete probability distribution that covers a case where an event will have a categorical outcome, e.g. K in {1, 2, 3, …, K}. For classification, this means that the model predicts the probability of an example belonging to each class label.
- Many algorithms used for binary classification can be used for multi-class classification.
- Popular algorithms that can be used for multi-class classification include:
- k-Nearest Neighbors.
- Decision Trees.
- Naive Bayes.
- Random Forest.
- Gradient Boosting.

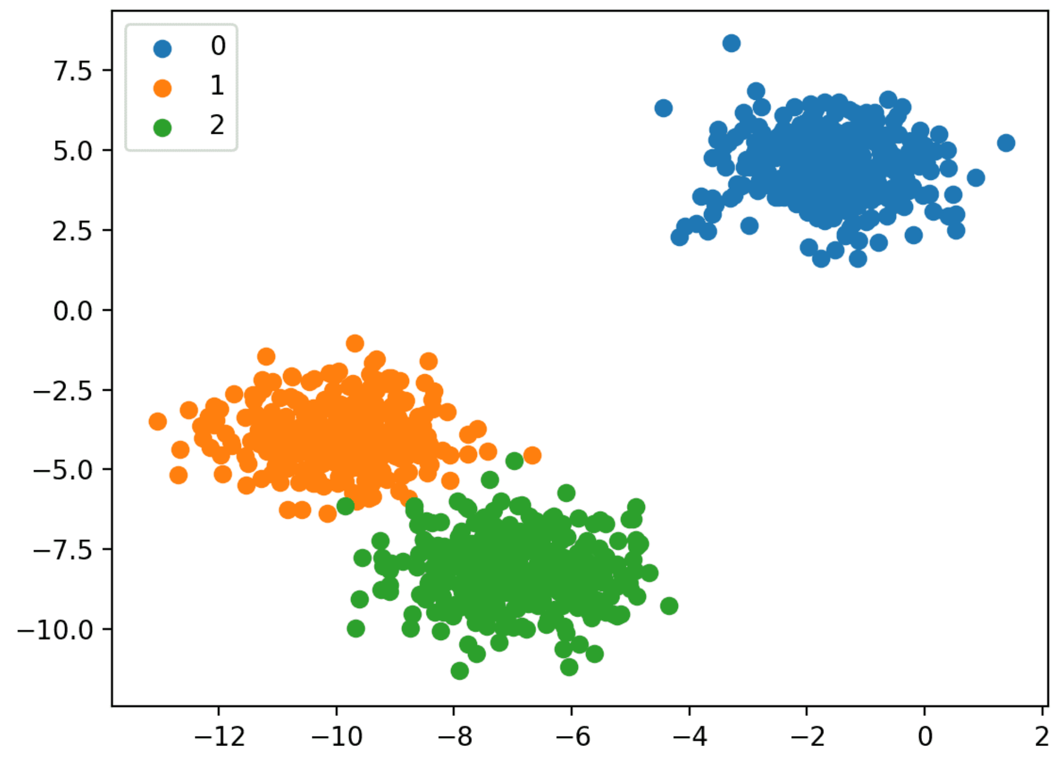
Multi-Class Classification

- Algorithms that are designed for binary classification can be adapted for use for multi-class problems.

- This involves using a strategy of fitting multiple binary classification models for each class vs. all other classes (called one-vs-rest) or one model for each pair of classes (called one-vs-one).

- One-vs-Rest: Fit one binary classification model for each class vs. all other classes.
- One-vs-One: Fit one binary classification model for each pair of classes.
- Binary classification algorithms that can use these strategies for multi-class classification include:

- Logistic Regression.
- Support Vector Machine.



Multi-Label Classification

- Multi-label classification refers to those classification tasks that have two or more class labels, where one or more class labels may be predicted for each example.
- Consider the example of photo classification, where a given photo may have multiple objects in the scene and a model may predict the presence of multiple known objects in the photo, such as “bicycle,” “apple,” “person,” etc.
- This is unlike binary classification and multi-class classification, where a single class label is predicted for each example.
- It is common to model multi-label classification tasks with a model that predicts multiple outputs, with each output taking predicted as a Bernoulli probability distribution. This is essentially a model that makes multiple binary classification predictions for each example

Multi-Label Classification

- Classification algorithms used for binary or multi-class classification cannot be used directly for multi-label classification. Specialized versions of standard classification algorithms can be used, so-called multi-label versions of the algorithms, including:

  - ✓ Multi-label Decision Trees
  - ✓ Multi-label Random Forests
  - ✓ Multi-label Gradient Boosting
  - ✓ Another approach is to use a separate classification algorithm to predict the labels for each class.

Imbalanced Classification

- Imbalanced classification refers to classification tasks where the number of examples in each class is unequally distributed.

- Typically, imbalanced classification tasks are binary classification tasks where the majority of examples in the training dataset belong to the normal class and a minority of examples belong to the abnormal class.

- Examples include:

  - ✓Fraud detection.
  - ✓Outlier detection.
  - ✓Medical diagnostic tests.
- These problems are modeled as binary classification tasks, although may require specialized techniques.
- Specialized techniques may be used to change the composition of samples in the training dataset by undersampling the majority class or oversampling the minority class.
- Examples include:

  - ✓Random Undersampling.
  - ✓SMOTE Oversampling.

Imbalanced Classification

- Specialized modeling algorithms may be used that pay more attention to the minority class when fitting the model on the training dataset, such as cost-sensitive machine learning algorithms.

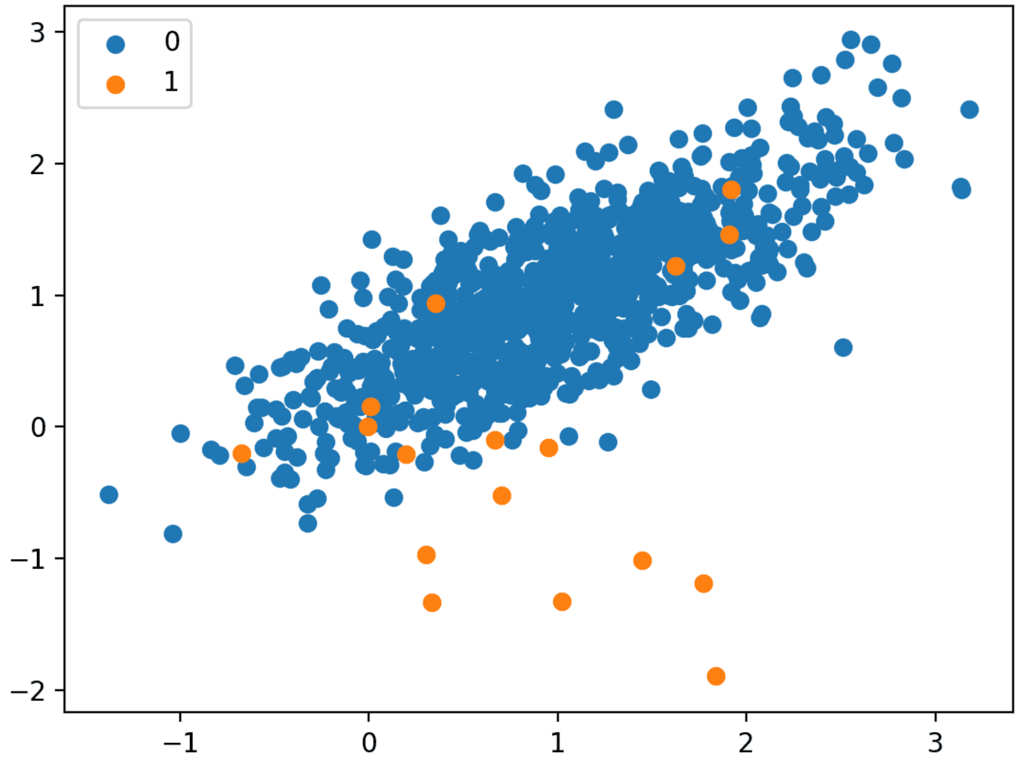
- Examples include:

  - ✓Cost-sensitive Logistic Regression.
  - ✓Cost-sensitive Decision Trees.
  - ✓Cost-sensitive Support Vector Machines.

- Finally, alternative performance metrics may be required as reporting the classification accuracy may be misleading.

- Examples include:

  - ✓Precision.
  - ✓Recall.
  - ✓F-Measure.



Assessment

- A dataset that requires a numerical prediction is a regression problem.
- An algorithm that is fit on a regression dataset is a regression algorithm.
- A model fit using a regression algorithm is a regression model.

REFERENCES

1. Tom M. Mitchell, "Machine Learning", McGraw-Hill Education (India) Private Limited, 2013.
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Springer; Second Edition, 2009.

THANK YOU