**Pattern Recognition**

**Pattern** is everything around in this digital world. A pattern can either be seen physically or it can be observed mathematically by applying algorithms.  
**Example:** Mathematics: AP, GP Series. Biology: DNA, RNA. Chemistry: Elements showing a set of specific properties (S, P, D, F Blocks). The colours on the clothes, speech pattern etc. In computer science, a pattern is represented using vector features values.

**What is Pattern Recognition ?**

**Pattern recognition** is the process of recognizing patterns by using machine learning algorithm. Pattern recognition can be defined as the classification of **data based** on **knowledge** already gained or on **statistical information extracted** from **patterns** and/or their **representation**. One of the important aspects of the pattern recognition is its application potential.

**Examples:** Speech recognition, speaker identification, multimedia document recognition (MDR), automatic medical diagnosis.  
In a typical pattern recognition application, the raw data is processed and converted into a form that is amenable for a machine to use. Pattern recognition involves classification and cluster of patterns.

* In classification, an appropriate class label is assigned to a pattern based on an abstraction that is generated using a set of training patterns or domain knowledge. Classification is used in supervised learning.
* **Classification** uses predefined classes in which objects are assigned
* Clustering generated a partition of the data which helps decision making, the specific decision making activity of interest to us. Clustering is used in an unsupervised learning.
* **Clustering** identifies similarities **between** objects, which it groups according to those characteristics in common and which **differentiate** them from other.

**Features** may be represented as continuous, discrete or discrete binary variables. A feature is a function of one or more measurements, computed so that it quantifies some significant characteristics of the object.  
**Example:** consider our face then eyes, ears, nose etc are features of the face.

A set of features that are taken together, forms the **features vector**.

**Pattern recognition possesses the following features:**

* Pattern recognition system should recognise familiar pattern quickly and accurate
* Recognize and classify unfamiliar objects
* Accurately recognize shapes and objects from different angles
* Identify patterns and objects even when partly hidden
* Recognise patterns quickly with ease, and with automaticity.

**Training and Learning in Pattern Recognition**

**Learning** is a phenomena through which a system gets trained and becomes adaptable to give result in an accurate manner. Learning is the most important phase as how well the system performs on the data provided to the system depends on which algorithms used on the data. Entire dataset is divided into two categories, one which is used in training the model i.e. Training set and the other that is used in testing the model after training, i.e. Testing set.

* **Training set:**

Training set is used to build a model. It consists of the set of images which are used to train the system. Training rules and algorithms used give relevant information on how to associate input data with output decision. The system is trained by applying these algorithms on the dataset, all the relevant information is extracted from the data and results are obtained. Generally, 80% of the data of the dataset is taken for training data.

* **Testing set:**

Testing data is used to test the system. It is the set of data which is used to verify whether the system is producing the correct output after being trained or not. Generally, 20% of the data of the dataset is used for testing. Testing data is used to measure the accuracy of the system. Example: a system which identifies which category a particular flower belongs to, is able to identify seven category of flowers correctly out of ten and rest others wrong, then the accuracy is 70 %



A pattern is a physical object or an abstract notion. While talking about the classes of animals, a description of an animal would be a pattern. While talking about various types of balls, then a description of a ball is a pattern. In the case balls considered as pattern, the classes could be football, cricket ball, table tennis ball etc. Given a new pattern, the class of the pattern is to be determined. The choice of attributes and representation of patterns is a very important step in pattern classification. A good representation is one which makes use of discriminating attributes and also reduces the computational burden in pattern classification.  
An obvious representation of a pattern will be a **vector**. Each element of the vector can represent one attribute of the pattern. The first element of the vector will contain the value of the first attribute for the pattern being considered.

**Advantages:**

* Pattern recognition solves classification problems
* Pattern recognition solves the problem of fake bio metric detection.
* It is useful for cloth pattern recognition for visually impaired blind people.
* It helps in speaker diarization.
* We can recognise particular object from different angle.

**Disadvantages:**

* Syntactic Pattern recognition approach is complex to implement and it is very slow process.
* Sometime to get better accuracy, larger dataset is required.
* It cannot explain why a particular object is recognized.

**Applications:**

* **Image processing, segmentation and analysis**  
  Pattern recognition is used to give human recognition intelligence to machine which is required in image processing.
* **Computer vision**  
  Pattern recognition is used to extract meaningful features from given image/video samples and is used in computer vision for various applications like biological and biomedical imaging.
* **Seismic analysis**  
  Pattern recognition approach is used for the discovery, imaging and interpretation of temporal patterns in seismic array recordings. Statistical pattern recognition is implemented and used in different types of seismic analysis models.

(Analyzing Building structures during Earthquake).

* **Radar signal classification/analysis**  
  Pattern recognition and Signal processing methods are used in various applications of radar signal classifications like AP mine detection and identification.
* **Speech recognition**  
  The greatest success in speech recognition has been obtained using pattern recognition paradigms. It is used in various algorithms of speech recognition which tries to avoid the problems of using a phoneme level of description and treats larger units such as words as pattern
* **Finger print identification**  
  The fingerprint recognition technique is a dominant technology in the biometric market. A number of recognition methods have been used to perform fingerprint matching out of which pattern recognition approaches is widely used.

**Basics and Design Principles**

**Pattern Recognition System**  
Pattern is everything around in this digital world. A pattern can either be seen physically or it can be observed mathematically by applying algorithms.

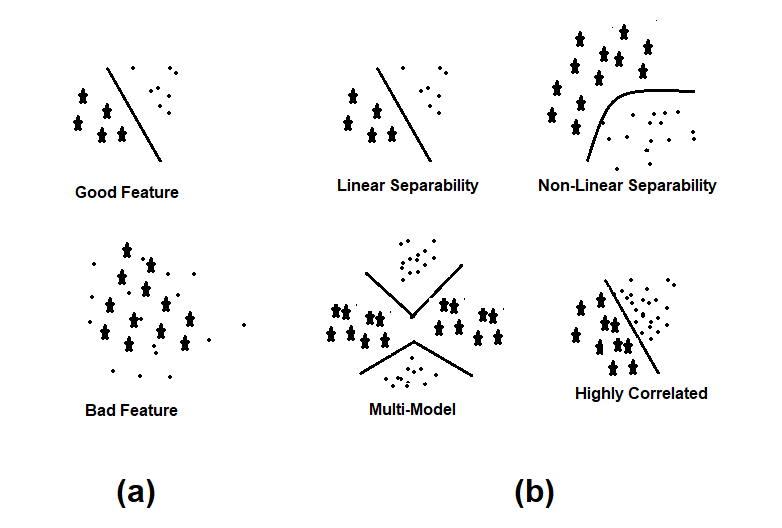
In **Pattern Recognition**, pattern is comprises of the following two fundamental things:

* Collection of observations
* The concept behind the observation

**Feature Vector:**

The collection of observations is also known as a feature vector. A feature is a distinctive characteristic of a good or service that sets it apart from similar items. **Feature vector** is the combination of n features in n-dimensional column vector. The different classes may have different features values but the same class always has the same features values.

**Example:**

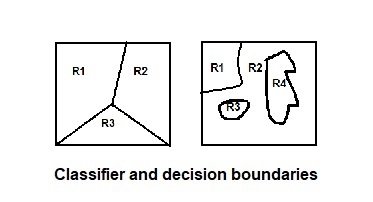


1. Differentiate between good and bad features.
2. Feature properties.

**A metric space is a set where a distance(called a metric) is defined b/w elements of set.**

**Classifier and Decision Boundaries:**

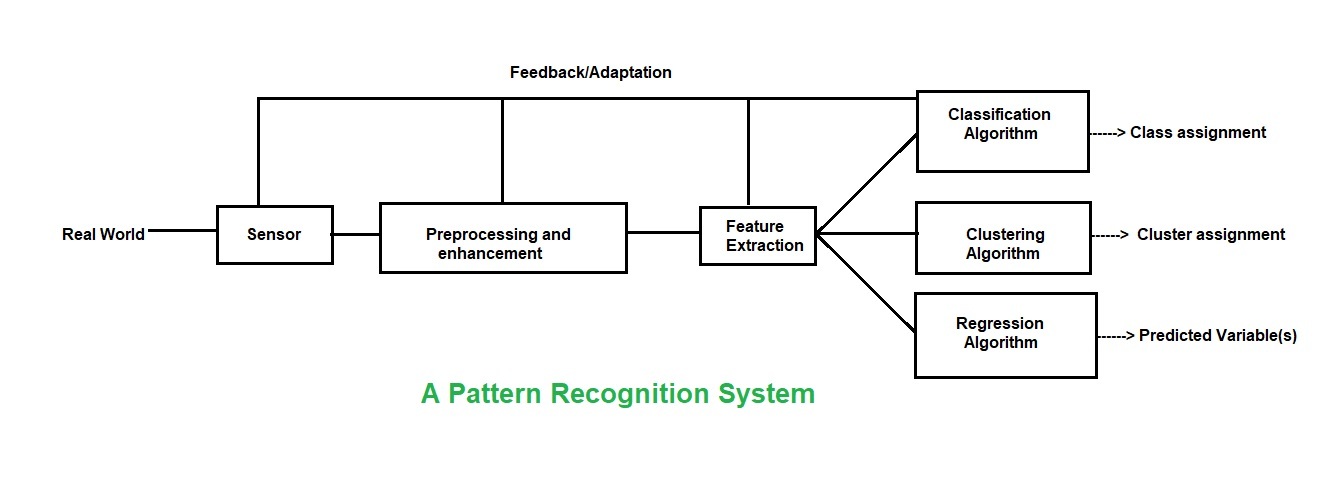
* In a statistical-classification problem, a **decision boundary** is a hypersurface that partitions the underlying vector space into two sets. A decision boundary is the region of a problem space in which the output label of a classifier is ambiguous. **Classifier** is a hypothesis or discrete-valued function that is used to assign (categorical) class labels to particular data points.
* **Classifier** is used to partition the feature space into class-labeled decision regions. While **Decision Boundaries** are the borders between decision regions.



Positive and Negative Prediction

**Components in Pattern Recognition System:**  
A pattern recognition systems can be partitioned into components. There are five typical components for various pattern recognition systems. These are as following:

* **A Sensor :** A sensor is a device used to measure a property, such as pressure, position, temperature, or acceleration, and respond with feedback.
* **A Preprocessing Mechanism :** Segmentation is used and it is the process of partitioning a data into multiple segments. It can also be defined as the technique of dividing or partitioning an data into parts called segments.
* **A Feature Extraction Mechanism :** feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. It can be manual or automated.
* **A Description Algorithm :** Pattern recognition algorithms generally aim to provide a reasonable answer for all possible inputs and to perform “most likely” matching of the inputs, taking into account their statistical variation
* **A Training Set :** Training data is a certain percentage of an overall dataset along with testing set. As a rule, the better the training data, the better the algorithm or classifier performs.



**Design Principles of Pattern Recognition**

In pattern recognition system, for recognizing the pattern or structure two basic approaches are used which can be implemented in different techniques. These are –

* Statistical Approach and
* Structural Approach

**Statistical Approach:**

Statistical methods are mathematical formulas, models, and techniques that are used in the statistical analysis of raw research data. The application of statistical methods extracts information from research data and provides different ways to assess the robustness of research outputs.

Two main statistical methods are used :

1. **Descriptive Statistics:** It summarizes data from a sample using indexes such as the mean or standard deviation.
2. **Inferential Statistics:** It draw conclusions from data that are subject to random variation.

**Structural Approach:**

The Structural Approach is a technique wherein the learner masters the pattern of sentence. Structures are the different arrangements of words in one accepted style or the other.

Types of structures:

* Sentence Patterns
* Phrase Patterns
* Formulas
* Idioms

**Difference Between Statistical Approach and Structural Approach:**

| **Sr. No.** | **Statistical Approach** | **Structural Approach** |
| --- | --- | --- |
| 1 | Statistical decision theory. | Human perception and cognition. |
| 2 | Quantitative features. | Morphological primitives |
| 3 | Fixed number of features. | Variable number of primitives. |
| 4 | Ignores feature relationships. | Captures primitives relationships. |
| 5 | Semantics from feature position. | Semantics from primitives encoding. |
| 6 | Statistical classifiers. | Syntactic grammars. |

# Supervised and Unsupervised learning

**Supervised learning**

Supervised learning as the name indicates the presence of a supervisor as a teacher. Basically supervised learning is a learning in which we teach or train the machine using data which is well labeled that means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples(data) so that supervised learning algorithm analyses the training data(set of training examples) and produces a correct outcome from labeled data.

Supervised learning classified into two categories of algorithms: 

* **Classification**: A classification problem is when the output variable is a category, such as “Red” or “blue” or “disease” and “no disease”.
* **Regression**: A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

Supervised learning deals with or learns with “labeled” data. Which implies that some data is already tagged with the correct answer.

**Types:-**

* Regression
* Logistic Regression
* Classification
* Naïve Bayes Classifiers
* Decision Trees
* Support Vector Machine

**Advantages:-**

* Supervised learning allows collecting data and produce  data output from the previous experiences.
* Helps to optimize performance criteria with the help of experience.
* Supervised machine learning helps to solve various types of real-world computation problems.

**Disadvantages:-**

* Classifying big data can be challenging.
* Training for supervised learning needs a lot of computation time. So, it requires a lot of time.

**Unsupervised learning**

Unsupervised learning is the training of machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of machine is to group unsorted information according to similarities, patterns and differences without any prior training of data.

Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore machine is restricted to find the hidden structure in unlabeled data by our-self.

Unsupervised learning classified into two categories of algorithms: 

* **Clustering**: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.
* **Association**: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Types of Unsupervised Learning:-

**Clustering**

1. Exclusive (partitioning)
2. Agglomerative
3. Overlapping
4. Probabilistic

**Clustering Types:-**

1. Hierarchical clustering
2. K-means clustering
3. K-NN (k nearest neighbors)
4. Principal Component Analysis
5. Singular Value Decomposition
6. Independent Component Analysis

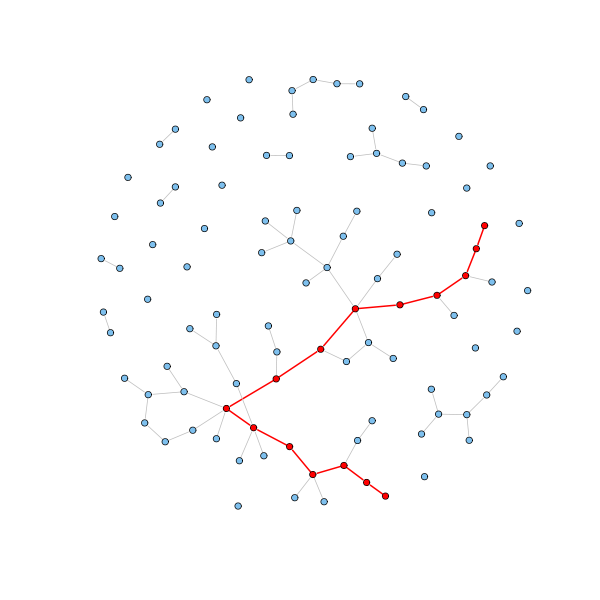
# Metric Space Method

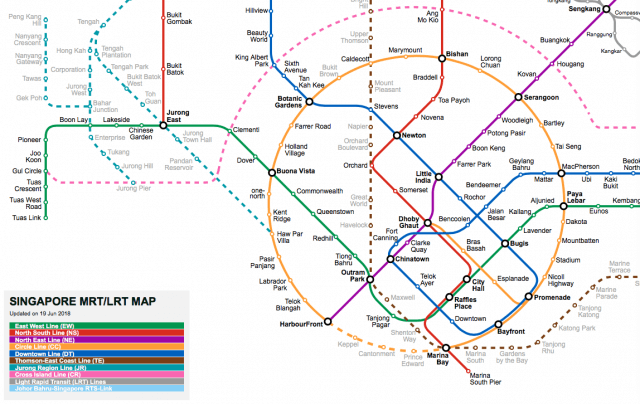
Mathematical Derivation and Problem has been already solved.

In mathematics, a metric space is a set where a distance (called a metric) is defined between elements of the set.  Metric space methods have been employed for decades in various applications, for example in internet search engines, image classification etc.

## What is a Distance?

The creation of a metric space requires the definition of a dissimilarity distance, which measures the dissimilarity between two different models.  The distance measure has two main requirements.  First, since the distance must be calculated between each model pair, it should be rapid to calculate for large ensembles of models. Second, the distance measure must be designed for the purpose of the study to be undertaken.  No single distance measure is applicable to all situations.  Finally, the distance measure must be easy to understand, in order to understand the results of the study.





**UNIT: 2**

# ****What is Classification?****

Classification is technique to categorize our data into a desired and distinct number of classes where we can assign label to each class.

In classification tasks, your job is to build a function that takes in a vector of **features X** (also called “inputs”) and predicts a **label** Y (also called the “class” or “output”). Features are things you know, and the label is what your algorithm is trying to figure out; for example, the label might be a binary variable indicating whether an animal is a cat or a dog, and the features might be the length of the animal’s whiskers, the animal’s weight in pounds, and a binary variable indicating whether the animal’s ears stick up or are droopy. Your algorithm needs to tell dogs and cats apart (Y) using only this information about weight, whiskers, and ears (**X**).

**Applications of Classification are:** speech recognition, handwriting recognition, biometric identification, document classification etc.

## Classifiers can be:

**Binary classifiers:** Classification with only 2 distinct classes or with 2 possible outcomes

example: Male and Female

example: classification of spam email and non spam email

example: classification of author of book

example: positive and negative sentiment

**Multi-Class classifiers**: Classification with more than two distinct classes.

example: classification of types of soil

example: classification of types of crops

example: classification of mood/feelings in songs/music

# 1). Naive Bayes (Classifier):

Naive Bayes is a probabilistic classifier inspired by the Bayes theorem. Under a simple assumption which is the attributes are conditionally independent.

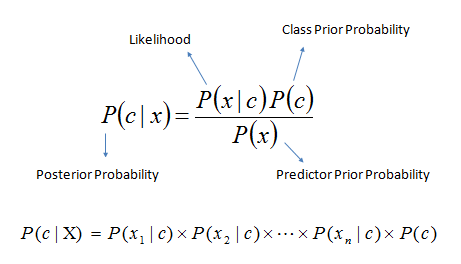


Fig: Naïve Bayes

The classification is conducted by deriving the maximum posterior which is the maximal P(Ci|**X**) with the above assumption applying to Bayes theorem. This assumption greatly reduces the computational cost by only counting the class distribution. Even though the assumption is not valid in most cases since the attributes are dependent, surprisingly Naive Bayes has able to perform impressively.

Naive Bayes is a very simple algorithm to implement and good results have obtained in most cases. It can be easily scalable to larger datasets since it takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.

Naive Bayes can suffer from a problem called the zero probability problem. When the conditional probability is zero for a particular attribute, it fails to give a valid prediction. This needs to be fixed explicitly using a Laplacian estimator.

**Advantages:** This algorithm requires a small amount of training data to estimate the necessary parameters. Naive Bayes classifiers are extremely fast compared to more sophisticated methods.

**Disadvantages:** Naive Bayes is is known to be a bad estimator.

## **Steps for Implementation:**

* Initialise the classifier to be used.
* Train the classifier: All classifiers in scikit-learn uses a fit(X, y) method to fit the model(training) for the given train data X and train label y.
* Predict the target: Given an non-label observation X, the predict(X) returns the predicted label y.
* Evaluate)\* the classifier model

Bayes Theorem: P(A/B) = P(B/A)\*P(A)

P(B)

Dataset: x

x ={x1, x2, x3…………… xn} = {y}

F1  F2 F3……… Y

x1 x2 x3 y1

P(y/x1,x2,x3,……..xn)= P(x1/y)\* P(x2/y)\*P(x3/y)\*………. P(xn/y) \* P(y)

P(x1)\* P(x2)\*P(x3)\*………. P(xn)

P(y/x1,x2,x3,……..xn) = P(y) \* πi =1nP(xi/y)

P(x1)\* P(x2)\*P(x3)\*………. P(xn)

P(x1)\* P(x2)\*P(x3)\*………. P(xn) ∝ P(y) \* πi =1nP(xi/y)

y = argmax P(y) πi =1nP(xi/y)

for yes = 0.7

no = 0.3

Example:

Finding probability that the player can play outside or not, Depending upon the weather.

Outlook:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Yes | No | P(Y) | P(N) |
| Sunny | 2 | 3 | 2/9 | 3/5 |
| Overcast | 4 | 0 | 4/9 | 0/5 |
| Rainy | 3 | 2 | 1/3 | 2/5 |
| Total | 9 | 5 | 100% | 100% |

Temperature:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Yes | No | P(Y) | P(N) |
| Hot | 2 | 2 | 2/9 | 2/5 |
| Mild | 4 | 2 | 4/9 | 2/5 |
| Cold | 3 | 1 | 1/3 | 1/5 |
| Total | 9 | 5 | 100% | 100% |

Using the above data we need to find that a player can play outside today or not?

Weather for today is sunny and hot.

|  |  |  |
| --- | --- | --- |
|  |  | P(Y) / P(N) |
| Yes | 9 | 9/14 |
| No | 5 | 5/14 |
| Total | 14 | 100% |

P(Y/Today) = P(OC/Yes) \* P(Cold/Yes) \* P(Yes)

P(Today)

P(Y/Today) ∝ P(OC/Yes) \* P(Cold/Yes) \* P(Yes)

= 4/9 \* 1/3 \* 9/14 = 0.0987

P(N/Today) ∝ P(OC/No) \* P(Cold/No) \* P(No)

= 0/5 \* 1/5 \* 5/14 = 0

Normalization

P(Yes) = 0.0987 = 1

0.0987 + 0

P(No) = 1 - P(Yes) = 0

# ****Random Forest Classifier:****

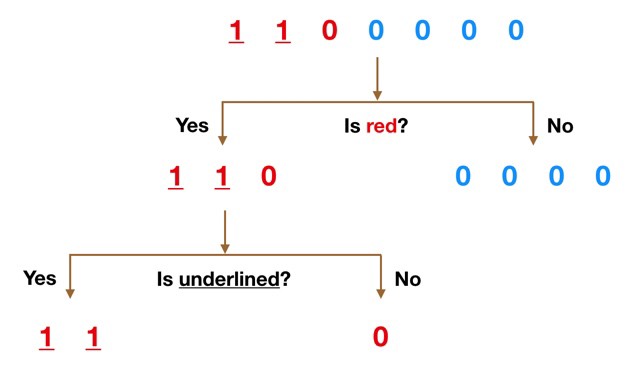
It is an ensemble tree-based learning algorithm. The Random Forest Classifier is a set of decision trees from randomly selected subset of training set. It **aggregates the votes from different decision trees** to decide the final class of the test object.

# ****Ensemble Algorithm :****

Ensemble algorithms are those which **combines more than one algorithms of same or different kind for classifying objects**. For example, running prediction over Naive Bayes, SVM and Decision Tree and then taking vote for final consideration of class for test object.

# Decision Trees

Decision trees are the building blocks of the random forest model. It’s probably much easier to understand how a decision tree works through an example.



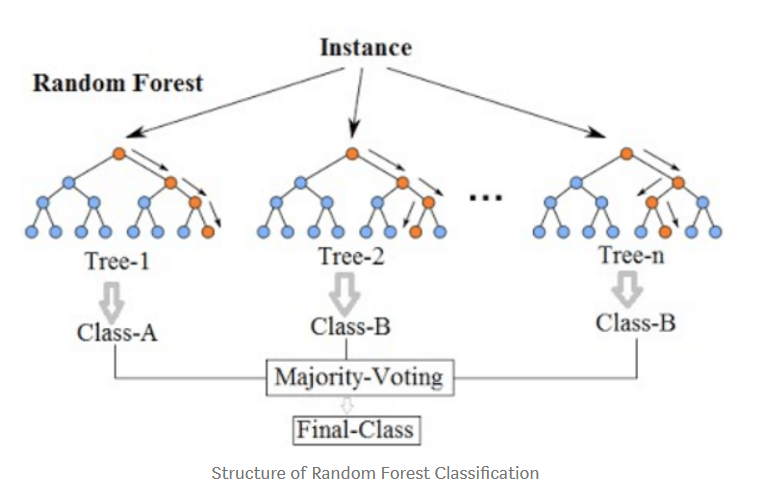
Simple Decision Tree Example

Imagine that our dataset consists of the numbers at the top of the figure to the left. We have two 1s and five 0s (1s and 0s are our classes) and desire to separate the classes using their features. The features are color (red vs. blue) and whether the observation is underlined or not. So how can we do this?

Color seems like a pretty obvious feature to split by as all but one of the 0s are blue. So we can use the question, “Is it red?” to split our first node. You can think of a node in a tree as the point where the path splits into two — observations that meet the criteria go down the Yes branch and ones that don’t go down the No branch.

The No branch (the blues) is all 0s now so we are done there, but our Yes branch can still be split further. Now we can use the second feature and ask, “Is it underlined?” to make a second split.

The two 1s that are underlined go down the Yes subbranch and the 0 that is not underlined goes down the right subbranch and we are all done. Our decision tree was able to use the two features to split up the data perfectly.



# Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model’s prediction

# 

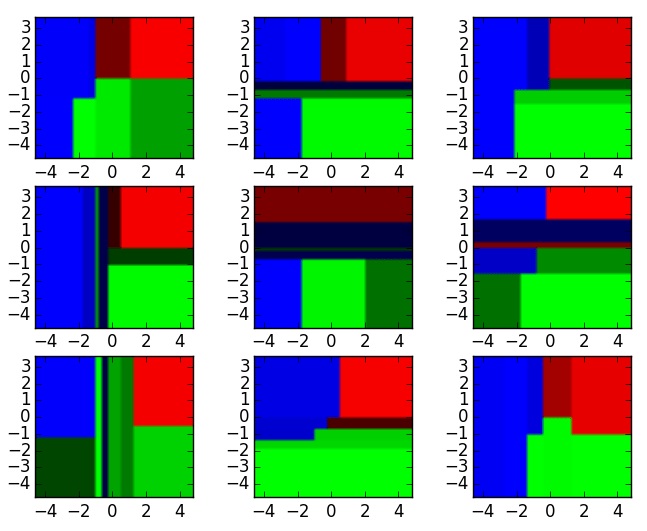
# ****Types of Random Forest models:****

1. Random Forest Prediction for a **classification problem**:

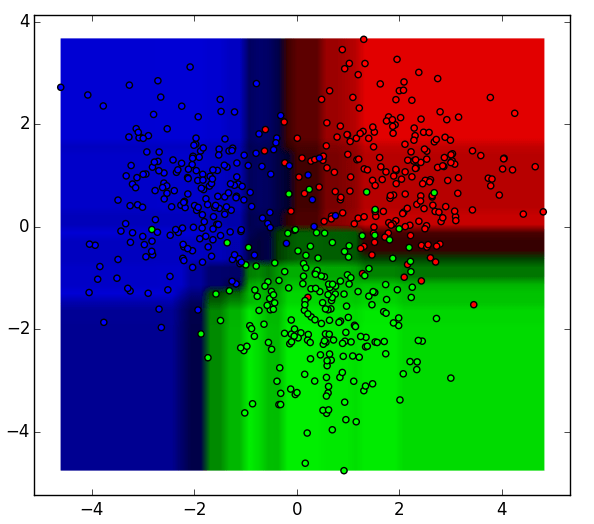
f(x) = majority vote of all predicted classes over B trees

2. Random Forest Prediction for a **regression problem**:

f(x) = sum of all sub-tree predictions divided over B trees



Nine Different Decision Tree Classifier



Aggregated Result based on above classifier

The 9 decision tree classifiers shown above can be aggregated into a random forest ensemble which **combines their input**. The horizontal and vertical axes of the above decision tree outputs can be thought of as features x1 and x2. At certain values of each feature, the decision tree outputs a classification of “blue”, “green”, “red”, etc.

These above **results are aggregated**, through **model votes or averaging**, into a single  
ensemble model that ends up outperforming any individual decision tree’s output.

**Features and Advantages of Random Forest :**

1. It is one of the most accurate learning algorithms available. For many data sets, it produces a **highly accurate classifier**.
2. It runs efficiently on large databases.
3. It can **handle thousands of input variables** without variable deletion.
4. It gives estimates of what variables that are important in the classification.
5. It generates an internal **unbiased estimate of the generalization error** as the forest building progresses.
6. It has an **effective method for estimating missing data** and maintains accuracy when a large proportion of the data are missing.

**Disadvantages of Random Forest:**

1. Random forests have been observed to **overfit for some datasets** with noisy classification/regression tasks.
2. For data including categorical variables with different number of levels, **random forests are biased in favor of those attributes with more levels**. Therefore, the variable importance scores from random forest are not reliable for this type of data.