

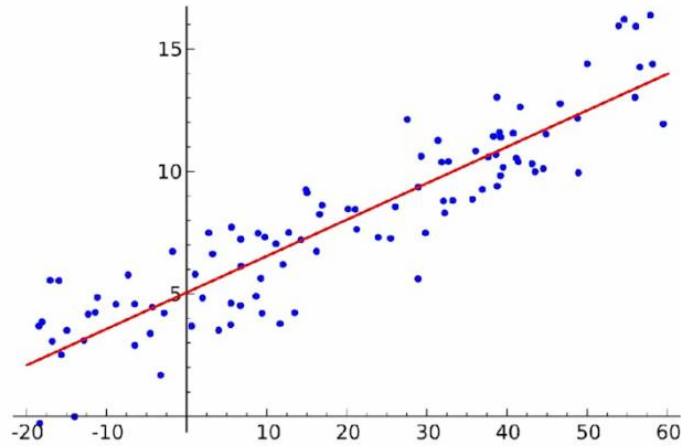
SUPERVISED:

A. TYPES OF REGRESSION MODELS:

1. **LINEAR REGRESSION:** Finding the line which fits the data
 - We find target value based on independent predictor
 - i.e., use to find relationship between dependent and independent variable
 - Output is continuous
 - Types of Regression:
 - # Multiple Linear Regression
 - Finding a plain for best fit
 - # Polynomial Regression
 - Finding a curve for best fit

Types of Regression Models

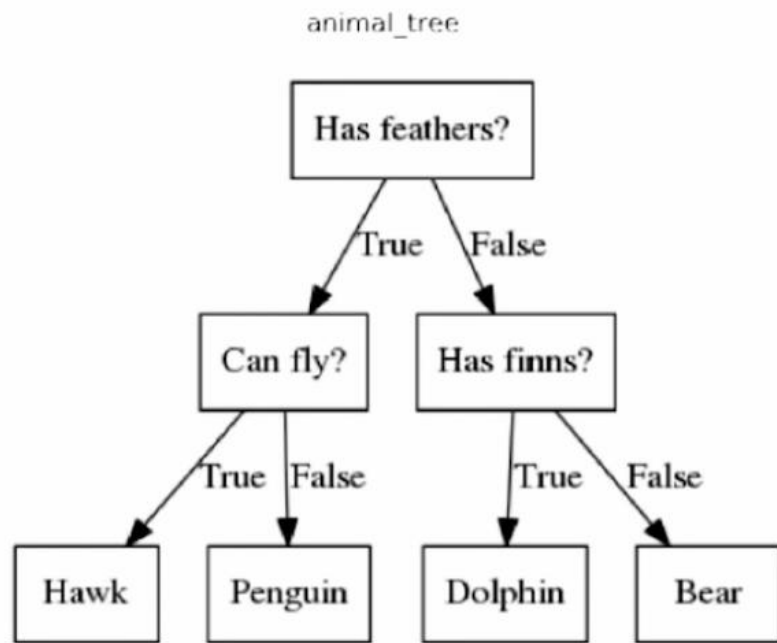
I. Linear Regression



- Multiple Linear
- Polynomial

2. DECESSION TREE:

II. Decision Tree

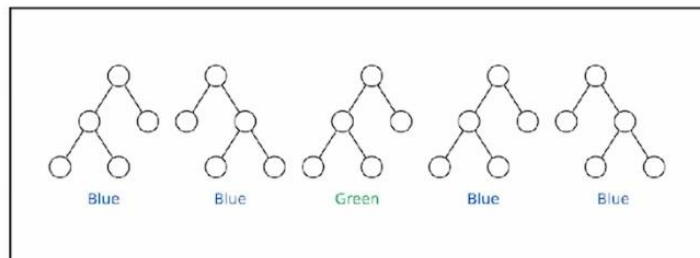


In picture each square above is node, the more node we have the more accurate decision tree we have

3. RANDOM FORESTS:

III. Random Forests

Ensemble learning technique



↓
Blue

“Majority Wins” Model

Ensemble learning technique that builds over decision tree. It involves creating multiple decision trees using bootstrapped data sets of original data and randomly selecting a subset of variables at each step of decision tree.

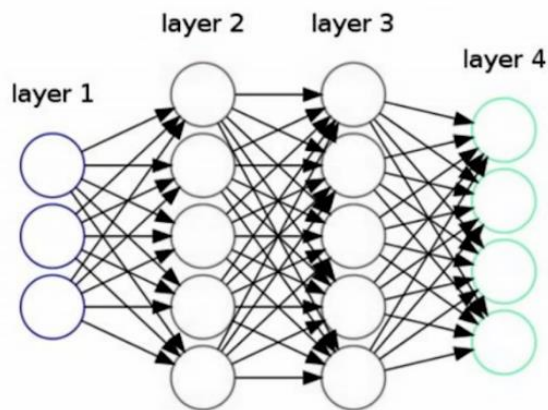
The model then selects the mode of all the predictions of each decision tree and by relying on ‘Majority Wins’ model it produces the risk of error from individual tree.

4. NEURAL NETWORK:

Multilayer model inspired by human mind

- In below image circle represents node
- Blue circle represents an input layer
- Black circle represents a hidden layer
- Green circle represents output layer
- Each layer in hidden layer represents the function that input goes through ultimately leading to output in the green circles

IV. Neural Network



Layer 1 = Input Layer
Layer 2 = Hidden Layer
Layer 3 = Hidden Layer
Layer 4 = Output Layer

B. TYPES OF CLASSIFICATION: Output is discrete

1. LOGISTIC REGRESSION:

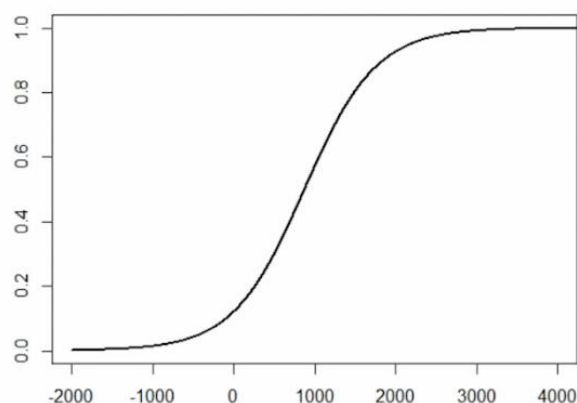
Use to model the probability of finite number of outcomes like ranges from 0 to 1

1.2 Classification

discrete

I. Logistic regression

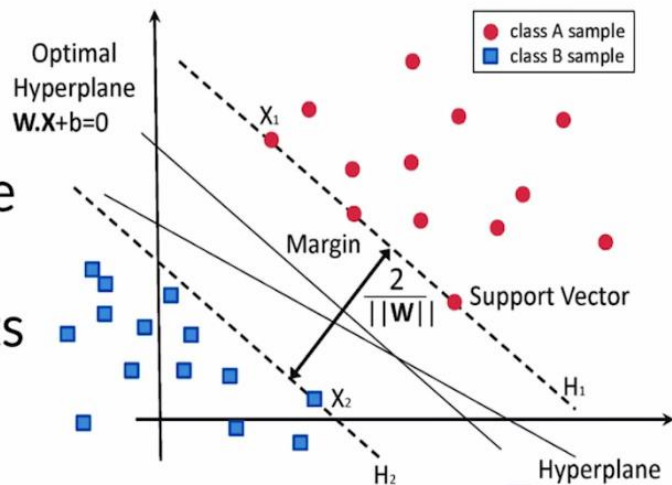
the output values can only be between 0 and 1



2. SUPPORT VECTOR MACHINE:

II. Support Vector Machine

N-dimensional space
that distinctly
classifies data points



It carries an objective to find a hyperlink in n dimensional space distinctly classified the data points.

3. NAÏVE BAYES:

III. Naive Bayes

$$P(B | A) = \frac{P(A | B)P(B)}{P(A)}$$

Likelihood points to $P(A | B)$
Prior probability points to $P(B)$
Posterior probability points to $P(B | A)$
Predictor points to $P(A)$

IV. Decision Tree, Random Forest, Neural Network

A classifier which acts as a probabilistic machine learning model used for classification tasks. The cracks of classifier are based on base theorem.

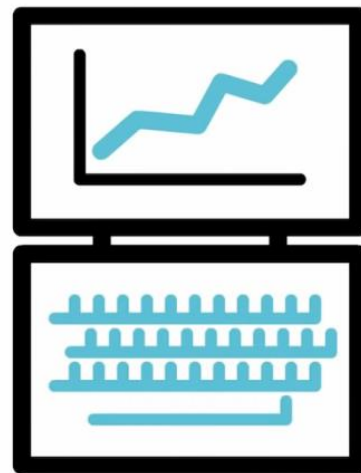
**** For classification; Decision tree, Random Forest, Neural Network models follow the same logic as regression, but only difference is that in classification output is discrete rather than continuous**

UNSUPERVISED:

2. Unsupervised Learning

Patterns from input data
without references to
labeled outcomes

- > Clustering
- > Dimensionality Reduction



C. CLUSTERING:

- > Involves grouping of data points.
- > Frequently used for customer segmentation, fraud detection and document classification.

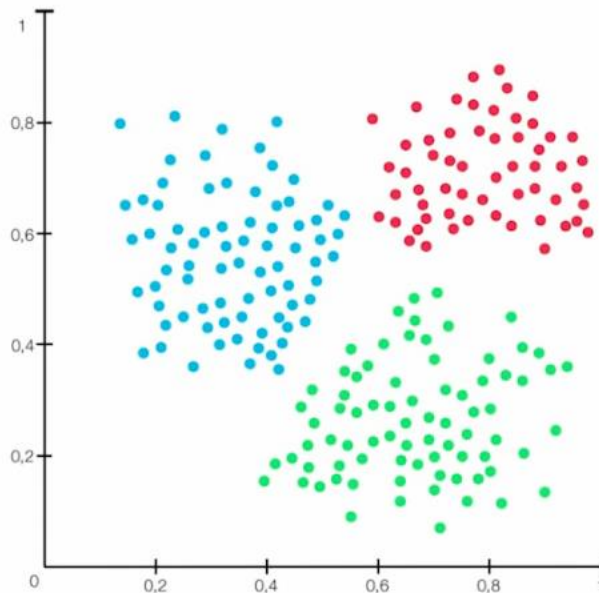
Common clustering includes:

- K-means
- Hierarchical
- Mean shift
- Density based

Each technique has different method in finding clusters,

But all have same goal to achieve that is to find clusters.

2.1 Clustering



- > K-means
- > Hierarchical
- > Mean shift
- > Density-based

D. DIMENSIONALITY REDUCTION:

Use to reduce number of features

2.2 Dimensionality Reduction

process of reducing the dimension
of your feature set

- > feature elimination
- > feature extraction

PRINCIPAL COMPONENT ANALYSIS (PCA)