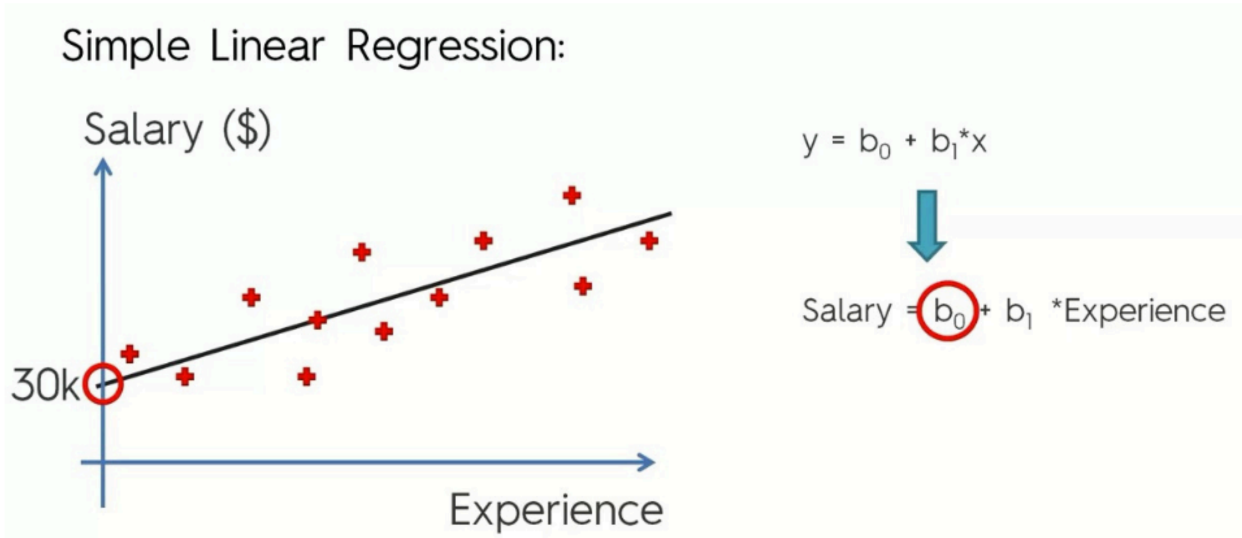
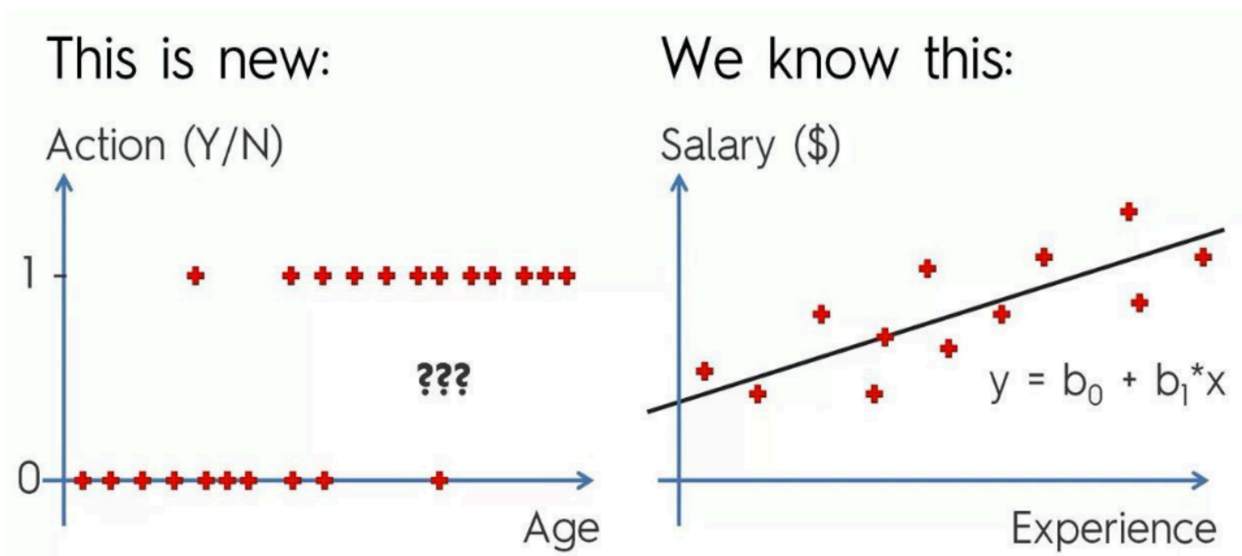
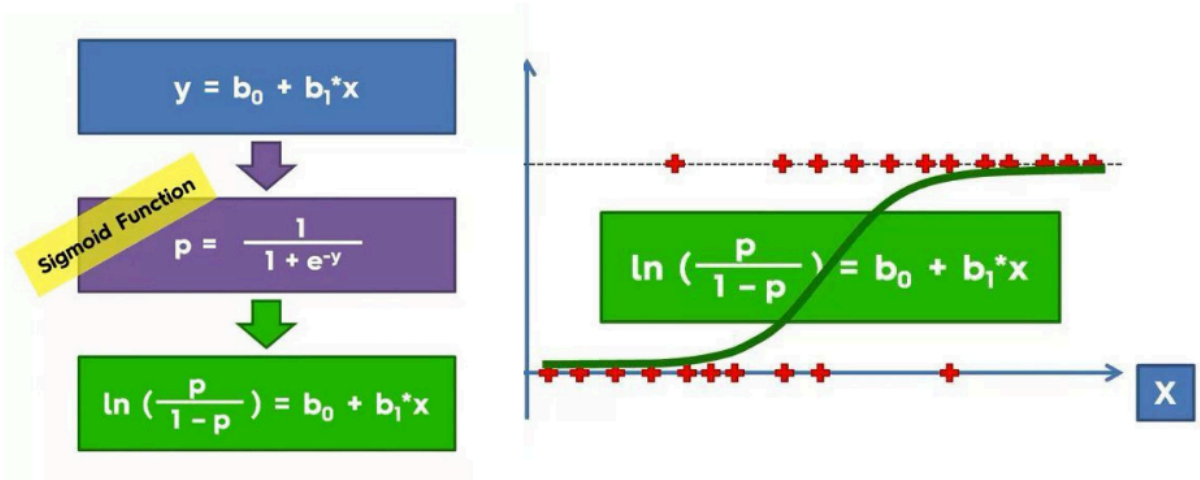
Logistic Regression



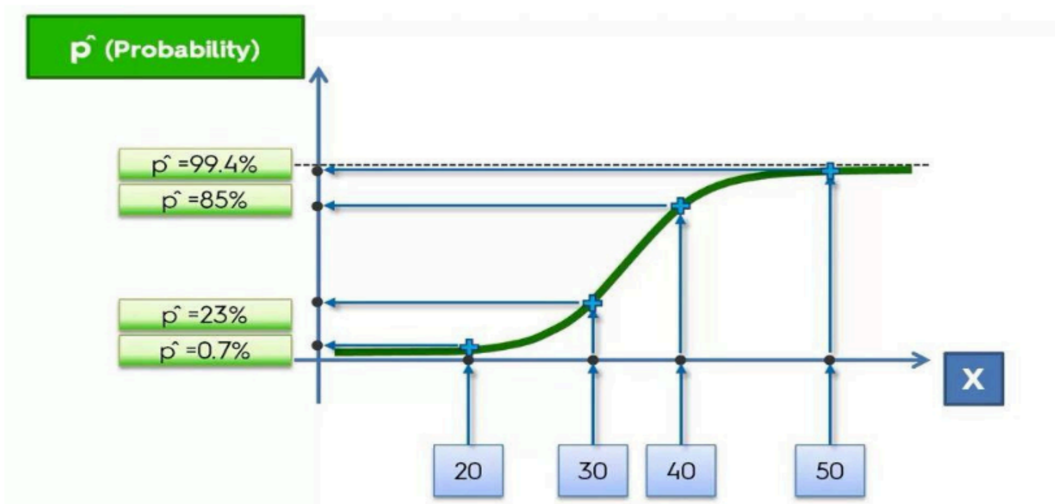
Logistic R vsLinear R



Logistic Regression Formula



Logistic Regression working

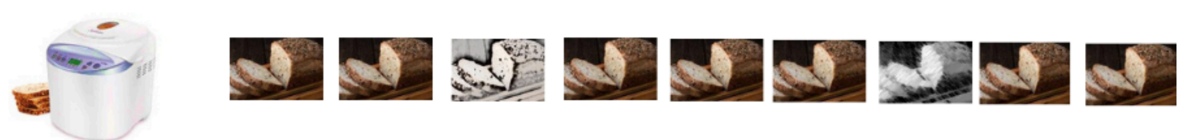


Bayes' theorem

In probability theory and statistics, Bayes' theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event

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

Bayes' theorem



Machine 1
30 Bread per hour

Machine 2
20 Bread per hour

Machine1: 30 Breads / hr
Machine2: 20 Breads /hr
Out of all product parts:
We can SEE that 1% are defective
Out of all defective parts:
We can See that 50% came form Machine1
And 50% came from Machine2
What is probability that a part produced by machine1
Is defective =?

Bayes' theorem

=> P(Machine1) = 30/50 = 0.6
=> P(Machine2) = 20/50 = 0.4
=> P(Defect) = 1% => P(Machine1|Defect) = 50%
=> P(Machine2|Defect) = 50%

=> P(Defect | Machine2) = ?

$$P(\text{Defect} \mid \text{Machine2}) = \frac{P(\text{Machine2}|\text{Defect}) * P(\text{Defect})}{P(\text{Machine2})}$$

$$P(\text{Defect} \mid \text{Machine2}) = \frac{0.5 * 0.01}{0.4} = 0.0125 \Rightarrow 1.25\%$$

Naïve Bayes' theorem

It is aclassification technique based onBayes' Theorem with an assumption of independence among predictors. 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.

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

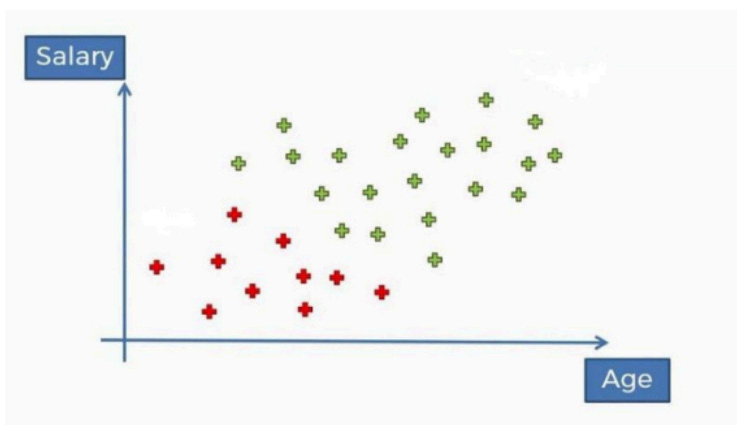
Likelihood 0.3

Prior Probability
0.33

Posterior Probability

$$P(\,ClassA \mid X) = \frac{P(X \mid ClassA) \; * \; P(\,ClassA\,)}{P(X)} = 0.75$$

Marginal Probability
0.13



P(ClassA) = 10 /30 = 0.33

Using Circle P(X) = 4/30 = 0.13

Using Ciriclewith Red Data
P(X | ClassA) = 3 /10 = 0.3

P(ClassB) = 20 /30 = 0.66

Using Circle P(X) = 4/30 = 0.13

Using Ciriclewith Red Data
P(X | ClassB) = 1 /20 = 0.05

Posterior Probability

Likelihood

Prior Probability

$$P(\,ClassB \mid X) = \frac{P(X \mid ClassB) \; * \; P(\,ClassB\,)}{P(X)} = 0.25$$

Marginal Probability

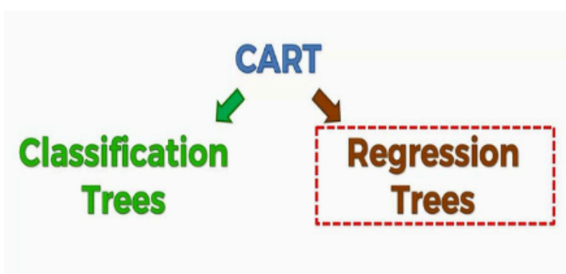
0.75 < 0.25

New Data will be allocated to ClassA

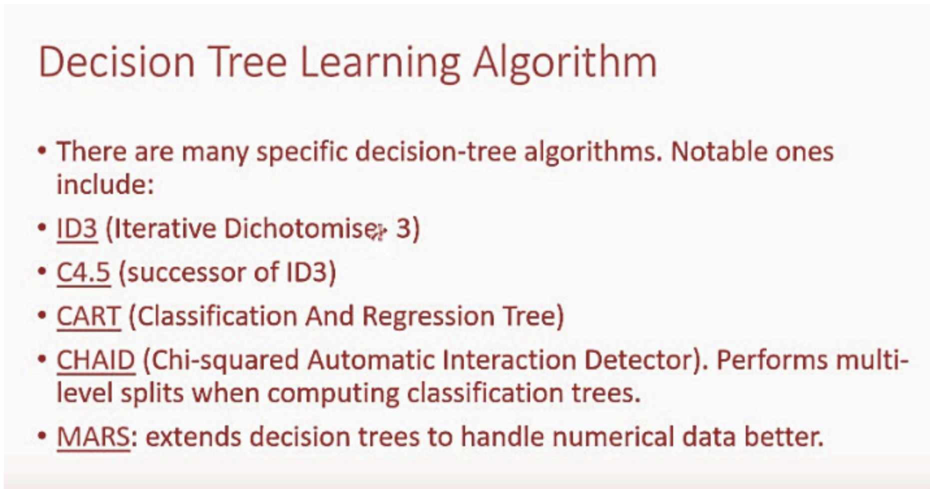
There are three types of Naive Bayes model under the scikit-learn library: •Gaussian: It is used in classification and it assumes that features follow a normal distribution. •Multinomial: It is used for discrete counts. For example, let's say, we have a text classification problem. Here we can consider Bernoulli trials which is one step further and instead of "word occurring in the document", we have "count how often word occurs in the document", you can think of it as "number of times outcome number x_i is observed over the n trials". •Bernoulli: The binomial model is useful if your feature vectors are binary (i.e. zeros and ones). One application would be text classification with 'bag of words' model where the 1s & 0s are "word occurs in the document" and "word does not occur in the document" respectively.

Decision Tree Classifier

A Classification And Regression Tree (CART), is a predictive model, which explains how an outcome variable's values can be predicted based on other values. ACART output is a decision tree where each fork is a split in a predictor variable and each end node contains a prediction for the outcome variable

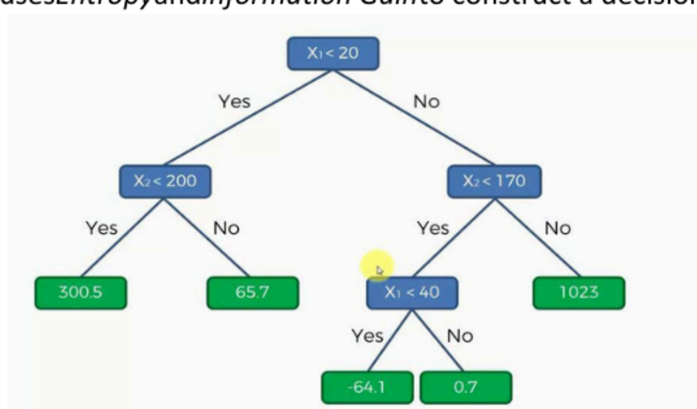


Decision Tree Classifier



ID3 Decision Tree Classifier

ID3: The core algorithm for building decision trees is called**ID3.**Developedby J. R. Quinlan, this algorithm employs a top-down, greedy search through the space of possible branches with no backtracking. ID3 uses*Entropy*and*Information Gain*to construct a decision tree.



Concept to map Decision Tree

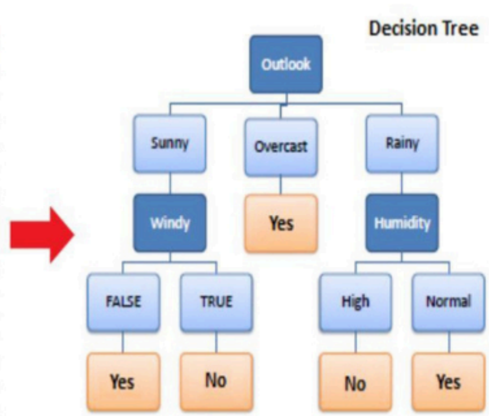
Information Gain: Theinformation gainis based on the decrease in entropy after a data-set is split on an attribute. Constructing adecision treeis all about finding attribute that returns the highestinformation gain(i.e., the most homogeneous branches)

OR [a measure of the decrease in disorder achieved by partitioning the original dataset]

Entropy :Entropy, as it relates tomachine learning, is a measure of the randomness in the information being processed. The higher theentropy, the harder it is to draw any conclusions from that information. Flipping a coin is an example of an action that provides information that is random. ... This is the essence ofentropy.
. *Or [is a measure of disorder in a dataset]*

ID3 Decision Tree Classifier

Predictors				Target
Outlook	**Temp**	**Humidity**	**Windy**	**Play Golf**
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No



Concept to map Decision Tree

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

Play Golf	
Yes	No
9	5

Entropy(PlayGolf) = Entropy (5,9)
= Entropy (0.36, 0.64)
= - (0.36 $\log_2$ 0.36) - (0.64 $\log_2$ 0.64)
= 0.94

=-p / p + n * log (p / p+ n) -n / p + n * log(n / p + n)
=-9 / 5 + 5 * log (9 / 9+ 5) -5 / 9 + 5 * log(5 / 9 + 5)
Here p = 9, n = 5   also for log2 = log(?) / log(2)

Concept to map Decision Tree

$$E(T, X) = \sum_{c \in X} P(c)E(c)$$

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
				14

E(PlayGolf, Outlook) = **P**(Sunny)***E**(3,2) + **P**(Overcast)***E**(4,0) + **P**(Rainy)***E**(2,3)
= (5/14)*0.971 + (4/14)*0.0 + (5/14)*0.971
= 0.693

$-\frac{2}{5}\log \frac{2}{5} - \frac{3}{5}\log 3/5 = 0.970$

$-\frac{4}{5}\log \frac{4}{5} - \frac{3}{5}\log 0 = 0$

$-\frac{3}{5}\log \frac{3}{5} - \frac{2}{5}\log 2/5 = 0.970$

Concept to map Decision Tree

$$E(T,X) = \sum_{c \in X} P(c)E(c)$$

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
				14

E(PlayGolf, Outlook) = **P**(Sunny)***E**(3,2) + **P**(Overcast)***E**(4,0) + **P**(Rainy)***E**(2,3)

= (5/14)*0.971 + (4/14)*0.0 + (5/14)*0.971

= 0.693

Entropy E(A) = pi + ni/ p + n * (I (p,n))

Concept to map Decision Tree

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
Gain = 0.247			

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1
Gain = 0.029			

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1
Gain = 0.152			

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3
Gain = 0.048			

Information Gain

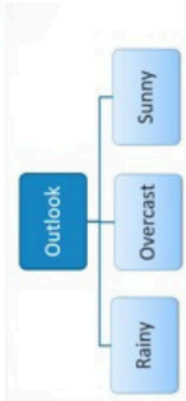
$$Gain(T,X) = Entropy(T) - Entropy(T,X)$$

G(PlayGolf, Outlook) = **E**(PlayGolf) – **E**(PlayGolf, Outlook)

= 0.940 – 0.693 = 0.247

Concept to map Decision Tree

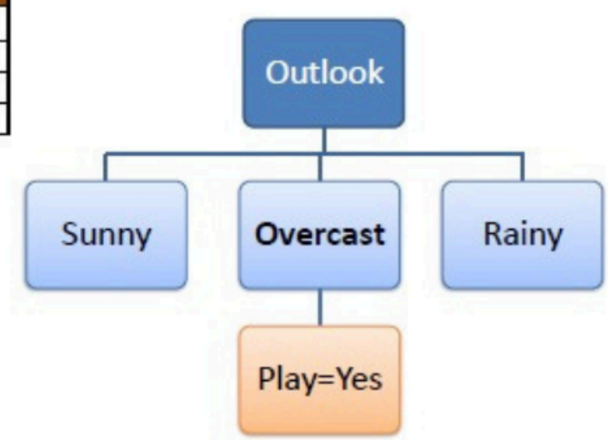
		Play Golf	
★		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
Gain = 0.247			



Outlook	Temp	Humidity	Windy	Play Golf
Sunny	Mild	High	FALSE	Yes
Sunny	Cool	Normal	FALSE	Yes
Sunny	Cool	Normal	TRUE	No
Sunny	Mild	Normal	FALSE	Yes
Sunny	Mild	High	TRUE	No
Overcast	Hot	High	FALSE	Yes
Overcast	Cool	Normal	TRUE	Yes
Overcast	Mild	High	TRUE	Yes
Overcast	Hot	Normal	FALSE	Yes
Rainy	Hot	High	FALSE	No
Rainy	Hot	High	TRUE	No
Rainy	Mild	High	FALSE	No
Rainy	Cool	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes

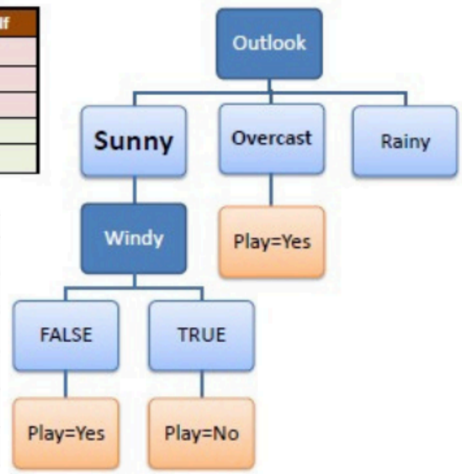
Concept to map Decision Tree

Temp	Humidity	Windy	Play Golf
Hot	High	FALSE	Yes
Cool	Normal	TRUE	Yes
Mild	High	TRUE	Yes
Hot	Normal	FALSE	Yes



Concept to map Decision Tree

Temp	Humidity	Windy	Play Golf
Mild	High	FALSE	Yes
Cool	Normal	FALSE	Yes
Mild	Normal	FALSE	Yes
Cool	Normal	TRUE	No
Mild	High	TRUE	No



	WIND		
wind	pi	ni	E
TRUE	0	2	0
FALSE	3	0	0
E	0		
G	0.97		

	Temp		
Temp	pi	ni	E
hot	0	0	0
mild	2	1	0.9
cold	1	1	1
E	0.951		
G	0.019		

Concept to map Decision Tree

A decision tree can easily be transformed to a set of rules by mapping from the root node to the leaf nodes one by one.

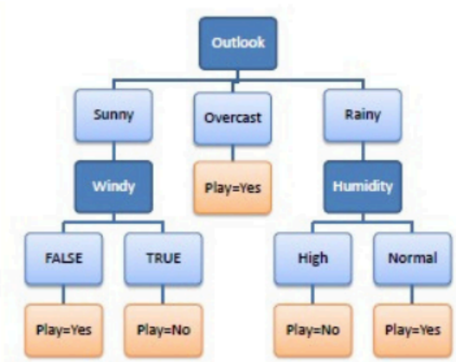
R_1: IF (Outlook=Sunny) AND (Windy=FALSE) THEN Play=Yes

R_2: IF (Outlook=Sunny) AND (Windy=TRUE) THEN Play=No

R_3: IF (Outlook=Overcast) THEN Play=Yes

R_4: IF (Outlook=Rainy) AND (Humidity=High) THEN Play=No

R_5: IF (Outlook=Rain) AND (Humidity=Normal) THEN Play=Yes



Concept to map Decision Tree

Gini Index

Giniindex says, if we select two items from a population at random then they must be of same class and probability for this is 1 if population is pure.

It can be used only if the target variable is a binary variable Classification and regression Tree (CART).

Concept to map Decision Tree

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2,$$

$$Gini(D) = 1 - \left(\frac{9}{14}\right)^2 - \left(\frac{5}{14}\right)^2 = 0.459$$

RandomForestTree Classifier

In statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone.

Steps to make RandomForestTree Classifier

STEP 1: Pick at random K data points from the Training set.
STEP 2: Build the Decision Tree associated to these K data points.
STEP 3: Choose the number Ntreeof trees you want to build and repeat STEPS 1 & 2
STEP 4: For a new data point, make each one of your Ntreetrees predict the category to which the data points belongs, and assign the new data point to the category that wins the majority vote.

Steps to make RandomForestTree Classifier

STEP 1: Pick at random K data points from the Training set. STEP 2: Build the Decision Tree associated to these K data points. STEP 3: Choose the number Ntreeof trees you want to build and repeat STEPS 1 & 2 STEP 4: For a new data point, make each one of your Ntreetrees predict the category to which the data points belongs, and assign the new data point to the category that wins the majority vote.