Machine Learning

Dr.Venkataramana Veeramsetty
NVIDIA DLI Ambassador & Head of CAIDL
Assistant Professor in Dept. of EEE
SR Engineering College Warangal

July 22, 2020

Email ID: (vvr.nitw@ieee.org)

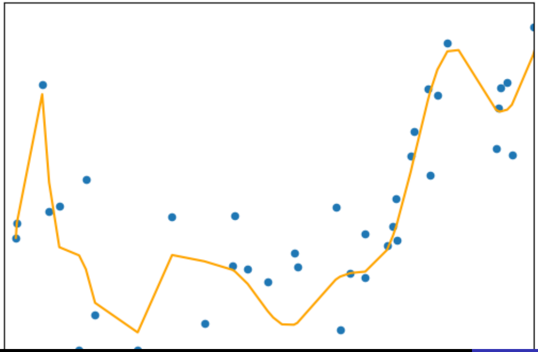
Machine Learning Techniques

- √ Simple Linear Regression
- √ Multiple Linear Regression
- √ Polynomial Linear Regression
- √ Logistic Regression
- √ Support Vector Machines (SVM)
- √ Decision Trees
- √ Random Forest

Email ID: (vvr.nitw@ieee.org)

Polynomial Linear Regression

- Linear regression requires the relation between the dependent variable and the independent variable to be linear.
- How we approach if data scattered like below ?



Email ID: (vvr.nitw@ieee.org)

- If we try to use Linear Regression for above data then the model will be under fitting
- Then we go with Polynomial Regression. It is nothing but we need to develop a linear model which relates higher order inputs to output

$$y = ax + b \tag{1}$$

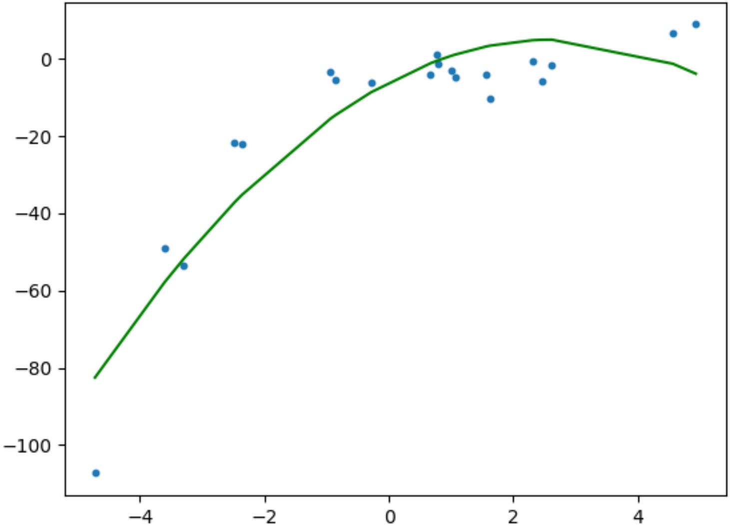
$$y = a_n x^n + a_{n-1} x^{n-1} + ... + a_1 x + b \tag{2}$$

Email ID: (vvr.nitw@ieee.org)

- $\surd$ To convert the original features into their higher order terms we will use the **PolynomialFeatures** class provided by **scikit-learn**.
- $\surd$ Next, we train the model using Linear Regression.

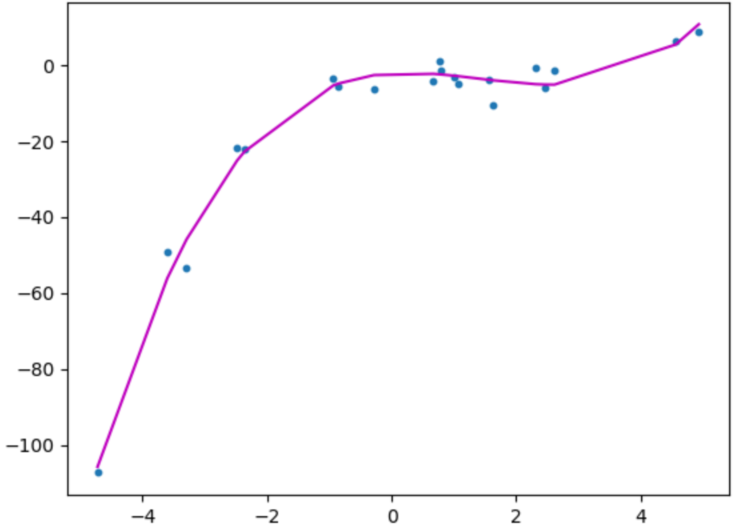
Email ID: (vvr.nitw@ieee.org)

Regression With Degree=2



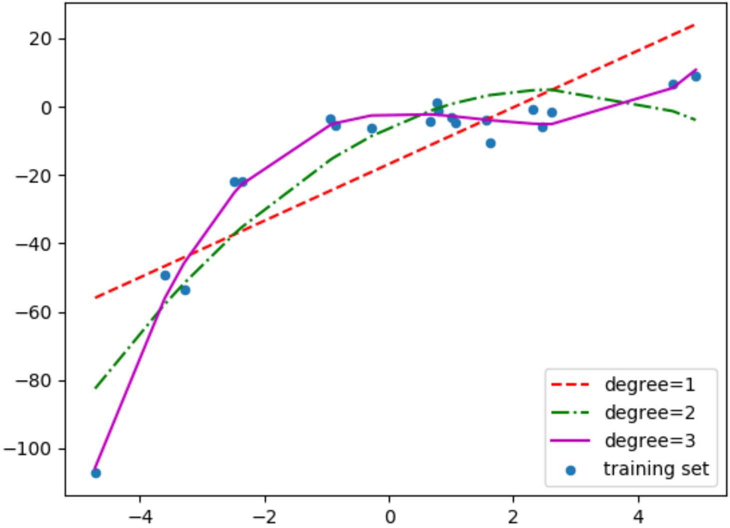
Email ID: (vvr.nitw@ieee.org)

Regression With Degree=3



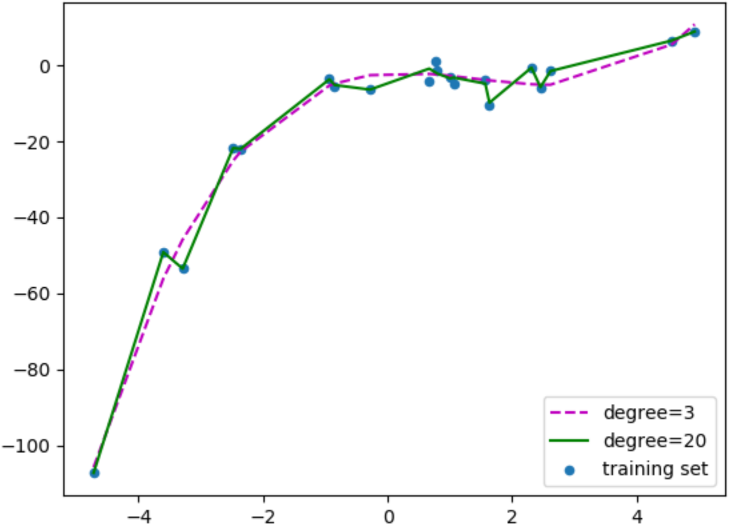
Email ID: (vvr.nitw@ieee.org)

Regression Comparison



Email ID: (vvr.nitw@ieee.org)

Regression With Degree=20



Email ID: (vvr.nitw@ieee.org)

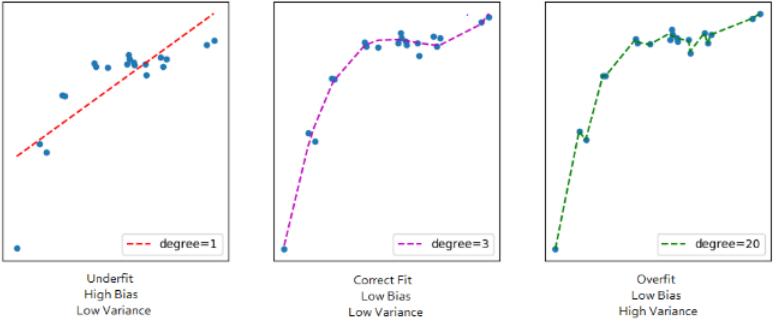
- For degree=20, the model is capturing all the data points. This is an example of over-fitting. Even though this model passes through most of the data, it will fail to generalize on unseen data.
- To prevent over-fitting, we can add more training samples so that the algorithm doesn't learn the noise in the system and can become more generalized.

Email ID: (vvr.nitw@ieee.org)

Bias Vs Variance

- √ **Bias** refers to the error due to the model's simplistic assumptions in fitting the data. A **high bias** means that the model is unable to capture the patterns in the data and this results in **under-fitting**.
- √ **Variance** refers to the error due to the complex model trying to fit the data. **High variance** means the model passes through most of the data points and it results in **over-fitting** the data.

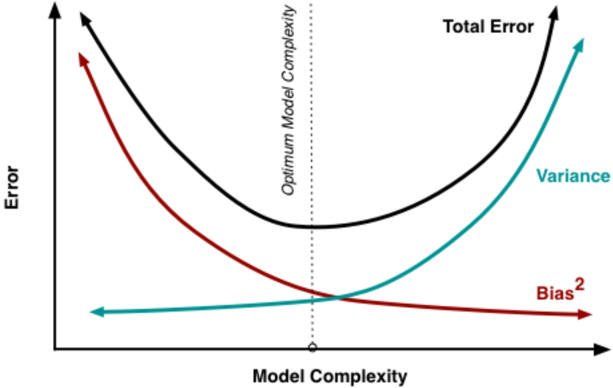
Email ID: (vvr.nitw@ieee.org)



- From the picture shown in previous slide we can observe that as the model complexity increases, the bias decreases and the variance increases and vice-versa.

Email ID: (vvr.nitw@ieee.org)

- Ideally, a machine learning model should have low variance and low bias. But practically it's impossible to have both.
- Therefore to achieve a good model that performs well both on the train and unseen data, a trade-off is made.



Email ID: (vvr.nitw@ieee.org)

Decision Tree Algorithm

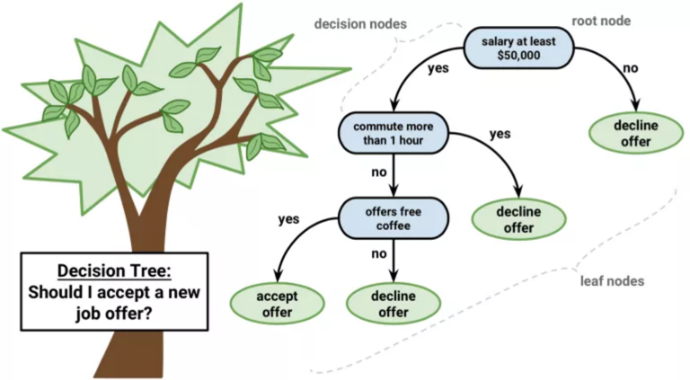
- √ It is a supervised algorithm
- √ It can be used for both regression and classification problems
- √ The main goal of this algorithm is to develop a training model which can predict or classify the target variable based on learning decision rules which were generated from historical data
- √ Decision tree has two nodes
  - Internal node - Represents attribute or feature
  - Leaf node - Represents class label or value

Email ID: (vvr.nitw@ieee.org)

Decision Tree Algorithm - Pseudo code

- √ Place best attribute/feature in the dataset as the root of the tree
- √ Split the training set into subsets. Subsets should be made in such a way that each subset contains data with the same value for an attribute
- √ Repeat step 1 and step 2 on each subset until you find leaf nodes in all the branches of the tree.

Email ID: (vvr.nitw@ieee.org)



Email ID: (vvr.nitw@ieee.org)

- √ In decision trees, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.
- √ We continue comparing our record's attribute values with other internal nodes of the tree until we reach a leaf node with predicted class value.

Email ID: (vvr.nitw@ieee.org)

Challenges

- The primary challenge in the decision tree implementation is to identify which attributes do we need to consider as the root node and each level
- If dataset consists of "n" attributes then deciding which attribute to place at the root or at different levels of the tree as internal nodes is a complicated step.
- By just randomly selecting any node to be the root can't solve the issue. If we follow a random approach, it may give us bad results with low accuracy.

Email ID: (vvr.nitw@ieee.org)

Attribute Selection measures

- √ The popular attribute selection measures:
  - Information gain
  - Gini Index
- √ The values are sorted, and attributes are placed in the tree by following the order i.e, the attribute with a high value(in case of information gain) is placed at the root.
- √ While using information Gain as a criterion, we assume attributes to be categorical, and for gini index, attributes are assumed to be continuous.

Email ID: (vvr.nitw@ieee.org)

Construct a Decision Tree by using "information gain" as a criterion

Features				Target	Features				Target
A	B	C	D	E	A	B	C	D	E
4.8	3.4	1.9	0.2	P	7	3.2	4.7	1.4	N
5	3	1.6	0.2	P	6.4	3.2	4.5	1.5	N
5	3.4	1.6	0.4	P	6.9	3.1	4.9	1.5	N
5.2	3.5	1.5	0.2	P	5.5	2.3	4	1.3	N
5.2	3.4	1.4	0.2	P	6.5	2.8	4.6	1.5	N
4.7	3.2	1.6	0.2	P	5.7	2.8	4.5	1.3	N
4.8	3.1	1.6	0.2	P	6.3	3.3	4.7	1.6	N
5.4	3.4	1.5	0.4	P	4.9	2.4	3.3	1	N

Email ID: (vvr.nitw@ieee.org)

√ Convert continuous information of each attribute in to categorical as information gain can be calculated on categorical attributes only

Table 1 : Categorical conversion

Features			
A	B	C	D
≥ 5	≥ 3	≥ 4.2	≥ 1.4
< 5	< 3	< 4.2	< 1.4

Email ID: (vvr.nitw@ieee.org)

Steps to calculate IG

- √ Step1:Calculate entropy of Target.
- √ Step2:Calculate entropy of all attributes.
- √ Step3:IG(Attribute)=Entropy(Target) - Entropy(Attribute)

$$Entropy(X) = -\sum_{x \in X} p(x) log p(x) \qquad (3)$$

X may be target or attribute

Email ID: (vvr.nitw@ieee.org)

Calculate Entropy of Target

Target	
P=8	N=8
p(P)=8/16=0.5	p(N)=8/16=0.5

Entropy(Target)=-p(P)log(p(P))-p(N)log(p(N))

$=-0.5*log_2(0.5)-0.5*log_2(0.5)=1$

Email ID: (vvr.nitw@ieee.org)

Information gain of A

Category	P	N	p(P)	p(N)
$A \geq 5$	5	7	$\frac{5}{12}$	$\frac{7}{12}$
$A < 5$	3	1	$\frac{3}{4}$	$\frac{1}{4}$

Entropy($A \geq 5$)=-p(P)log_2(p(P))-p(N)log_2(p(N))=-(0.4)log_2(0.4)-(0.58)log_2(0.58)=0.98

Entropy($A < 5$)=-p(P)log_2(p(P))-p(N)log_2(p(N))=-(0.75)log_2(0.75)-(0.25)log_2(0.25)=0.81

Entropy(Target,A)=p($A \geq 5$)Entropy($A \geq 5$)+p($A < 5$)Entro

IG(A)=Entropy(Target)-Entropy(Target,A)=1-0.9377=0.0623

Note: $log_2 a = \frac{log_{10} a}{log_{10} 2}$

Email ID: (vvr.nitw@ieee.org)

IG of attribute A

$A \geq 5$
- $P : 5 \rightarrow P(P) : \frac{5}{12}$
- $N : 7 \rightarrow P(N) : \frac{7}{12}$

$A < 5$
- $P : 3 \rightarrow P(P) : \frac{3}{4}$
- $N : 1 \rightarrow P(N) : \frac{1}{4}$

$$\text{Entropy}(\geq 5) : -\left[\frac{5}{12}\log_2\left(\frac{5}{12}\right) + \frac{7}{12}\log_2\left(\frac{7}{12}\right)\right] = 0.9799$$

$$\text{Entropy}(< 5) : -\left[\frac{3}{4}\log_2\left(\frac{3}{4}\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right] = 0.81128$$

$$\text{Entropy}(\text{Target}, A) = P(\geq 5) * \text{Entropy}(\geq 5) + P(<5) * \text{Entropy}(<5)$$

$$= \frac{12}{16} * 0.9799 + \frac{4}{16} * 0.81128$$

$$= 0.9377$$

$$IG(A) = \text{Entropy}(\text{Target}) - \text{Entropy}(\text{Target}, A) = 1 - 0.9377 = 0.06225$$

Email ID: (vvr.nitw@ieee.org)

IG of attribute B

$B >= 3$
- $\rightarrow P: 8 \rightarrow P(P): \frac{8}{12}$
- $\rightarrow N: 4 \rightarrow P(N): \frac{4}{12}$

$B < 3$
- $\rightarrow P: 0 \rightarrow P(P): \frac{0}{4}$
- $\rightarrow N: 4 \rightarrow P(N): \frac{4}{4}$

$$\text{Entropy}(\geq 3): -\left[\frac{8}{12}\log_2\left(\frac{8}{12}\right) + \frac{4}{12}\log_2\left(\frac{4}{12}\right)\right] = 0.39054$$

$$\text{Entropy}(<3): -\left[\frac{0}{4}\log_2\left(\frac{0}{4}\right) + \frac{4}{4}\log\left(\frac{4}{4}\right)\right] = 0$$

$$\text{Entropy}(\text{Target}, B) = P(B \geq 3)\, E(\geq 3) + P(B<3)\, E(<3)$$

$$= \left[\frac{12}{16}\right](0.39054) + \frac{4}{16}\left[0\right] = 0.2929$$

$$IG(B) = \text{Entropy}(\text{Target}) - \text{Entropy}(\text{Target}, B) = 1 - 0.2929 = 0.7070$$

Email ID: (vvr.nitw@ieee.org)

IG of attribute C

$$C \geq 4.2 \rightarrow \begin{cases} P: 0 \rightarrow P(P) = \frac{0}{6} = 0 \\ N: 6 \rightarrow P(N) = \frac{6}{6} = 1 \end{cases}$$

$$C < 4.2 \rightarrow \begin{cases} P: 8 \rightarrow P(P) = \frac{8}{10} = 0.8 \\ N: 2 \rightarrow P(N) = \frac{2}{10} = 0.2 \end{cases}$$

$$\text{Entropy}(\geq 4.2) = -\left[0 \log_2 0 + 1 \log_2 1\right] = 0$$

$$\text{Entropy}(< 4.2) = -\left[0.8 \log_2 0.8 + 0.2 \log_2 0.2\right] = 0.72193$$

$$\text{Entropy}(\text{Target}, C) = P(C \geq 4.2)\, E(\geq 4.2) + P(< 4.2)\, E(< 4.2)$$

$$= \frac{6}{16} * 0 + \frac{10}{16} * 0.72193 = 0.4512$$

$$IG(C) = \text{Entropy}(\text{Target}) - \text{Entropy}(\text{Target}, C)$$

$$= 1 - 0.4512 = 0.5488$$

Email ID: (vvr.nitw@ieee.org)

IG of attribute D

$$D(>=1.4) \rightarrow \begin{cases} P = 0 \rightarrow p(P) = \frac{0}{5} = 0 \\ N = 5 \rightarrow p(N) = \frac{5}{5} = 1 \end{cases}$$

$$D(<1.4) \rightarrow \begin{cases} P = 8 \rightarrow p(P) = \frac{8}{11} \rightarrow 0.7272 \\ N = 3 \rightarrow p(N) = \frac{3}{11} \rightarrow 0.2727 \end{cases}$$

$$\text{Entropy}(D \geq 1.4) \rightarrow -\left[0 \log_2 0 + 1 \log_2 1\right] = 0$$

$$\text{Entropy}(D < 1.4) \rightarrow -\left[0.7272 \log_2 \frac{8}{11} + 0.2727 \log_2 \frac{3}{11}\right] = 0.8453$$

$$\text{Entropy}(\text{Target}, D) = P(D \geq 1.4) * E(D \geq 1.4) + P(D < 1.4) * E(D < 1.4)$$

$$= \left(\frac{5}{16}\right)(0) + \left(\frac{11}{16}\right)(0.8453) = 0.5811575$$

$$IG(D) = \text{Entropy}(\text{Target}) - \text{Entropy}(\text{Target}, D)$$

$$= 1 - 0.5811575 = 0.41189$$

Email ID: (vvr.nitw@ieee.org)

Information gain chart

		Target	
		Positive	Negative
A	>= 5.0	5	7
	<5	3	1
Information Gain of A = 0.062255			

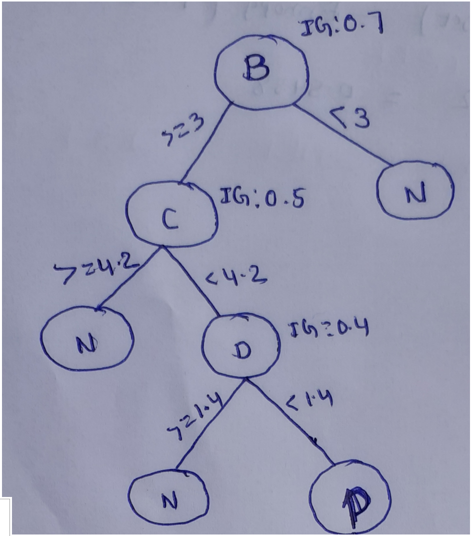
		Target	
		Positive	Negative
B	>= 3.0	8	4
	< 3.0	0	4
Information Gain of B= 0.7070795			

		Target	
		Positive	Negative
C	>= 4.2	0	6
	< 4.2	8	2
Information Gain of C= 0.5488			

		Target	
		Positive	Negative
D	>= 1.4	0	5
	< 1.4	8	3
Information Gain of D= 0.41189			

Email ID: (vvr.nitw@ieee.org)

Decision Tree



Email ID: (vvr.nitw@ieee.org)

Construct a Decision Tree by using "Gini Index" as a criterion

- √ Gini Index is a metric to measure how often a randomly chosen element would be incorrectly identified. It means an attribute with lower gini index should be preferred.

Email ID: (vvr.nitw@ieee.org)

Gini Index of A

Category	P	N	p(P)	p(N)	Gini
$A \geq 5$	5	7	$\frac{5}{12}$	$\frac{7}{12}$	$1\text{-}((\frac{5}{12})^2 + (\frac{7}{12})^2)=0.486$
$A < 5$	3	1	$\frac{3}{4}$	$\frac{1}{4}$	$1\text{-}((\frac{3}{4})^2 + (\frac{1}{4})^2)=0.375$

$$\text{Gini(Target,A)}=p(A \geq 5)*\text{Gini}(A \geq 5)+p(A<5)*\text{Gini}(A<5)$$
$$=(12/16)*0.486+(4/16)*0.375=0.3645+0.0938=0.4582$$

Gini Index of B

Category	P	N	p(P)	p(N)	Gini
B$\geq$3	8	4	$\frac{8}{12}$	$\frac{4}{12}$	1-$((\frac{8}{12})^2 + (\frac{4}{12})^2)$=0.446
B$<$3	0	4	$\frac{0}{4}$	$\frac{4}{4}$	1-$((\frac{0}{4})^2 + (\frac{4}{4})^2)$=0

Gini(Target,B)=p(B$\geq$3)*Gini(B$\geq$3)+p(B$<$3)*Gini(B$<$3)
=(12/16)*0.446+(4/16)*0=0.3345+0=0.3345

Email ID: (v[illegible]nitw@ieee.org)

Gini Index of C

Category	P	N	p(P)	p(N)	Gini
C$\geq$4.2	0	6	$\frac{0}{6}$	$\frac{6}{6}$	1-$((\frac{0}{6})^2+(\frac{6}{6})^2)$=0
C$<$4.2	8	2	$\frac{8}{10}$	$\frac{2}{10}$	1-$((\frac{8}{10})^2+(\frac{2}{10})^2)$=0.32

Gini(Target,C)=p(C$\geq$4.2)*Gini(C$\geq$4.2)+p(C$<$4.2)*Gini(

=(6/16)*0+(10/16)*0.32=0.2

Email ID: (vvr.nitw@ieee.org)

Gini Index of D

Category	P	N	p(P)	p(N)	Gini
D$\geq$1.4	0	5	$\frac{0}{5}$	$\frac{5}{5}$	1-$((\frac{0}{5})^2+(\frac{5}{5})^2)$=0
D$<$1.4	8	3	$\frac{8}{11}$	$\frac{3}{11}$	1-$((\frac{8}{11})^2+(\frac{3}{11})^2)$=0.3966

Gini(Target,D)=p(D$\geq$1.4)*Gini(D$\geq$1.4)+
p(D$<$1.4)*Gini(D$<$1.4)
=(5/16)*0+(11/16)*0.3966=0.273

Email ID:(vvr.nitw@ieee.org)

GINI Index of attributes

		wTarget	
		Positive	Negative
A	>= 5.0	5	7
	<5	3	1
Ginin Index of A = 0.45825			

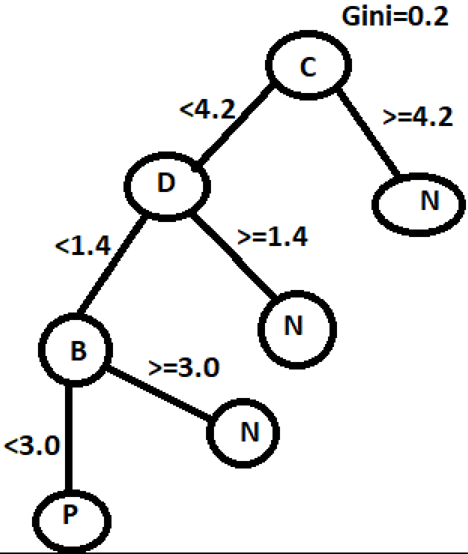
		Target	
		Positive	Negative
B	>= 3.0	8	4
	< 3.0	0	4
Gini Index of B= 0.3345			

		Target	
		Positive	Negative
C	>= 4.2	0	6
	< 4.2	8	2
Gini Index of C= 0.2			

		Target	
		Positive	Negative
D	>= 1.4	0	5
	< 1.4	8	3
Gini Index of D= 0.273			

Email ID: (vvr.nitw@ieee.org)

Decision Tree - Using Gini



Email ID: (vvr.nitw@ieee.org)

References

- √ https://towardsdatascience.com/polynomial-regression-bbe8b9d97491
- √ https://dataaspirant.com/how-decision-tree-algorithm-works/

Email ID: (vvr.nitw@ieee.org)

Email ID: (vvr.nitw@ieee.org)