Unit :- 3

Introduction to Machine Learning

Page No.
Date : 6 / 1 / 25

⇒ We wish to program computers so that they can learn from inputs available to them.

⇒ Roughly, learning is the process of converting experience into expertise / knowledge.

Input (Training data) → Learning Algo → Output (expertise/knowledge)

Input (Training data) represent experience

Learning Algo usually takes the form of another computer program that can perform task

- What is the training data?
- How can the process of learning be automated?
- How can we evaluate the success of each process?

☆ Need of ML :-

1. The problem's complexity
2. Adaptivity

Too Complex to Program :- Tasks human perform routinely

- Driving
- Speech Recognition
- image understanding

- Astronomical data
- Turning medical archives into medical knowledge
- weather prediction
- Analysis of genomic data
- Web search engines
- electronic commence

Beyond Human capabilities

⇒ Adaptivity :- One limiting feature of programming tool is Rigidity.

** • Applications of M.L (Technical elaboration)

H.W Challenges in the field of M.L

* Types of Learning
1) Supervised learning
2) Unsupervised learning
3) Reinforcement learning (Critic Based learning)

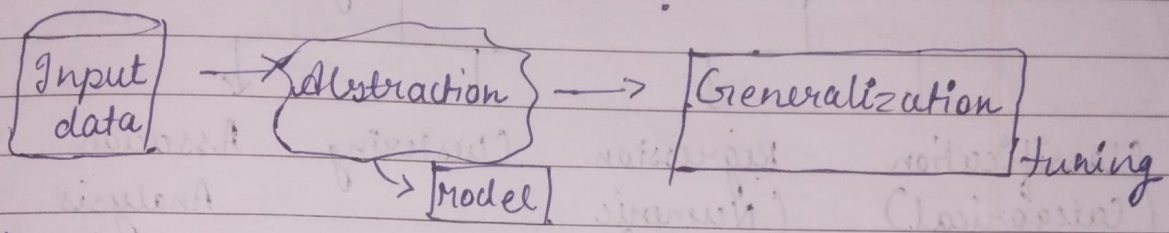
* Challenges
1. Inadequate Training Data
2) Data Quality Issues
3) Model complexity
4) Optimization problems
5) Bias and Fairness
6) Infrastructure costs
7) Ethical AI Governance
8) Digital Divide
9) Societal changes
10) ~~Public~~ Transfer Running

* Limitation :- Multi- Modelling (image, text, audio).

Technical Dfn :- Machine Learning :- A computer program is said to learn from experience E with respect to some ~~to~~ class of tasks T

and performance measure P if its performance at tasks in T as measured by P improves with experience E. (by Tom M. Mitchell)

- How do machine's learn ?



⇒ The basic ML process can be divided into 3 parts : Input data, Abstraction and Generalization.

⇒ ~~Parse~~ Past data / information is utilized as a basis for future decision making.

⇒ The input data is represented in a broader way through the underlane algorithm. ~~and~~

⇒ Generalization : The abstracted representation is generalized to form a framework for making decisions.

Model ——→ ① If / else rules
② Mathematical equations
③ tree / graphs (DS)
④ Logical grouping

* Well-Posed learning Problem :-
→ What is the Problem ? T, P, E
→ Why does the Problem need to be solved ? Benefits
→ How to solve the Problem ?
→ How to do manually ? → Automate.

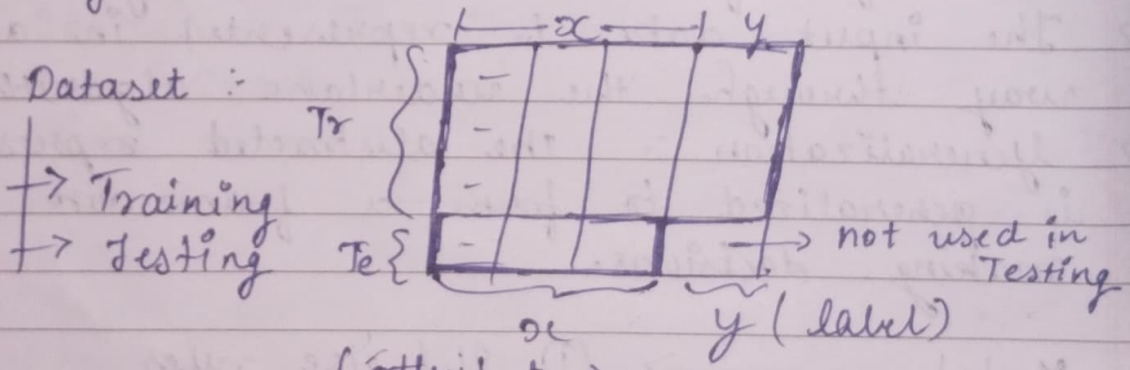
Machine Learning

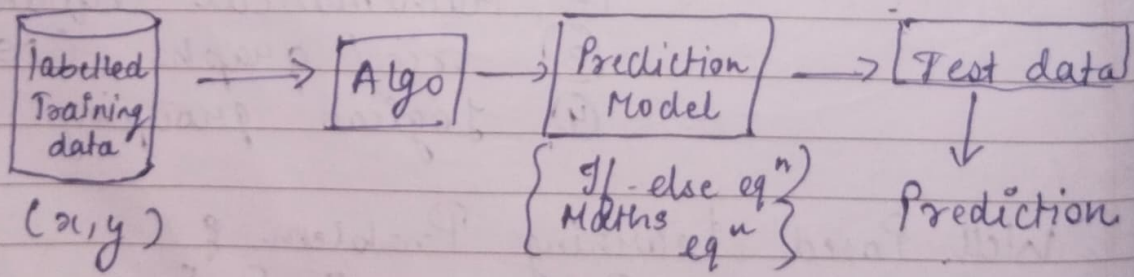
- Supervised Learning
  - Classification (Categorical O/P)
  - Regression (Numeric O/P)
- Unsupervised Learning
  - Clustering
  - Assocaian Analysis
- Reinforcement Learning

- Supervised Learning

(i) Classification :- (Non- numeric O/P)

(ii) Regression :- (Numeric O/P)





Algorithms used :-

1. KNN (K- Nearest Neighbor)
2. Decision Tree
3. Random forest

④ Support Vector Machine
⑤ Naive Bayes Classifier
⑥ N.N (Neuron Network)

* Algorithms :-

1) K- Nearest Neighbour

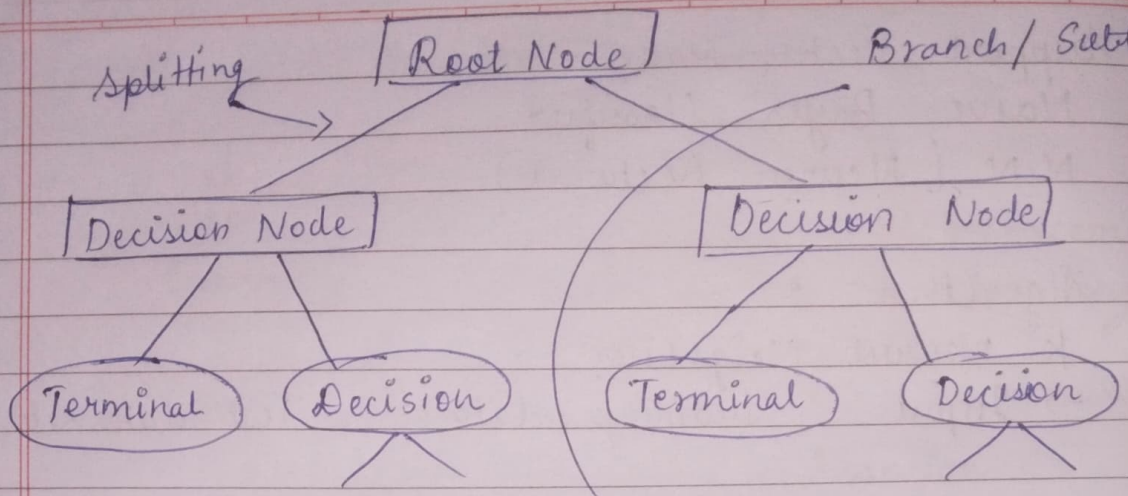
-> Input :- Training dataset ; Test data set, value of k

-> Steps : * Do for all test data points
- -> Calculate the distance of the test data point from the different training data points.
- -> Find the closest K training data points that is training data points whose distances are least from the test data point.
- -> If K=1
  THEN assign class label of the training data point to the test data point.
  Else whichever class label is predominantly present in the training data points, assign that class label to the test data point.

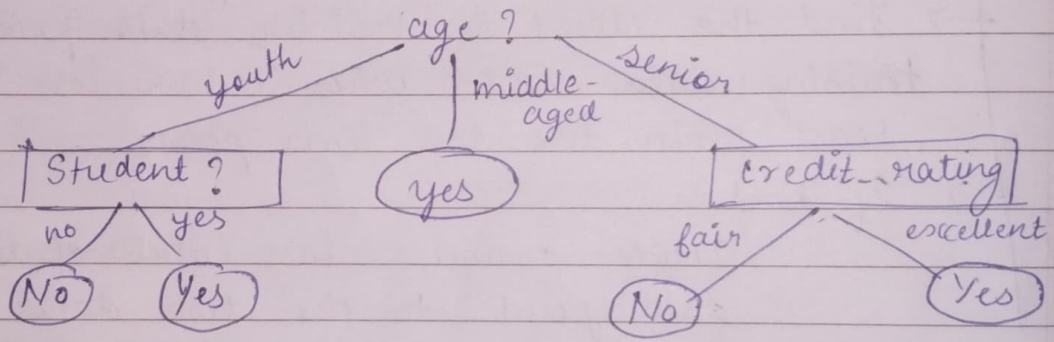
END do.

* learners —— lazy learner -> do not perform task
  \ eager learner -> perform task

2) Decision Tree Terminologies :-
- Root Node
- Decision Node
- Leaf Node
- Sub-Tree
- Pruning (to cut out something)
- Branch / Subtree
- Parent / Child Node.



Q: Whether a person is likely to purchase a computer or not ?



Types of Learning

(i) Supervised :- The computer is presented with example inputs and desired outputs given by a "teacher" and the goal is to learn a general rule that maps inputs to outputs

① Classification :- Eg :- Differentiating between low-risk and high-risk customers from their income and savings i.e categorical output.

② Regression :- Numerical output (Continuous o/p)
Eg :- Price of a used car

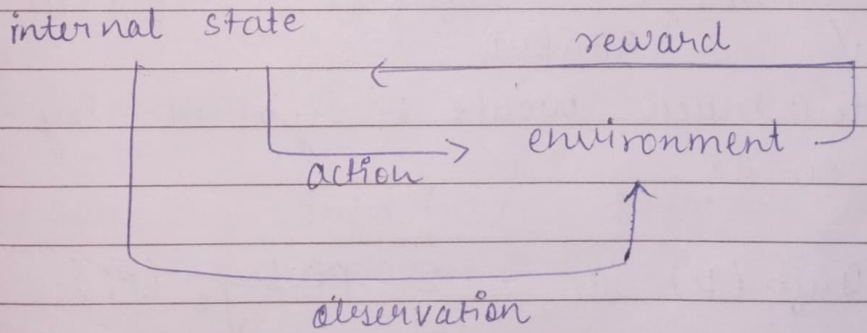
(ii) Unsupervised Learning :- No labels are given to the learning algorithm, leaving on its own to find structure in its input.

- This learning can be a goal in itself or a means towards an end.

① Clustering :- Method of grouping the objects into clusters such that object with most similarities remains into group & has less or no similarities with objs of another group.

② Association :- It is used for finding the relationships between variables in the large database. (Eg :- Market - Basket Analysis)

(iii) Reinforcement Learning :- It is a learning method that interacts with its environment by producing actions and discover error or rewards.



Applications of ML :-

1. Image Recognition
2. Speech "
3. Traffic Prediction
4. Product recommendations
5. Email spam and Malware Filtering
6. Self driving cars

Entropy :- Impurity

☆ Attribute Selection Measure

① Information gain (Entropy of D)
② Gain Ratio
③ Gini Index

=> The expected info needed to classify a tuple in D is given by info (D)

$$Info(D) = -\sum_{i=1}^{m} Pi \log_2(Pi)$$

where Pi is the probability that a tuple in D belongs to class Ci

$$Pi = \frac{|Ci, D|}{|D|}$$

$$Info_A(D) = \sum_{j=1}^{v} \frac{|Dj|}{|D|} \times Info(Dj)$$

$$Gain(A) = Info(D) - Info_A(D)$$

↑ Attribute

How much would be gained by branching on A?

Eg:- $$Info(D) = -\sum_{i=1}^{m} Pi \log_2(Pi)$$

buys_comp = {yes, no}

$$Info(D) = -\sum_{i=1}^{2} Pi \log_2(Pi)$$

$$= -P_{yes} \log_2(P_{yes}) - P_{no} \log_2 P_{no}$$

$$= -\frac{9}{14} \log_2\left(\frac{9}{14}\right) - \frac{5}{14} \log_2\left(\frac{5}{14}\right)$$

High gain → strong attribute $\log_2$

$$= -\frac{1}{14}\left(9\log_2\left(\frac{9}{14}\right) + 5\log_2\left(\frac{5}{14}\right)\right)$$

$$= -\frac{1}{14}\left(9\log_2(0.64) + 5\log_2(0.35)\right)$$

$$= -\frac{1}{14}\left(-5.794 + (-3.21)\right) \quad (-7.57)$$

$$= +0.407 + (+0.227)$$

$$= 0.940 \text{ bits}$$

$$Info_A(D) = \sum_{j=1}^{3} \frac{|Dj|}{D} \times Info(Dj)$$

age = {y, m, s}

1 - y - y (2), n (3)

2 - m - y (4), n (0)

3 - s - y (3), n (2)

$$= \frac{5}{14} \times \left(\frac{-2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right)$$

$$+ \frac{4}{14} \times \left(\frac{-4}{4}\log_2\left(\frac{4}{4}\right) - \frac{0}{5}\log_2\left(\frac{0}{4}\right)\right)$$

$$+ \frac{5}{14} \times \left(\frac{-3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right)$$

$$= 0.694 \text{ bits}$$

Gain (age) = Info (D) - Info age (D)

= 0.940 - 0.694 = 0.246 bits

Find Gain (student), Gain (Income)

$$Info_s(D) = \sum_{j=1}^{2} \frac{|Dj|}{D} \times Info(Dj)$$

student = {y, n}

1 - y - y (6), n (1)

2 - n - y (3), n (4)

$$= \frac{7}{14}\left(-\frac{6}{7}\log_2\frac{6}{7} + \left(-\frac{1}{7}\log_2\frac{1}{7}\right)\right) +$$

$$7/14\left(-3/7\log_2 3/7 + (-4/7\log_2 4/7)\right)$$

:= 0.892

G = 0.1518

☆ Grain Ratio :

$$\text{Split Info}_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{D} \times \log_2\left(\frac{|D_j|}{|D|}\right) \quad \text{— (1)}$$

Min value attribute selected

$$\text{Grain Ratio}(A) = \frac{\text{Grain}(A)}{\text{SplitInfo}(A)} \quad \text{— (2)}$$

☆ Gini Index

$$\text{Gini}(D) = 1 - \sum_{i=1}^{m} P_i^2 \quad \text{— (1)}$$

Min value attribute selected

$$\text{Gini}_A(D) = \frac{|D_1|}{|D|}\text{Gini}(D_1) + \frac{|D_2|}{|D|}\text{Gini}(D_2) \quad \text{— (2)}$$

$$\Delta\,\text{Gini}(A) = \text{Gini}(D) - \text{Gini}_A(D) \quad \text{— (3)}$$

Continuation of the sum :-

$$\text{Info}_{\text{income}}(D) = \sum_{j=1}^{3} \frac{|D_j|}{D} \times \text{Info}(D_j)$$

H — y(2), N(2)

M — y(4), N(2)

L — y(3), N(1)

$$= \frac{4}{14} \times \left(-\frac{2}{4}\log_2\frac{2}{4} + \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right) + \frac{6}{14} \times \left(-\frac{4}{5}\log_2\frac{4}{5} + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right) + \frac{4}{14} \times \left(-\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right)$$

$$= \frac{4}{14}(1.0) + \frac{5}{14}(0.918) + \frac{4}{14}(0.811)$$

$$= 0.911$$

$$\text{Grain} = 0.940 - 0.911$$

$$= 0.029 \quad \text{(Lowest)}$$

$$\log_2(0.5) = \frac{\log(0.5)}{\log 2}$$

Page No.
Date: / /

$$\text{Info}_{CR}(D) = \sum_{j=1}^{2} \frac{|Dj|}{D} \text{Info}(Dj)$$

$$= \frac{8}{14}\left(-\frac{6}{8}\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right)\right) + \frac{6}{14}\left(-\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right)\right)$$

F → Y (6), N (2)

E → Y (3), N (3)

$$= \frac{8}{14}(0.3112 + 0.5) + \frac{6}{14}(0.5 + 0.5)$$

$$= 0.462 + 0.428 = 0.890$$

$$\text{Gain} = 0.940 - 0.890 = 0.048$$

Information Gain for Entropy D are:

Age : 0.246 — Student : 0.1518

Credit_rating : 0.048 — Income : 0.029

$$\text{Gain Ratio (Age)} = \frac{\text{Split Info}}{-\sum_{j=1}^{v} \frac{|Dj|}{D} \times \log_2\left(\frac{|Dj|}{|D|}\right)}$$

$$-\left(\frac{5}{14}\log_2\left(\frac{5}{14}\right) + \frac{4}{14}\log_2\left(\frac{4}{14}\right) + \frac{5}{14}\log_2\left(\frac{5}{14}\right)\right)$$

$$\Rightarrow 1.5774$$

$$\text{Gain Ratio} = \frac{0.246}{1.5774} = 0.1568$$

Split Info (Income)

$$= -\left(\frac{4}{14}\log_2\left(\frac{4}{14}\right) + \frac{6}{14}\log_2\left(\frac{6}{14}\right) + 4/14\log_2(4/14)\right)$$

$$= 0.0188$$

$$\text{Gain Ratio} = \frac{0.029}{0.0188} = 1.542$$

Split Info (student) :-

$$-\left(\frac{7}{14}\log_2\left(\frac{7}{14}\right) + \frac{7}{14}\log_2\left(\frac{7}{14}\right)\right)$$

$$= -0.5 - (-0.5) = 1$$

$$\text{Gain Ratio} = \frac{0.1518}{1} = 0.1518$$

Split Info (credit Rating) :-

$$-\left(\frac{8}{14}\log_2\left(\frac{8}{14}\right) + \frac{6}{14}\log_2\left(\frac{6}{14}\right)\right)$$

$$= -(-0.46 + (-0.53))$$

$$= 0.99$$

$$\text{Gain Ratio} = \frac{0.048}{0.99} = 0.04$$

Gini Index (Age)

$$\text{Gini}(D) = 1 - \left(\left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2\right) + \left(1 - \left(\frac{4}{4}\right)^2 + 0\right) + \left(1 - \left(\frac{3}{5}\right)^2 + \left(\frac{2}{5}\right)^2\right)$$

$$= (1 - 0.16) + (1 - 0.36) + (1 - 0.64)$$

$$= 0.84 + 0.64 + 0.36$$

$$= 1.84$$

~~Gini A (D)~~ =

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

$$= 1 - (0.411 + 0.125)$$

$$= 1 - 0.536$$

$$= 0.464$$

Page No.
Date : / /

Δ Gini (A) = 0.464 - 1.84 = -1.376

Gini Index (Student)

$$= 1 - \left(\left(\frac{6}{7}\right)^2 + \left(\frac{1}{7}\right)^2\right) + \left(1 - \left(\frac{3}{7}\right)^2 + \left(\frac{4}{7}\right)^2\right)$$

$$= 1 - (0.72 + 0.02) + (1 - (0.18 + 0.325))$$

$$= (1 - 0.74) + (1 - 0.505)$$

$$= 0.26 + 0.495 = 0.755$$

$$= \frac{7}{14}(0.26) + \frac{7}{14}(0.490) = 0.367$$

Gini Index (Credit Rating) :-

$$= \frac{8}{14}\left(1 - \left(\left(\frac{6}{8}\right)^2 + \left(\frac{2}{8}\right)^2\right)\right) + \frac{6}{14}\left(1 - \left(\left(\frac{3}{6}\right)^2 + \left(\frac{3}{6}\right)^2\right)\right)$$

$$= 8/14\ (0.375) + 6/14\ (0.5)$$

$$= 0.2143 + 0.2143 = 0.429$$

Gini Index (Income) :-

$$= 4/14\left(1 - \left(\left(2/4\right)^2 + \left(2/4\right)^2\right)\right) + 6/14\left(1 - \left(\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2\right)\right) + \frac{4}{14}\left(1 - \left(\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right)\right)$$

$$= 4/14\ (0.5) + 6/14\ (0.444) + 4/14\ (0.375)$$

$$= 0.440$$

Gini Index (Age) :- $\frac{5}{14}\left(1 - \left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2\right) + \frac{4}{14}(0) + \frac{5}{14}\left(1 - \left(\left(\frac{3}{5}\right)^2 + \left(\frac{2}{5}\right)^2\right)\right)$

$$= \frac{5}{14}(0.48) + \frac{4}{14}\left(1 - \left(\frac{4}{4}\right)^2 + 0\right) + \frac{5}{14}(0.48)$$

$$= 0.1714 + 0 + 0.1714$$

$$= 0.343$$

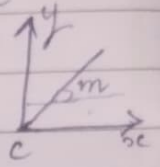
* Pruning :-

Regularization → Prepruning
→ Postpruning

* Equation of straight line $y = mx + c$

m = slope

c = constant



$y = mx + c$

independent variable

dependent

↑ intercept

$y = a + bx$

$$b = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

Eg:- A Teacher beleives that if the grade of Internal exam is high the grade of external exam will also be high. A random sample of 15 students in class were selected. Find the eq^n of regression or find out the model.

Internal exam	15	23	18	23	24	22	22	19	19	16	24	11	24	16	23
External exam	49	63	58	60	58	61	60	63	60	52	62	30	59	49	68

x	y	$x - \bar{x}$	$y - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
15	49	-4.93	-7.8	38.454	24.3049
23	63	3.07	6.2	19.034	9.4249
18	58	-1.93	1.2	2.316	3.7249
23	60	3.07	3.2	9.824	9.4249
24	58	4.07	1.2	4.884	16.5649
22	61	2.07	4.2	8.694	4.2849
22	60	2.07	3.2	6.624	4.2849

X	Y	$x-\bar{x}$	$y-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$
19	63	-0.93	6.2	-5.766	~~24.3049~~ 0.8649
19	60	-0.93	3.2	-2.976	0.8649
16	52	-3.93	-4.8	18.864	15.4449
24	62	4.07	5.2	21.164	16.5649
11	30	-8.93	-26.8	239.324	79.7449
24	59	4.07	2.2	8.954	16.5649
16	49	-3.93	-7.8	30.654	15.4449
23	68	3.07	11.2	34.384	9.4249

$$\bar{x} = \frac{15+23+18+23+24+22+22+19+19+16+24+11+24+16+23}{15} = 19.93$$

$$\bar{y} = 56.8$$

$$\Sigma (x_i-\bar{x})(y_i-\bar{y}) = 429.8$$

$$\Sigma (x_i-\bar{x})^2 = 226.93$$

$$b = \frac{429.8}{226.93} = 1.89$$

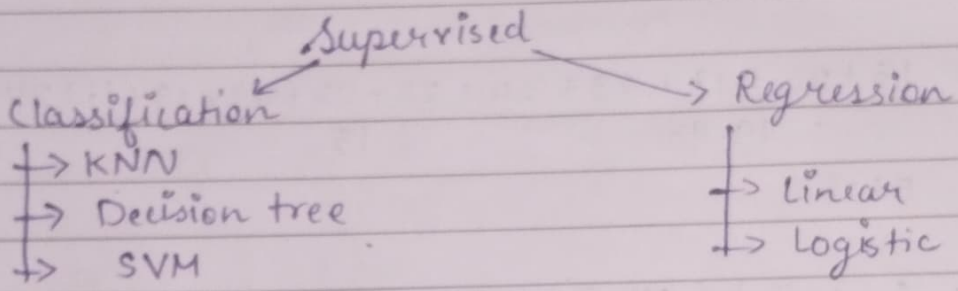
$$a = 56.8 - 1.89\,(19.93)$$

$$= 19.05$$

$$\boxed{y = 19.05 + 1.89x}$$

★ Common Regression Algorithms

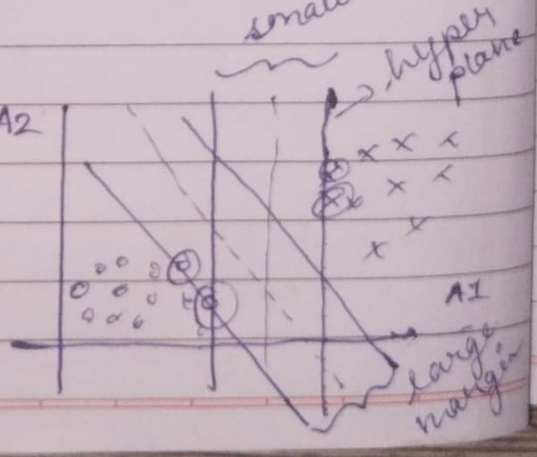
=> Simple linear regression
=> Multiple linear regression
=> Polynomial regression
=> Multivariate adaptive regression splines (MARS)
=> Logistic regression
=> Maximum likelihood estimation (MLE)



★ Support Vector Machines (SVMs):

=> Linearly separable problem
=> Non- " "
let dataset D given as (X_1, Y_1) (X_2, Y_2), (X_3, Y_3) ... $(X_{|D|}, Y_{|D|})$
where Xi is the set of training tuples with associated class labels yi
=> Each yi can take one of two values +1 or -1

=> Hyperplane
=> Maximum Marginal Hyperplane
-> Support vectors
-> Leaf decease problem



Page No.
Date : / /

Hyperplane :- It refers to the decision boundary that we are seeking regardless of the number of input attributes.

=> The hyperplane with largest margin which known as maximum marginal hyperplane (MMH).

=> SVM search for maximum marginal hyper-plane for the problem / to solve the problem.

=> A

★ SVM (contd)

The distance from separating hyperplane to any point on H_1 is $\frac{1}{\|w\|}$

[If $w = \{w_1, w_2 \dots w_n\}$ $\|w\|$

If $w = \{w_1, w_2, \dots w_n\}$ then

$w \cdot w = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$.

$\|w\|$ is the euclidean norm of w, $\sqrt{w \cdot w}$

∴ Maximum Margin is $\frac{2}{\|w\|}$

=> The case when data are linearly inseparable (Non-linear)

Use of Kernel functions :-

increasing dimensionality

1) Polynomial Kernel
2) Graussion basis function Kernel
3) Sigmoid Kernel

★ Model Training

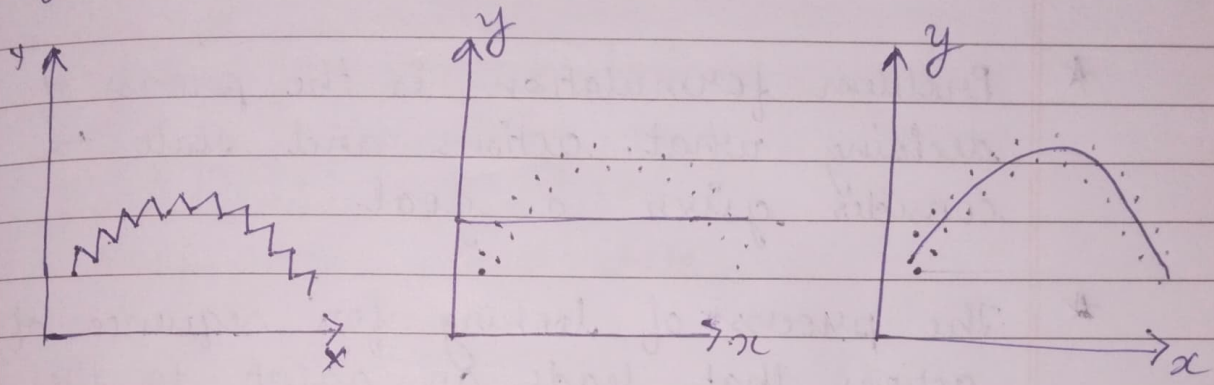
- Bias (error)
- Variance
- Overfitting
- Underfitting
- Bias - Variance Trade Off

• Bias : Bias is the difference between the average prediction of our model and the correct value which we are trying to predict.

low Bias low Variance (for Ideal model)

- Variance : Model with high variance face a lot of attention to training data and does not generalize on the data which it has not seen before.

=> Variance is the variability of the model prediction for a given data point which tells us spread of our data



Overfitting Underfitting Good Balance