DIVYA. S

UNIT - 1 : INTRODUCTION TO MACHINE LEARNING

1. MACHINE LEARNING:

→ Ability to automatically learn and improve from experience without being explicitly programmed.

→ Focuses on development of computer programs that can access data and use it to learn for themselves.

→ The process of learning begins with observations of data such as examples, direct experience or instruction in order to look for patterns in data and make better decisions in the future based on examples that we provide.

1.1 EXAMPLES OF ML:

a) IMAGE RECOGNITION:

→ In Black & White image, each pixel is served as one of the measurements.

→ In colored images, each pixel provides 3 measurements of intensities in 3 different colors (RGB)

→ Also handwritten and printed letter classification can be done.

b) SPEECH RECOGNITION:

→ Translation of spoken words into text.

→ Example: Apple's Siri, Amazon Alexa, Microsoft's Cortana, Google Assistant, Nuance's Dragon Assistant.

Scanned by CamScanner

c) MEDICAL DIAGNOSIS:

→ Therapy planning & patient monitoring.

d) STATISTICAL ARBITRAGE:

→ Stock market

e) LEARNING ASSOCIATION

f) CLASSIFICATION:

→ Before a bank decides to distribute loans, it assesses the customers on their ability to pay loans

g) PREDICTION:

→ Loan example : To compute probability of a fault, the system will need to classify the available data in groups. It is defined by a set of rules prescribed by analysis.

h) EXTRACTION:

→ Getting structured information from unstructured data. Ex: Web pages, Articles, Blogs, Business reports and emails.

1.2 METHODS OF ML:

1.2.1 → Supervised Learning

1.2.2 → Unsupervised Learning

1.2.3 → Semi Supervised Learning.

Scanned by CamScanner

1.2.1 SUPERVISED LEARNING:

→ Supervised Machine Learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events.

→ Starting from the analysis of a known training dataset, the learning algorithms produce an inferred function to make predictions about the output values.

→ The System is able to provide targets for any new input after sufficient training.

→ The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly

USEFULNESS:

1.2.1.1 Classification problems ask the algorithm to predict the discrete values, identifying the input data as the member of a particular class or group.

1.2.1.2 Regression problems look at continuous data like linear regression - Given 'x' value and finding 'y' value.

↳ Example: An algorithm that predicts the price of an apartment in Shozhinganallur based on square footage, location and proximity to public transport.

Scanned by CamScanner

1.2.2 UNSUPERVISED LEARNING:

→ It is used when the information used to train is neither classified nor labeled.

→ It studies how the system can infer a function to describe a hidden structure from unlabeled data.

→ The System doesnot figure out the right output but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

USEFULNESS:

1.2.2.1 Clustering - To look at a collection of bird photos and separate them roughly by species relying on cues like feather's color, size or beak shape.

1.2.2.2 Anamoly Detection - Bank detect fraudlent transaction by looking for unsual patterns in customer purchasing behavior.

↳ Example: Same credit card is used in two different locations.

1.2.2.3 Association - Amazon shows relevant items of your purchase pattern.

1.2.3 SEMI SUPERVISED LEARNING:

→ The learning method have less labeled and more unlabeled.

→ Improve learning accuracy

→ It is chosen when acquired labeled data requires skilled and relevant resources in order to train it /learn from it.

Scanned by CamScanner

USEFULNESS:

↳ Useful when extracting relevant features from the data is difficult & labeling examples is a time intensive task.

Example: To label all the Scan/x-ray images of tumor

1.2.4 REINFORCEMENT LEARNING

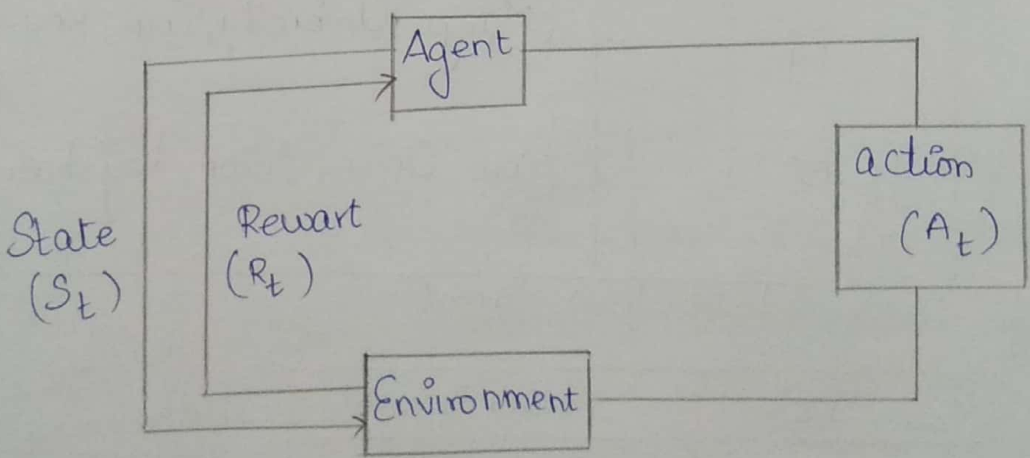
→ Interacts with its environment by producing actions and discovers error or rewards.

→ Trial & Error method and delayed reward are the most relevant characteristics of reinforcement learning.

→ This method allows machines & software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance.

→ Simple reward feedback is required for the agent to learn which action is best.

→ Example: video games.



Scanned by CamScanner

2. LEARNING ASSOCIATION RULE:

→ Learning Association Rule basically means finding the items that are purchased together more frequently than others.

→ Used for data mining.

→ Example for learning association rule is Apriori algorithm.

→ Suppose you have records of large number of transactions at a shopping centre as follows.

Transaction ID	Items Bought.
T1	{Mango, Onion, Net, keychain, Eggs, Yo-Yo}
T2	{Doll, Onion, Net, keychain, Eggs, Yo-Yo}
T3	{Mango, Apple, key chain, Eggs}
T4	{Mango, Umbrella, Corn, Key-chain, Yo-Yo}
T5	{Corn, Onion, Onion, keychain, Icecream, Egg}

Change table into below format:

Tr. ID	Items Bought
T1	{M O N K E Y}
T2	{D O N K E Y}
T3	{M A K E}
T4	{M U C K Y}
T5	{C O O K I E}

Scanned by CamScanner

STEP 1: Count the number of transactions in which each item occurs. Note: O = Onion is bought 4 times in total, but it occurs only in 3 transactions.

Items	No. of transactions
M	3
O	3
N	2
K	5
E	4
Y	3
D	1
A	1
U	1
C	2
I	1.

STEP 2: From the above table remove the items that has frequency less than '3' times.

Items	No. of trans.
M	3
O	3
K	5
E	4
Y	3

STEP 3: Start making pairs like MO, MK, ME, MY and then start with the second item with other items. OK, OE, OY and KE, KY, EY

Scanned by CamScanner

STEP 4 : Count how many items each pair is bought together. (See from the initial table).

Item pairs	No. of trans
MO	1
MK	3
ME	2
MY	2
OK	3
OE	3
OY	2
KE	4
KY	3
EY	2

STEP 5 : Remove the item pairs that has frequency less than 3. (from the above table)

Item pairs	No. of trans.
MK	3
OK	3
OE	3
KE	4
KY	3

STEP 6 : From step 5 make three combinational pairs

i.e., From OK and OE ⇒

From KE and KY ⇒

Item pair	No. of tran.
OKE	3 ✓
KEY.	2.

STEP 7 : From the above, we can conclude that OKE are the three items bought together.

× × ×

Scanned by CamScanner

3. REGRESSION :

→ Regression is basically a statistical approach to find the relationship between variables.

→ In Machine learning, this is used to predict the outcome of an event based on the relationship between variables obtained from the data set.

→ Linear regression is one type regression used in Machine Learning.

→ Regression is a technique from statistics that is used to predict values of a desired target quantity when the target quantity is continuous.

Example:

x	y
1	2
2	4
3	5
4	4
5	5

Line, $y = mx + c$

$$\text{Slope} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

$\bar{x}$ = mean of x
$\bar{y}$ = mean of y

$\bar{x} = 3$ $\bar{y} = 4$

Find 'y' value when $x = 6$

Scanned by CamScanner

STEP 1: Find the slope with the given dataset.

$$\text{Slope} = \frac{(1-3)(2-4)+(2-3)(4-4)+(3-3)(5-4)+(4-3)(4-4)+(5-3)(5-4)}{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}$$

$$= \frac{6}{10} = 0\cdot 6$$

Slope $= 0\cdot 6$

STEP 2:

$$c = \bar{y} - m\bar{x}.$$

$$c = 4 - (0.6)(3)$$

$$c = 2\cdot 2$$

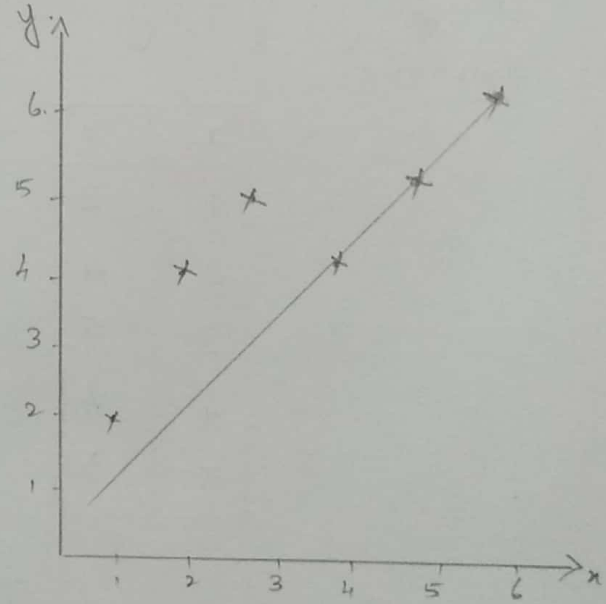
STEP 3:

$$y = mx + c$$

Here, $x = 6$ $y = ?$

$$y = (0\cdot 6)(6) + 2\cdot 2$$

$$y = 5\cdot 8$$



$\therefore$ $Y = 5\cdot 8$ when $X = 6$

XXX

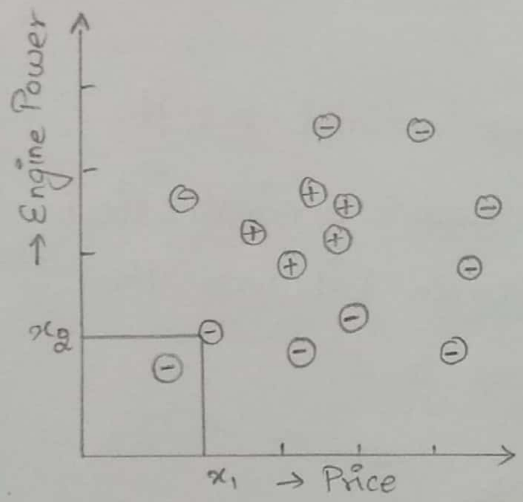
Scanned by CamScanner

4. HYPOTHESIS :

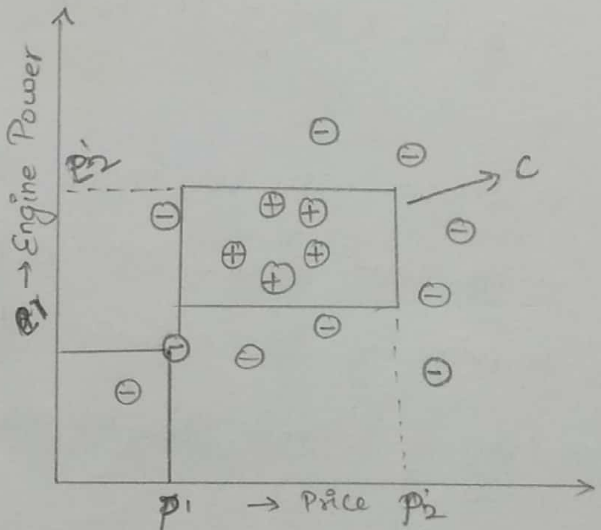
LEARNING A CLASS FROM EXAMPLES: Two class Problem

→ We are given a training set of labeled examples (positive and negative) and want to learn a classifier that we can use to predict unseen examples, or to understand the data

→ Input representation: we need to decide what attributes (features) to use to describe the input patterns. This implies ignoring other attributes as irrelevant.



"Training set for a family car"



"Hypothesis class"

$$(p_1 \leq \text{price} \leq p_2) \text{ AND } (e_1 \leq \text{engine power} \leq e_2)$$

where $p_1, p_2, e_1, e_2 \in R$.

Scanned by CamScanner

→ Training Set : $X = \{(x_n, y_n)\}_{n=1}^{N}$

where $x_n \in R$ is the n^{th} input vector.

$y_n \in \{0, 1\}$ → class label.

→ Hypothesis (model) class H : The set of classifier functions we will use. Ideally, the true class distribution C can be represented by a function in H (exactly, or with a small error).

→ Having selected H, learning the class reduces to finding an optimal $h \in H$. Approximate error by empirical.

$$E(h; X) = \sum_{n=1}^{N} (h(x_n) \neq y_n)$$ → Number of misclassified instances.

→ There may be more than one optimal $h \in H$.

→ In that case, we achieve better generalization by maximizing the margin (the distance between the boundary of h and the instances closest to it).

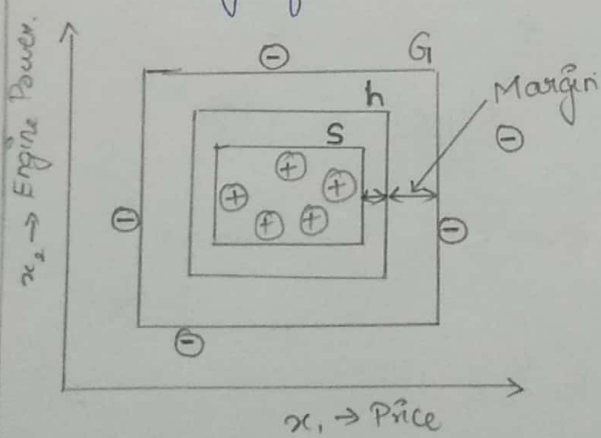


Fig 1: The hypothesis with the largest margin

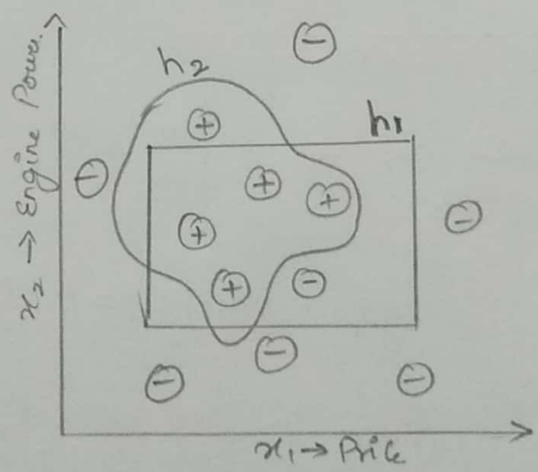


Fig 2: Noise and a more complex hypothesis

Scanned by CamScanner

In Fig 1:

→ 'S' represents the most specified class which have only the positive data sets.

→ 'h' represents the randomly taken Hypothesis

→ 'G' represents Generalised class which will lie inside the boundary of negative data set. Not even a -ve data will come in this region. Also, no +ve is left out.

In Fig 2:

→ In this, there are two hypothesis region.

→ h_1 → Simple and little noisy hypothesis, as it encloses a negative data in it.

→ h_2 → Complex and noiseless hypothesis.

→ In general, any hypothesis we take should be simple.

What is Noise?

→ Noise is any unwanted anomaly in the data.

Scanned by CamScanner

→ It can be due to the following:

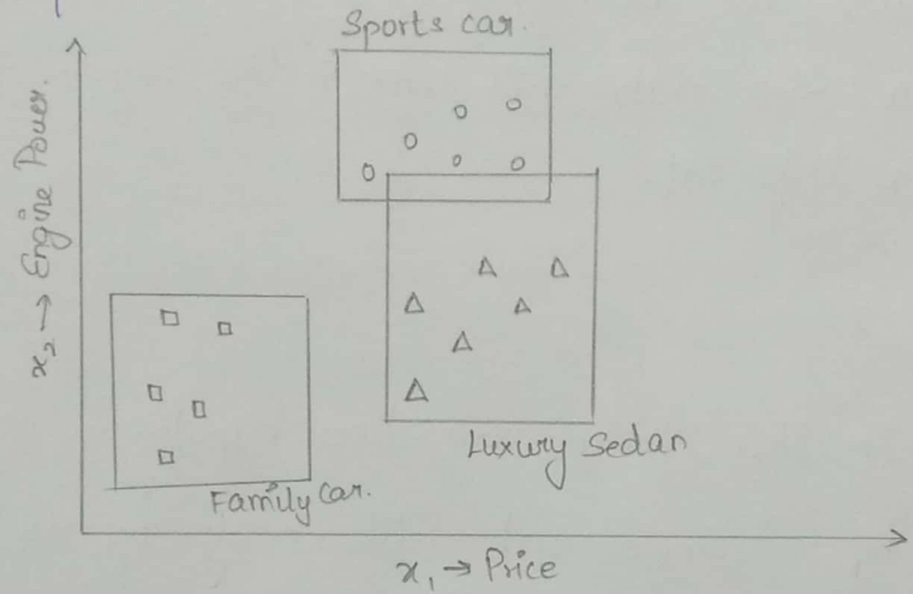
↳ Imprecision in recording the input attributes : x_n

↳ Errors in labeling the input vectors : y_n.

↳ Attributes not considered that affects the label.

→ Noises makes learning harder.

LEARNING MULTIPLE CLASSES:



In Multiple class learn, the hypothesis will have more than one datapoints such as

$h(x) \Rightarrow h(x_1) \quad h(x_2) \quad h(x_3)$

x_1 → Family Car

x_2 → Luxury Sedan

x_3 → Sports Car.

× × ×

Scanned by CamScanner

5. PAC LEARNING:

→ Probably Approximately Correct Learning

→ Normally, we cannot expect a learner to learn a concept exactly.

→ Also we cannot expect to learn a close approximation to the target concept.

→ Therefore, the only realistic expectation of a good learner is that with high probability it will learn a close approximation to the target concept.

→ In PAC Learning, one requires that given small parameters ε and δ, with probability at least $(1-\delta)$ a learner produces a hypothesis with error at most ε.

5.1 Approximately (error rate) Learning:

$$\varepsilon = 0 \leq \varepsilon \leq 0.5$$

5.2 Probability Learning: (failure)

$$\delta = 0 \leq \delta \leq 0.5$$

When, $\boxed{P(c \oplus h) \leq \varepsilon}$ // we can say it is approximately correct learning.

Also when, $\boxed{P_r(P(c \oplus h)) \leq \varepsilon) \geq \underbrace{1-\delta}_{\text{Succes.}}}$ // we can say it is Probably Approximately Correct learning model.

Scanned by CamScanner

U1

5.3 OCCAM'S RAZOR : ⁂ 2m.

claim : The probability that there exists a hypothesis $h \in H$ that

i) is consistent with 'm' examples.

ii) satisfies error $(h) > \varepsilon$.

$$m > \frac{1}{\varepsilon}\left\{\ln(|H|) + \ln(1/\delta)\right\}$$

× × ×

6. MODEL SELECTION AND GENERALIZATION:

→ ML problems (Classification, Regression and Others) are typically ill-posed :- the observed data is finite and does not uniquely determine the classification or regression function.

→ In order to find a unique solution and learn something useful, we must take assumptions.

→ We can always enlarge the class of functions that can be learned by using a larger hypothesis class H.

→ The goal of ML is not to replicate the training data, but to predict unseen data well i.e., generalize.

→ For best generalization, we should match the complexity of the hypothesis class H with the complexity of the function underlying the data.

Scanned by CamScanner

→ If H is less complex, it is underfitting and if H is more complex, it is overfitting.

→ In ML algorithms there is a tradeoff between 3 factors:

↳ The complexity $c(H)$ of the hypothesis class.

↳ The amount of training data N.

↳ The generalization error E.

CROSS VALIDATION:

→ Divide the available data set into three parts.

Training Set:

→ Used to train. i.e, to fit a hypothesis $h \in H$.

→ Usually done with an optimization algorithm.

Validation Set:

→ Used to minimize the generalization error.

→ Usually done with a "grid search".

Test Set:

→ Used to report the generalization error.

→ We optimize nothing on it, we just evaluate the final model it.

×××

Scanned by CamScanner

7. DIMENSIONS OF A SUPERVISED ML ALGORITHM:

We have a Sample $X = \{(x_n, y_n)\}_{n=1}^{N}$ (usually independent and ideally distributed) drawn from an unknown distribution.

We want to learn a useful approximation to the underlying function that generated the data.

We must choose

→ A model $h(x;\Theta)$
where Θ → parameter which determines a particular hypothesis in the class.

→ A loss function $L(\ )$ to compute the difference between the desired output (y_n) i.e., label and our prediction to it $h(x_n;\Theta)$.

$$E(\Theta; X) = \sum_{n=1}^{N} L(y_n, h(x_n;\Theta)) = \text{Sum of errors over instances.}$$

where y_n → Label

$h(x_n;\Theta)$ → model

L → Loss function

Scanned by CamScanner

→ Different algorithms differ in any of these choices.

→ The model, loss and learning algorithm are chosen by the ML system designer so that:

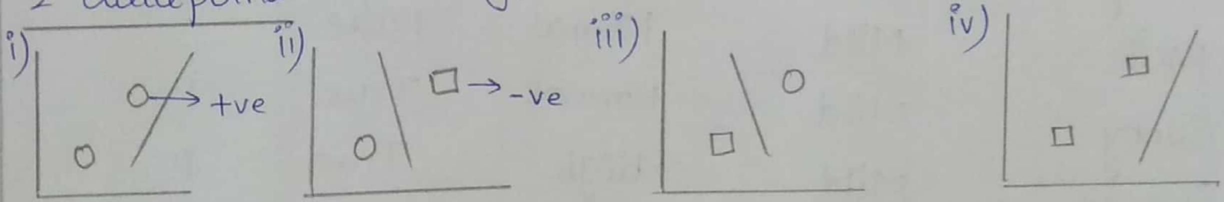
↳ The model class is large enough to contain a good approximation to the underlying function that generated the data in X in a noisy form.

↳ Learning algorithm is efficient and accurate.

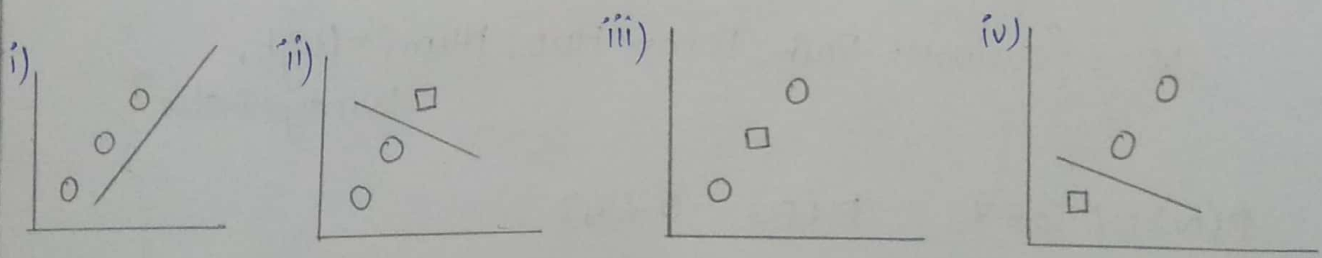
↳ We must have sufficient training data to pinpoint the right model.

VC DIMENSION:

2 datapoints: Line hypothesis can satisfy all four cases.



3 datapoints: Line hypothesis cannot satisfy all the cases. In some cases it may fail.



In case (iii) we cannot classify positive and negative data using a single line hypothesis.

× × ×

Scanned by CamScanner

OTHER QUESTIONS:

1) Predicting a class label using Naive Bayesian Classification (UNIT-2)

Outlook	Temperature	Humidity	Windy	Class
Sunny	Hot	High	False	N
Sunny	Hot	High	True	N
Overcast	Hot	High	False	P
Rain	Mild	High	False	P
Rain	Cool	Normal	False	P
Rain	Cool	Normal	True	N
Overcast	Cool	Normal	True	P
Sunny	Mild	High	False	N
Sunny	Cool	Normal	False	P
Rain	Mild	Normal	False	P
Sunny	Mild	Normal	True	P
Overcast	Mild	High	True	P
Overcast	Hot	Normal	False	P
Rain	Mild	High	True	N

X = {outlook = Rain, Temp = Hot, Humi = High, Windy = False}

Sol:

P(N) = 0.357 P(P) = 0.643

P(outlook=Rain/N) = 0.4 P(outlook=Rain/P) = 0.33

P(Temp = Hot/N) = 0.4 P(Temp=Hot/P) = 0.22

P(Humi = High/N) = 0.8 P(Humi=High/P) = 0.33

P(Windy = False/N) = 0.4 P(Windy=False/P) = 0.67

P(X/N) = 0.0512 P(X/P) = 0.01605

P(X/N) P(N) = 0.01827 P(X/P) P(P) = 0.0103

Scanned by CamScanner

2) Predicting a class label using Naive Bayesian classification. (UNIT-2)

X = {Owns Home = Yes, Married = No, Gender = Female, Employed = Yes}

Owns Home	Married	Gender	Employed	Risk class.
Yes	Yes	Male	Yes	B
No	No	Female	Yes	A
Yes	Yes	Female	Yes	C
Yes	No	Male	No	B
No	Yes	Female	Yes	C
No	No	Female	Yes	A
No	No	Male	No	B
Yes	No	Female	Yes	A
No	Yes	Female	Yes	C
Yes	Yes	Female	Yes	C

Scanned by CamScanner

OTHER QUESTIONS : REGRESSION

1.

(year) x	2005	2006	2007	2008	2009
(sales) y	12	19	29	37	45

$\bar{x} = 2007$ $\bar{y} = 28.4$

Slope = 8.4

2.

Height (In Inches)	Weight (In kg)
65	105
65	125
62	110
67	120
69	140
65	135
61	95
67	130

What is the predicted weight of a person who is 71 inches tall?

Scanned by CamScanner

OTHER PROBLEMS: (APRIORI ALGORITHM)

1)

T. ID	List of Item.
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

Scanned by CamScanner