Name: Abhishek. Milind. Patwardhan
ID : 811271359

Data Science II   HomeWork 2

Q.1] Linear Models

1) Training loss:

$$L(D,W) = \frac{1}{2} \sum_{(x,y \in D)} (y - f(x))^2 = \frac{1}{2} \sum_{(x,y \in D)} (y - W^T \cdot x)^2$$

Please write Pseudo code for stochastic gradient descent for training the linear model. Please provide the detailed math formula of gradient computation as well as how to update w.

→ Pseudo code for Stochastic Gradient Descent is :

```
initialize w randomly
set learning rate α
Set number of epochs T
for t=1 to T do
    for each example (x,y) in D do
        // compute gradient of the loss w.r.t. w
        gradient = -(y - wᵀ.x).x

        // update w using the gradient and learning rate
        w = w - α*gradient
    end for
end for
```

- Compute Gradient:

The gradient of the loss function with respect to w is computed for each training example (x, y) using the formula:

$$\nabla_w L(x, y, w) = -(y - w^T \cdot x) \cdot x$$

- Update weight

After computing the gradient, update the weight vector w using the update rule of SGD.

$$w_{t+1} = w_t - \alpha \cdot \nabla_w L(x, y, w)$$

Here, W_{t+1} = updated weight
W_t = current weight
α = learning rate
$\nabla_w L(x, y, w)$ = gradient of the loss function.

Name : Abhishek. Milind . Patwardhan
ID : 811271359

Data Science II HomeWork - 2

Q.2] Naive Baye's Classifiers.

No	Outlook	Temperature	Humidity	Play Golf
1	Sunny	Hot	High	N
2	Sunny	Hot	High	N
3	overcast	Hot	High	Y
4	rain	Mild	High	Y
5	rain	Cool	Normal	N
6	rain	Cool	Normal	N
7	Overcast	Cool	Normal	Y
8	Sunny	Cool	Normal	Y
9	Sunny	Mild	Normal	Y
10	Sunny	Mild	High	?

1) classify instance No. 10 using Naive Bayes classifier Include details of your NBC, probability calculations and how the final classification is decided.

→ Calculating the probabilities of "Yes" & "No"

$P(\text{Play Golf} = 'Y' = \text{"Yes"}) = \frac{5}{9} = 0.556$

$P(\text{Play Golf} = 'N' = \text{"No"}) = \frac{4}{9} = 0.444$

Calculating probabilities of different attributes for or with respect to target attribute

Outlook	Yes	No
Sunny	2/5	2/4
Overcast	2/5	0
rain	1/5	2/4.

Temperature	Yes	No
Hot	1/5	2/4
Mild	2/5	0
cool	2/5	2/4

Humidity	Yes	No
High	2/5	2/4
Normal	3/5	2/4

The above probabilities are calculated considering the attribute value & different target values i.e example consider outlook = "Sunny"
Then $\frac{2}{5}$ is the probability that out of all yes only 2 corrosponds to ~~Yes~~ Play Golf = Yes when the outlook is sunny.

New instance is,
(Outlook = Sunny, Temperature = Mild, Humidity = High)

Calculating Probabilities using Naive Bayes classifier

$V_{NB} = \text{argmax } P(v_j) \prod_i P(a_i | v_j)$

$= \underset{v_j \in \{Yes, No\}}{\text{argmax}} P(v_j) \quad P(\text{Outlook} = \text{sunny} | v_j) \times P(\text{Temperature} = \text{Mild} | v_j) \times P(\text{Humidity} = \text{High} | v_j)$

$V_{NB}(Yes) = P(Yes) . P(\text{sunny} | Yes) . P(\text{Mild} | Yes) . P(\text{High} | Yes)$

$= \frac{5}{9} \times \frac{2}{5} \times \frac{2}{5} \times \frac{2}{5} = 0.0355$

$V_{NB}(No) = P(No) . P(\text{sunny} | No) . P(\text{Mild} | No) . P(\text{High} | No)$

$= \frac{4}{9} \times \frac{2}{4} \times 0 \times \frac{2}{4} = 0$

- Calculating the Normalization Probabilities

$V_{NB}(Yes) = \frac{V_{NB}(Yes)}{V_{NB}(Yes) + V_{NB}(No)} = \frac{0.035}{0.035 + 0} = \underline{\underline{1}}$

$V_{NB}(No) = \frac{V_{NB}(No)}{V_{NB}(No) + V_{NB}(Yes)} = \frac{0}{0 + 0.035} = 0$

Probability of "Yes" is more than Probability of "No"
Hence the new instance is classified as "Yes"

2) What is the time complexity for training and testing Naive Bayes classifier, respectively?

→

- The time complexity for training a Naive Bayes classifier is generally O(nd), where 'n' is the number of samples in the training set and d is the number of features.
- The complexity is achieved from computing the probabilities for each feature and class combination.

- For testing, the time compexity is O(md), where 'm' is the number of samples in the test set. This complexity comes from applying the trained model to each sample in the test set and computing the posterior probabilities for each class given the features.

3) After a yearly checkup for a software developer, there are both bad news and good news from the doctor. The bad news is that developer has a test result positive for a disease, & the test is 98% accurate (i.e., if you have the disease, then the probability of testing positive is 0.98; if you do not have the disease, the probability of testing negative is also 0.98). The good news is that this is a rare disease, because only 1 in 20,000 people will have it. What are the chances that the developer actually has disease?

→ Given that the test is 98% accurate,

$\therefore$ P(Positive | Disease) = P(Pos | Dis) = 0.98

P(Negative | Non-Disease) = P(Neg | Non-Dis) = 0.98

The disease is rare because only 1 in 20,000

$\therefore$ P(Diseases) = P(Dis) $= \frac{1}{20{,}000} = 0.00005$

By using Bayes theorem,

we have to find the probability of developer having disease

$\therefore$ $P(\text{Disease} \mid \text{Positive}) = \frac{P(\text{Pos} \mid \text{Dis}) \times P(\text{Dis})}{P(\text{Pos})}$

But, $P(\text{Pos}) = P(\text{Pos} \mid \text{Dis}) . P(\text{Dis}) + P(\text{Pos} \mid \text{Non-Dis}) . P(\text{Non-Dis})$

$= (0.98) \times (0.00005) + (0.02) \times (0.99995)$

$= 0.000049 + 0.019999$

$= 0.020048$

$\therefore$ $P(\text{Disease} \mid \text{Positive}) = \frac{(0.98) \times (0.00005)}{0.020048}$

$P(\text{Disease} \mid \text{Positive}) = 0.00244$

Hence the chances that the developer actually has the disease are 0.00244.

Name: Abhishek Milind Patwardhan
ID : 811 27 13 59

Data Science II   HomeWork - 2

Q.3] **Decision Trees** :

S = "Small"   L = "Large"   M = "Medium"

No.	Posts	Friends	Photo	Real - Account
1	S	S	No	No
2	S	L	Yes	Yes
3	L	M	No	Yes
4	M	M	Yes	Yes
5	L	M	Yes	Yes
6	M	L	No	Yes
7	M	S	No	No
8	L	M	No	Yes
9	M	S	No	No
10	S	S	Yes	Yes

1) Compute the information Gain if we first choose "Friends" as the attribute to split Data.

→ Values (Friends) = Small, Medium, Large.

S = [Yes = 7 , No = 3]

S : Whole Data Set

$$\therefore \text{Entropy}(S) = -\frac{7}{10}\log_2\left(\frac{7}{10}\right) - \frac{3}{10}\log_2\left(\frac{3}{10}\right)$$

$$\text{Entropy}(S) = 0.8806$$

S_{small} = [Yes = 1, No = 3]

$$\therefore \text{Entropy}(S_{small}) = -\frac{1}{4}\log\left(\frac{1}{4}\right) - \frac{3}{4}\log\left(\frac{3}{4}\right)$$

$$= -\frac{1}{4}[-4] - \frac{1}{4}[\log 1 - \log 4] - \frac{3}{4}[\log 3 - \log 4]$$

$$= 0.811$$

$S_{Medium} = [\text{Yes} = 4, \text{No} = 0]$

$\therefore \text{Entropy}(S_{Medium}) = -\frac{4}{4}\log_2\left(\frac{4}{4}\right) - \frac{0}{4}\cdot\log_2\left(\frac{0}{4}\right)$

$= 0.$

$S_{Large} = [\text{Yes} = 2, \text{No} = 0].$

$\therefore \text{Entropy}(S_{Large}) = -\frac{2}{2}\log_2\left(\frac{2}{2}\right) - \frac{0}{2}\cdot\log_2\left(\frac{0}{2}\right)$

$= 0$

$\text{Gain}(S, \cancel{\text{Small}}\ \text{Friends}) = \text{Entropy}(S) - \sum_{v \in \text{Small, Medium, Large}} \frac{|S_v|}{|S|} \cdot \text{Entropy}(S_v)$

$= \text{Entropy}(S) - \frac{4}{10} \times (0.8811)\ \frac{4}{10} \times 0 - \frac{2}{10} \times 0.$

$\text{Gain}(S, \text{Friends}) = 0.8806 - \cancel{0.4}\ 0.3244 = \cancel{0.4806}.\ 0.5562$

$\therefore$ Information Gain of "Friends" attribute = ~~0.4806~~ 0.5562

2) Construct a decision tree from the given data. Show the computation steps

→ We first need to find the Information Gain of all attributes i.e "Posts", "Photo", and "Real-Account"

We ~~are~~ have already calculated the Information Gain of "Friends" attribute & will be using the same.

• Computing Information Gain of "Posts" attribute.

Values (Posts) = Small, Medium, Large.

Entropy (S) = 0.8806.

S_{small} = (Yes = 2, No = 1)

$\therefore$ Entropy (S_{small}) = $-\frac{2}{3} \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right)$

= 0.918

S_{Medium} = [Yes = 2, No = 2]

$\therefore$ Entropy (S_{Medium}) = $-\frac{2}{4} \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \log_2\left(\frac{2}{4}\right)$

= $-2 \times \frac{2}{4} \log_2\left(\frac{2}{4}\right)$ = 1

S_{large} = [Yes = 3, No = 0)

Entropy = 0

$$IG(S, Posts) = Entropy(S) - \sum_{v \in S,M,L} \frac{|S_v|}{|S|} \times Entropy(S_v)$$

$$= 0.8806 - \frac{3}{10} \times 0.918 - \frac{4}{10} \times 1 - \frac{3}{10} \times 0$$

$$= 0.8806 - 0.2754 - 0.4 = 0.2052$$

Computing Information Gain of "Photo" attribute

values (Photo) = Yes, No.

Entropy (S) = 0.8806.

S_{yes} = [Yes = 4, No = 0]

Entropy (S_{yes}) = 0

S_{NO} = [Yes = 3, No = 3]

$$Entropy(S_{NO}) = -\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\cdot\log_2\left(\frac{3}{6}\right)$$

$$= 1$$

$$\therefore IG(S, Photo) = Entropy(S) - \sum_{v \in Yes/No} \frac{|S_v|}{|S|} \times Entropy(S_v)$$

$$= 0.8806 - \frac{4}{10} \times 0 - \frac{6}{10} \times 1$$

$$= 0.8806 - 0.6$$

$$= 0.2806$$

$\therefore$ Gain (Friends) = 0. ~~4806~~ 5562 (Max).
Gain (Posts) = 0.2052
Gain (Photo) = 0.2806

$\therefore$ We will consider "friends" as the Root Node, because it is having maximum Gain.

NO	Posts	Photo	Real Account
1	S	No	No
7	M	No	No
9	M	No	No
10	S	Yes	Yes.

- Now calculating the Gain of "Posts" attribute

value (Post) = small, Medium.

S_{small} = Yes = 1, No = 1

$\therefore E(S_{small}) = -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)$

$= 1$

S_{Medium} = Yes = 0, No = 2

$\therefore$ Entropy (S_{Medium}) = 0.

~~$\therefore$ I.G (Posts) = Entrop~~

Entropy ($S_{small \to Friends}$) = Yes = 1, NO = 3

$\therefore = -\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4}$

$= 0.8112$

$\therefore$ IG (Posts) = E ($S_{small \to Friends}$) $- \sum_{v \in S,M}$ . . .

$= 0.8112 - \frac{2}{4} \times 1 - \frac{2}{4} \times 0.$

I.G (Post) = 0.3112

• Now calculating the Gain of "Photo" attribute.

~~V~~ values (Photo) = Yes, No.

$E(S) = 0.8112$

$\therefore S_{Yes} = [Yes = 1, No = 0]$

$\therefore E(S_{Yes}) = -\frac{1}{1}(\log_2(1)) - 0 = 0$

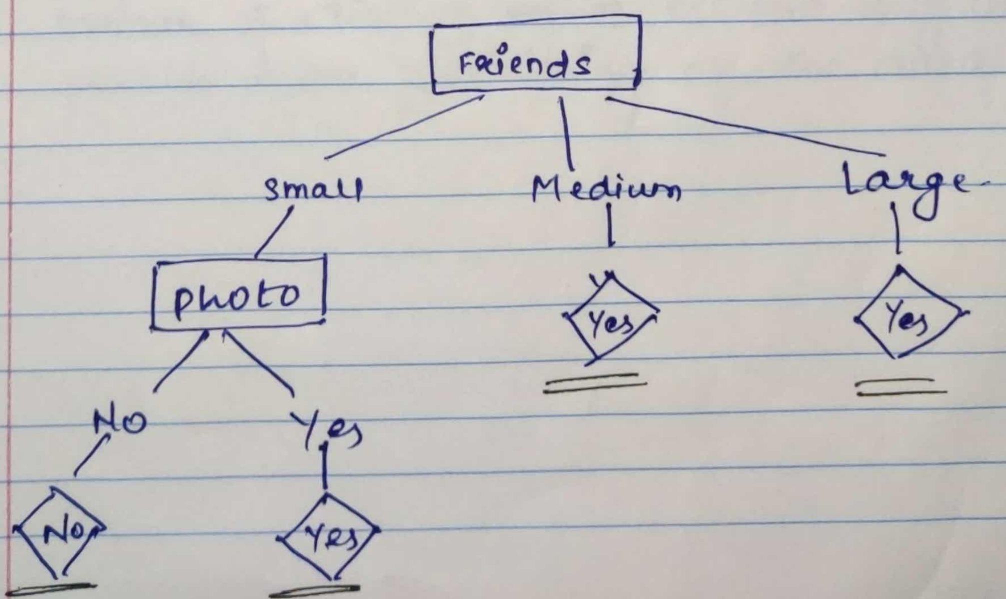
$S_{No} = [Yes = 0, No = 3]$

$\therefore E(S_{No}) = -\frac{0}{3} \cdot \log 0 - \frac{3}{3} \log\left(\frac{3}{3}\right) = 0$

$\therefore IG(photo) = 0.8112 - 0 - 0 = \underline{\underline{0.8112}}.$

$\therefore$ Gain (Post) = 0.3112
Gain (Photo) = 0.8112

The Decision Tree is as follows.



3) Explain the limitation of using Information Gain as the attribute splitting measure.

→ Limitation of using Information Gain as the attribute splitting measure are:

① Biasness :
Information Gain tends to favor attributes with a large number of distinct values. It may lead to Overfitting.

② Continuous attributes may not be handled well :
Information Gain is not well-suited for continuous ~~to~~ attributes without discreatization.

③ Irrelevant attributes :
Information Gain does not account for the relevance of an attribute to the target variable. It Only measures the reduction in entropy, regardless of wheather the attribute is actually useful for predicting the target variable

④ Information Gain tends to favor attributes with a large number of distinct values because they can potentially provide more partitioning of the data.