Name :- Niloy Mallick

ID :- 1731151

CSE 417 - Data Mining and Warehouse

Final Exam

Q2 → Section A

Supervised vs Unsupervised Learning

Supervised Learning (classification)

- Supervision: the training data (observations, measurements, etc) are accompanied by labels indicating the class of the observations
- New data is classified based on the training set

Unsupervised Learning (clustering)

- The class labels of training data is unknown
- Given a set of measurements, observations, etc with the aim of establishing the existence of classes or clusters in the data.

Evaluating classification Methods Issues

Accuracy

- classifer accuracy : predicting class label
- Predictor accuracy : guessing values of Predicted attributes.

speed

- time to construct the model (training time)
- time to use the model (classification / Prediction time)

Robustness

- handling noise and missing values

Scability

- efficiency in disk-resident databases

Interpretability

- understanding and insight provided by the model

other measures, e.g., goodness of rules, such as decision tree size or compactness of classification rules.

Q4 → Section B

The Bayes' theorem describes the Probability of an event based on prior knowledge of the conditions that might be relevant to the event. The Bayes' theorem is expressed in the following formula:

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

Where

* P(A|B) – the Probability of event A occurring, given event B has occurred
* P(B|A) – the Probability of event B occurring, given event A has occurred
* P(A) – the Probability of event A
* P(B) – the Probability of event B

One of the disadvantages of Naive-Bayes is that if you have no occurrences of a class label and a certain attribute value together then the frequency-based probability estimate will be zero. An approach to overcome this 'zero-frequency problem' in a Bayesian environment is to add one to the count for every attribute value-class combination when an attribute value doesn't occure with every class value.

Q5 → Section B

BackPropagation is a neural network learning algorithm. started by psychologist and neurobiologists to develop and test computational analogues of neurons.

A neural network is a set of connected input/output units where each connection has a weight associated with it during the learning phase, the network learns by adjusting the weights so as to be able to predict the correct class label of the input tuples.

The weakness and the strength of Neural network as a classifier are given below:-

Weakness

- long training time
- Require a number of parameters typically best determined empirically e.g. the network topology or structure

- Poore interepretability : difficult to interepret the symbolic meaning behind the learned weights and of 'hidden units" in the network.

Strength

- High tolerance to noisy data.
- Ability to classify untrained Patterns
- Well-suited for continuos-valued input and outputs
- Successful on a wide array of real world data
- Algorithms are inherently parallel
- Techniques have recently been developed for the extraction of roles from trained neural networks.

Q7 → Section C

* age

Age	Buy_Computer	
	Yes	NO
<=30	3 2	2 3
31...40	4	0
>40	3	2

* income

income	Buy_Computer	
	yes	NO
high	2	2
medium	4	2
low	3	1

* Student

Student	Buy_Computer	
	Yes	NO
Yes	6	1
NO	3	4

* Credit_rating

Credit rating	Buy_Computer	
	Yes	NO
fair	6	2
excellent	3	3

• P(Ci) : P(buys_computer = "Yes") = 9/14 = 0.643

P(buys_computer = "NO") = 5/14 = 0.357

Compute $P(x|C_i)$ for each class

$P(\text{age} = \text{"<=30"} \mid \text{buys\_computer} = \text{"yes"}) = 2/9 = 0.222$

$P(\text{age} = \text{"<=30"} \mid \text{buys\_computer} = \text{"NO"}) = 3/5 = 0.6$

$P(\text{income} = \text{"medium"} \mid \text{buys\_computer} = \text{"yes"}) = 4/9 = 0.444$

$P(\text{income} = \text{"medium"} \mid \text{buys\_computer} = \text{"NO"}) = 2/5 = 0.4$

$P(\text{student} = \text{"yes"} \mid \text{buys\_computer} = \text{"yes"}) = 6/9 = 0.667$

$P(\text{student} = \text{"yes"} \mid \text{buys\_computer} = \text{"NO"}) = 1/5 = 0.2$

$P(\text{credit\_rating} = \text{"fair"} \mid \text{buys\_computer} = \text{"yes"}) = 6/9 = 0.667$

$P(\text{credit\_rating} = \text{"fair"} \mid \text{buys\_computer} = \text{"NO"}) = 2/5 = 0.4$

$X = (\text{age} <= 30, \text{income} = \text{medium}, \text{student} = \text{yes}, \text{credit\_rating} = \text{fair})$

$P(X|C_i)$: $P(X \mid \text{buys\_computer} = \text{"yes"}) = 0.222 \times 0.444 \times 0.667 \times 0.667 = 0.044$

$P(X \mid \text{buys\_computer} = \text{"NO"}) = 0.6 \times 0.4 \times 0.2 \times 0.4 = 0.19$

$P(X|C_i) \times P(C_i)$: $P(X \mid \text{buys\_computer} = \text{"yes"}) \times P(\text{buys\_computer} = \text{"yes"})$

$= 0.044 \times 0.222 = 0.028$

$P(X \mid \text{buys\_computer} = \text{"NO"}) \times P(\text{buys\_computer} = \text{"NO"})$

$= 0.357 \times 0.19 = 0.007$

Q8 → Section C

We know,

$$I_j = \sum_i w_{ij} O_j + \theta_j,$$

$$O_j = \frac{1}{1+e^{-I_j}}$$

Net Input and Output Calculation:

Unit, j	Net Input I_j	Output O_j
4	$0.2+0-0.5-0.4=-0.7$	$1/(1+e^{0.7})=0.332$
5	$-0.3+0+0.2+0.2=0.1$	$1/(1+e^{-0.1})=0.525$
6	$(-0.3)(0.332)-(0.2)(0.525)+0.1=0.105$	$1/(1+e^{0.105})=0.4$

Error at Each Node :

$$Err_j = O_j(1 - O_j)(T_j - O_j)$$

calculation of the Error at Each Node

Unit, j	Err$_j$
4	$(0.332)(1-0.332)(0.1311)(-0.3) = -0.0087$
5	$(0.525)(1-0.525)(0.1311)(-0.2) = -0.0065$
6	$(0.474)(1-0.474)(1-0.0474) = 0.1311$

Weight and Bias updating

$$\Delta w_{ij} = (l)\, Err_j O_j$$
$$w_{ij} = w_{ij} + \Delta w_{ij}$$

$$\Delta \theta_j = (l)\, Err_j$$
$$\theta_j = \theta_j + \Delta \theta_j$$

Weight or Bias	New value
w_{46}	$-0.3 + (0.9)(0.1311)(0.332) = -0.261$
w_{56}	$-0.2 + (0.9)(0.1311)(0.525) = -0.138$
w_{14}	$0.2 + (0.9)(-0.0087)(1) = 0.192$
w_{15}	$-0.3 + (0.9)(-0.0065)(1) = -0.306$
w_{24}	$0.4 + (0.9)(-0.0087)(0) = 0.4$
w_{25}	$0.1 + (0.9)(-0.0065)(0) = 0.1$
w_{34}	$-0.5 + (0.9)(-0.0087)(1) = -0.508$
w_{35}	$0.2 + (0.9)(-0.0065)(1) = 0.194$
θ_6	$0.1 + (0.9)(0.1311) = 0.218$
θ_5	$0.2 + (0.9)(-0.0065) = 0.194$
θ_4	$-0.4 + (0.9)(-0.0087) = -0.408$