

Indian Institute of Information Technology, Allahabad
Data Mining C2 Examination Quiz

Name:
Roll No:
Section:

Date: 29.04.2024
Time: 25 mins
15 Marks

1. A training input x is used for a perceptron learning rule. The desired output is t , and the actual output is o . If the learning rate is η , the weight (w) update performed by the perceptron learning rule is described by?

- a. $w_i \leftarrow w_i + \eta(t - o)$
- b. $w_i \leftarrow w_i + \eta(t - o) x$**
- c. $w_i \leftarrow \eta(t - o) x$
- d. $w_i \leftarrow w_i + (t - o) x$

2. Taking transaction ID as a market basket, support for each itemset $\{e\}$, $\{b,d\}$, and $\{b,d,e\}$ is:

Customer ID	Transaction ID	Items Bought
1	1	{a,d,e}
1	2	{a,b,c,e}
2	3	{a,b,d,e}
2	4	{a,c,d,e}
3	5	{b,c,e}
3	6	{b,d,e}
4	7	{c,d}
4	8	{a,b,c}
5	9	{a,d,e}
5	10	{a,b,e}

- a. **0.8, 0.2, 0.2**
 - b. 0.3, 0.3, 0.4
 - c. 0.25, 0.25, 0.5
 - d. 1,0,0
3. A decision tree can be used to build models for
- a. Regression problems
 - b. Classification problems
 - c. Both of the above**
 - d. None of the above
4. Entropy value of ____ represents that the data sample is pure or homogenous and Entropy value of ____ represents that the data sample has a 50-50 split belonging to two categories. **0,1**

For questions 5-6, consider the following hypothetical data regarding the hiring of a person.

GPA	Effort	Confidence	Hire
Low	Some	Yes	No
Low	Lots	Yes	Yes
High	Lots	No	No
High	Some	No	Yes
High	Lots	Yes	Yes

5. Using Naïve Bayes determine whether a person with GPA=High, Effort=Some, and Confidence=Yes be hired:
- a. Yes**
 - b. No
 - c. The example cannot be classified.
 - d. Both classes are equally likely
6. Using Naïve Bayes determine whether a person with Effort=lots, and Confidence=No be hired:
- a. Yes**
 - b. No
 - c. The example cannot be classified.
 - d. Both classes are equally likely
7. Consider a binary classification problem with two classes C1 and C2. Class labels of ten other training set instances sorted in increasing order of their distance to an instance x is as follows: {C1, C2, C1, C2, C2, C2, C1, C2, C1, C2}. How will a K=5 nearest neighbor classifier classify x ?
- a. There will be a tie

Indian Institute of Information Technology, Allahabad
Data Mining C2 Examination Quiz

Name:
Roll No:
Section:

Date: 29.04.2024
Time: 25 mins
15 Marks

-
- b. C1
c. **C2**
d. Not enough information to classify
8. Issues with Euclidean measure are:
a. High dimensional data.
b. Can produce counter-intuitive results.
c. Shrinking density – sparsification effect
d. **All of the above.**
9. The SVM's are less effective when:
a. The data is linearly separable
b. The data is clean and ready to use
c. **The data is noisy and contains overlapping points**
d. None of the above
10. A neuron with 3 inputs has the weight vector $[0.2 \ -0.1 \ 0.1]^T$ and a bias $\theta = 0$. If the input vector is $X = [0.2 \ 0.4 \ 0.2]^T$, then the total input to the neuron is:
a. 0.2
b. **0.02**
c. 0.4
d. 0.10
11. Overfitting is expected when we observe that?
a. With training iterations, error on training set as well as test set decreases
b. **With training iterations, error on training set decreases but test set increases**
c. With training iterations, error on training set as well as test set increases
d. With training iterations, training set as well as test set error remains constant
12. The logic function that cannot be implemented by a perceptron having two inputs is?
a. AND
b. OR
c. NOR
d. **XOR**
13. How would a 3-NN classify the example $A1 = 0, A2 = 0, A3 = 0$ if the distance metric is Euclidean distance?
a. C1
b. C2
c. There will be a tie
d. Not enough information to classify
14. In a multiclass classification problem, Bayes classifier assigns an instance to the class corresponding to:
a. **Maximum aposteriori probability**
b. Maximum apriori probability
c. Lowest aposteriori probability
d. Lowest apriori probability
15. The Decision tree corresponding to the following is?
if C2 then
 if C1 then A3
 else A2
endif
else A1, A3
endif
- Answer**

