Practice Test for Mid-Term

- Due No due date
- Points 8
- Questions 6
- Available Feb 23 at 12am - Mar 3 at 11:59pm
- Time Limit 10 Minutes
- Allowed Attempts 3

Instructions

- This is a practice test for the mid-term exam on **modules 1, 2, 3, 4, 5, and 6.**
- You need to answer **5 MCQs** + **1 Short question** .
- You will have only **10 minutes** to complete your test in **one sitting.**
- **You have up to 3 attempts, and be aware that in the mid-term you have only one attempt**

Take the Quiz Again

Attempt History

	Attempt	Time	Score
LATEST	**Attempt 1**	7 minutes	3 out of 8 *

* Some questions not yet graded

* Some questions not yet graded
Submitted Mar 3 at 9:30pm

The First Part: MCQs

Question 1

1 / 1 pts

Which of the statement is not true about Data Mining?

○ Data mining activities can be subdivided into two major investigation streams, interpretation and prediction.

Correct!

◉

Data mining analysis is to draw a fresh conclusion without investigating the past data, observations and interpretations

○

The term data mining refer to the overall process consisting of data gathering and analysis, development of inductive learning models and adoption of practical decisions and consequent actions based on the knowledge acquired.

○ The data mining process is based on inductive learning methods

⋮⋮

Question 2

0 / 1 pts

Which attributes are numerical attributes that assume a finite number or a countable infinity of values.

Correct Answer

○ Discrete

○ Countable

You Answered

◉ Numerical

○ Ordinal

⋮⋮

Question 3

1 / 1 pts

Among the following which method guarantees that each observation of the dataset appears the same number of times in the training set and exactly once in the test set.

○ Holdout method and Repeated Random Sampling

○ Repeated Random Sampling

Correct!

◉ Cross Validation

○ Holdout method

⋮⋮

Question 4

0 / 1 pts

All of the following steps are part of Naïve Bayes method except:

Correct Answer

○ Assign that class to the old record D.

○ Express the probability as the product of $p(x1|y) \times p(x1|y) \dots p(xn|y)$

○ Determine what classes they all belong to and which is more prevalent

You Answered

◉ Find all the other records where the predictor values are same

⋮⋮

Question 5

1 / 1 pts

When using the k- nearest neighbor classifier, what is the problem of choosing very small value of k?

○ All the three

Correct!

◉ The classifier be capable of overfitting

○ The neighborhood may include points from other classes

○ Misclassification rate will be very high.

⋮

Second Part: Short Questions

⋮

Question 6

Not yet graded / 3 pts

What do we mean by *pruning* the decision tree? Given a decision tree, you have the option of (a) converting the decision tree to rules and then pruning the resulting rules, or (b) pruning the decision tree and then converting the pruned tree to rules. What advantage does (a) have over (b)?
Your Answer:

Once a decision tree has been constructed, it is a simple matter to convert it into an equivalent set of rules.

Converting a decision tree to rules before pruning has three main advantages:

1. "Converting to rules allows distinguishing among the different contexts in which a decision node is used" (Mitchell, 1997, p.72).
   - Each distinct path through the tree produces a distinct rule.
   - Therefore, a single path can be pruned, rather than an entire decision node.
   - If the tree itself were pruned, the only possible actions would be to remove an entire node, or leave it in its original form.
2. Unlike the tree, the rules do not maintain a distinction between attribute tests that occur near the root of the tree and those that occur near the leaves.
   - This allows pruning to occur without having to consider how to re-build the tree if root nodes are removed.
3. Rules are easier for people to read and understand.

To generate rules, trace each path in the decision tree from root node to leaf node. Record the test outcomes as antecedents and the leaf-node classification as the consequent.