

DATA SCIENCE (CDA)
CLASS ASSESSMENT 2 (UNITS 3 AND 4)
MODEL A

1. Which kind of task is this?

“Diagnose the presence of an illness given some symptoms”

- a) **Classification.**
- b) Regression.
- c) Association
- d) Clustering

2. Which kind of task is this?

“What types of failures do we have in a train engine?”

- a) Classification.
- b) Regression.
- c) Correlation
- d) **Clustering**

3. Why is a Naïve Bayes Classifier called Naïve?

- a) Because it gives very poor results.
- b) **Because it assumes that the attributes are independent.**
- c) Because it derives trivially from Bayes’ rule.
- d) Because it ignores the a priori probability.

4. Which of the following is NOT a distance?

- a) Manhattan.
- b) **Malmö.**
- c) Cosine.
- d) Edit.

5. Which kind of task is this?

“Estimate the pressure that a valve had just before a detected failure”

- a) Classification.
- b) **Regression.**
- c) Correlation
- d) Clustering

6. In the deployment and monitoring stages, after months of operation, if a regression model starts giving a much higher error than during evaluation, what can we say and do about this?
- a) The evaluation was incorrect and we need to re-evaluate the model with the new data.
 - b) The model may be obsolete, and retraining a new model with the new data could give good results again.
 - c) The evaluation was incorrect and we need to re-evaluate the model with the old data.
 - d) This is impossible, with a proper monitoring the error of the model should never increase.
7. Which kind of task is this?
- “Given a registry of marriages, what nationalities usually marry each other?”
- a) Classification.
 - b) Correlation.
 - c) Association
 - d) Clustering
8. Which of the following claims is TRUE?
- a) A dendrogram shows how examples are agglomerated according to their linkage distance.
 - b) The agglomerative hierarchical clustering algorithm works in a top-down manner: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.
 - c) The linkage criterion determines the distance between two observations.
 - d) K-means is an agglomerative clustering method.
9. A naive classifier that always predicts the majority class (ignoring minority classes) in a binary problem with 1% of negative instances, obtains good performance in terms of:
- a) Accuracy.
 - b) Macro-accuracy.
 - c) AUC.
 - d) Support and Confidence.
10. What is content-based filtering?
- a) The recommendations are produced by observing the preferences of similar users.
 - b) The recommendations are produced by observing the user's most similar items.
 - c) The recommendations are produced by observing the characteristics of the items (shape, price, colour, etc.).
 - d) The recommendations are produced by removing some forbidden contents.

ASSESSMENT
Answer Sheet

Surname:

Name:

Group in English: ☐

In the following table, circle the correct answer for each question.

Question	Answer			
1	a	b	c	d
2	a	b	c	d
3	a	b	c	d
4	a	b	c	d
5	a	b	c	d
6	a	b	c	d
7	a	b	c	d
8	a	b	c	d
9	a	b	c	d
10	a	b	c	d

The result will be calculated by the statistical correction formula:

$$(\text{Right} - \text{Wrong}/3) \times 1$$

which discounts the probability of getting a right answer by chance on a question with four possibilities.

The mark is between 0 and 10.

Remember that this assessment is just 10% of the final qualification for the course.