BABD

Masters in Business Analytics and Big Data

Exam sample questions

Correct answers are marked with ♦

Question 1

Which one(s) among the following answers is(are) correct?

1. Principal components are orthogonal only if attributes are normalized.
2. In a supervised learning task there is a target variable that can only be numerical.
3. The most only important criterion in the selection of a prediction model is its generalization capability.
4. Let X and Y be two numerical variables where X is unimodal and Y is bimodal, then the mean absolute deviation of X is smaller than the mean absolute deviation of Y.
5. If the covariance of two numerical attributes is equal to zero then the variables are not correlated.
6. ♦ None of the other answers are correct.

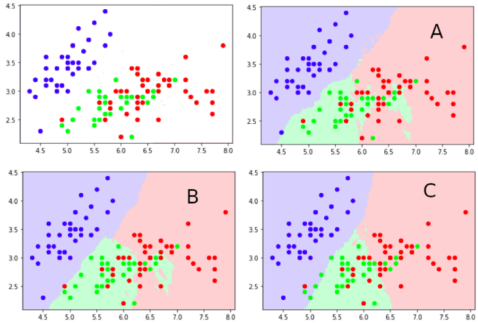
Question 2

Which one(s) among the following answers is(are) correct?

1. The split of a node in a classification tree must be done by maximizing the Gini index of the descendants.
2. The trajectory of a ROC curve is based on the combination of different classifiers.
3. ♦ In the Naive Bayesian method applied to a dataset in which all variables are categorical, the predicted values depend only on the observations of the training set.
4. ♦ In the Logistic regression, the standardization of the variables can have an impact on the interpretation of the variables weights.
5. In a Random Forest, the number of weak learners must be smaller than the number of explanatory features.
6. None of the other answers are correct.

Question 3

The figure depicts a training set of a multi-class classification task and the output of three k-NN models using different parameters.
The training set (upper-left) contains two numerical explanatory features (x and y) and a categorical target containing three categories denoted by C1 (red), C2 (green), and C3 (blue).
The colored regions in the outputs depict the prediction values (C1, C2, and C3) for new observations according to the three different models.



→

Question 3 - cont.

Regarding the figures, which one(s) among the following answers is(are) correct?

1. ♦ The precision of class C3 is greater in model A than in model C.
2. ♦ It is most likely that the third model uses a larger value of the parameter k.
3. In model B, the parameter k is likely to be equal to one.
4. In model A, better results can be obtained by modifying the regularization parameter to give more importance to the regularization term.[1]
5. Among the three models, it is expected that the second model will get better values for quality metrics on the test set.
6. *None of the other answers is correct*

[1] KNN has no regularization term.

Question 4

Referring to the clustering problem, which one(s) among the following answers is(are) correct?

1. ♦ Divisive algorithms are a top-down technique.
2. ♦ Clustering can be used to identify outliers.
3. ♦ A negative value of the silhouette coefficient for the observation x_i indicates that, the mean distance u_i between the points and the observations in its cluster, is greater than the minimum value v_i among the mean distances from observations of the other clusters.
4. Regarding the k-means algorithm, the squared error (EQ function) reaches its minimum value when the number of clusters (parameter k) is equal to one.
5. The standardization of explanatory features can influence the output of partition methods but not hierarchical methods.
6. None of the other answers are correct.