

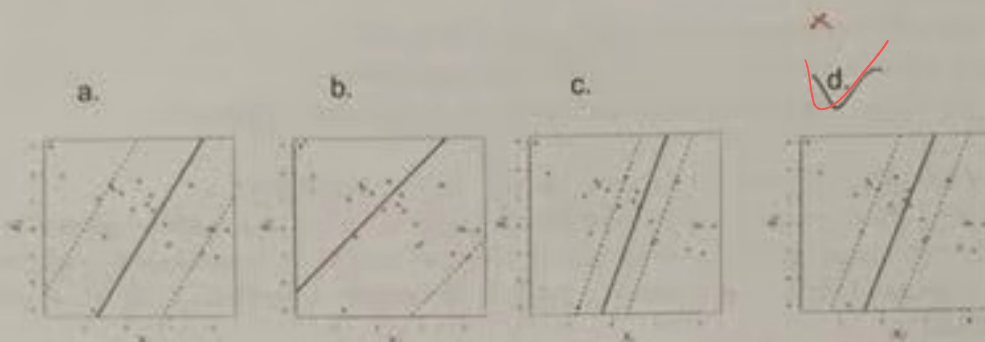
Multiple choice

Note: there is only ONE correct answer for each question.

1. Which of the following on logistic regression and KNN classification is FALSE?
- a. KNN is likely to have better performance than logistic regression when the decision boundary is highly non-linear. ✓
 - b. Neither KNN nor logistic regression require input parameters. ✓
 - c. KNN and logistic regression can both return class probabilities. ✗
 - d. Logistic regression models the logit function to be linear in feature space. ✓
2. Which of the following statement is TRUE for specificity and sensitivity?
- $Spe = \frac{TN}{TN+FP}$ (CN) $Sensitivity = \frac{TP}{TP+FN}$ (CP)
- a. In the cancer prediction problem (cancer as positive and normal as negative), sensitivity is the number of true cancer cases that are captured by the predictive model divided by all cancer cases. ✓
 - b. Again, in cancer prediction, if a classifier predicts all samples as normal, it has 0% specificity and 100% sensitivity. ✗
 - c. When evaluating classification models, specificity and sensitivity should be treated as equally important in all cases because they capture different aspects of model performance. ✗
 - d. Overall classification accuracy is the best metric to summarise specificity and sensitivity. ✗
3. Why would using selected features instead of all features improve classification accuracy in many classification applications?
- a. Feature selection scales the learning features so that the classification model can generalise well for unseen sample classification. ✗
 - b. Feature selection transforms the dataset into higher dimensions that makes samples more separable to classification models. ✗
 - c. Feature selection reduces classification model complexity so that it generalises well for unseen sample classification. 少的变量 ✓
 - d. Feature selection converts the classification problem to be more linearly separable and therefore improves model classification accuracy. ✗
4. Which of the following statement is TRUE?
- a. Forward stepwise feature selection is a type of greedy algorithm and therefore never yields the best model. ✗
 - b. Best feature subset selection guarantees the best model. ✓
 - c. Both forward and backward stepwise feature selection guarantees the best model. ✗
 - d. Forward stepwise feature selection is a type of filter method that ranks the importance of each feature without using a classification model. ✗
5. Which of the following is TRUE about Ridge and Lasso regression?
- a. Lasso can be formulated as $RSS + \lambda \sum_{i=1}^n \beta_i$ ✗
 - b. Ridge can be formulated as $RSS + \lambda \sum_{i=1}^n \beta_i^2$ ✓
 - c. Lasso often leads to sparse models whereas Ridge regression does not. ✓
 - d. Ridge regression often reduces coefficients in the model to zero so it can be used for model selection. ✗
- Lasso

- d 6. Which of the following support vector machine decision boundaries and margins correspond to the smallest C value of:

$$\begin{aligned} & \underset{\beta_0, \beta_1, \dots, \beta_p, \epsilon_1, \dots, \epsilon_n}{\text{maximize}} \quad M \quad \text{subject to} \quad \sum_{j=1}^p \beta_j^2 = 1, \\ & y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i), \\ & \epsilon_i \geq 0, \quad \sum_{i=1}^n \epsilon_i \leq C, \end{aligned}$$



- c 7. Which of the following best depicts the difference between Bagging, Random Forest, and AdaBoost ensembles?
- a. Bagging differs from AdaBoost in that Bagging combines base classifiers sequentially while AdaBoost combines them in parallel. ~~X~~
 - b. Bagging is an extension of Random Forest ensemble method in that it combines a bootstrap sampling procedure with random feature selection. ~~X~~
 - c. AdaBoost assigns different weights to base classifiers based on how much error they produce. In contrast, Bagging and Random Forest assign equal weights to all base classifiers. ☒
 - d. AdaBoost algorithm is most appropriately used to combine weak learners whereas Bagging and Random Forest can be used to combine tree classifiers only. ~~X~~

Word File Edit View Insert Format Tools Table Window Help

Home Insert Design Layout References Mailings Review View

Quiz_1_2018

1. Which of the following ways select the second to the fourth rows of a matrix X?

- ☒ a. $X[c(2:4)]$
- ☒ b. $X(2:4,)$
- ☒ c. $X[c(2, 3, 4),]$
- ☒ d. $X[2:4,]$

2. What is the Euclidean distance between the vectors $a=[5, 2, 5, 3, 8, 2, 3, 4]$ and $b=[1, 4, 5, 8, 2, 4, 5, 7]$?

a. 7.921

☒ b. 8.831

c. 6.482

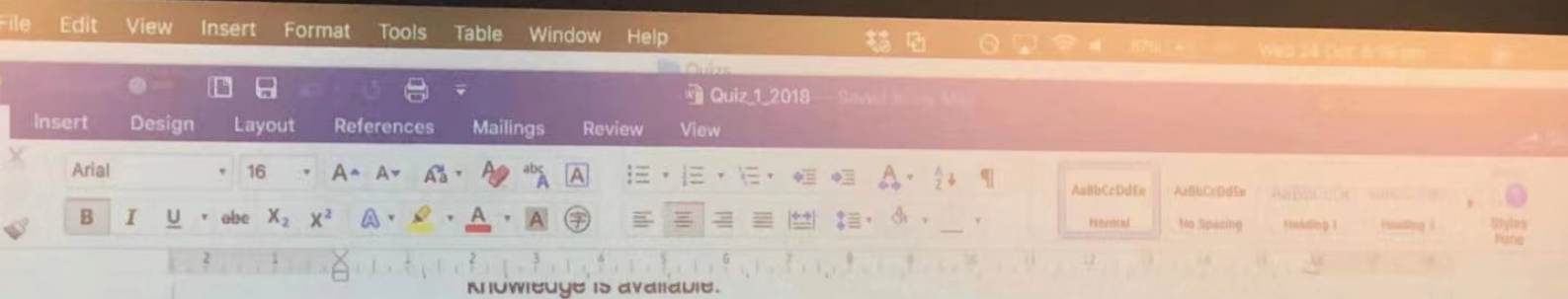
d. 10.132

$$\sqrt{1^2 + 4^2 + 5^2 + 8^2 + 2^2 + 4^2 + 5^2 + 7^2} = 8.831$$

☒ 3. Which following statement(s) about cluster validation is(are) true?

- ☒ a. Cluster validation are for finding appropriate number of clusters in the data
- ☒ b. Cluster validation guarantee for finding optimal number of clusters in the data
- ☒ c. Domain and prior knowledge can be useful for cluster validation.
- ☒ d. Data structure-based clustering validation are more appropriate when prior knowledge is available.

4. What does a smoothing parameter (bandwidth) in a normal (Gaussian) kernel controls for density estimation?



4. What does a smoothing parameter (bandwidth) in a normal (Gaussian) kernel controls for density estimation?
- a. Proportion of nearby points to be considered
 - ☒ b. Number of nearby points to be considered
 - c. Weights of all points to be considered
 - ☒ d. The trade-off between the bias and the variance of the normal kernel
5. Which of the follow are characteristics of a kernel function in smoothing?
- ☒ a. a frequency function from a histogram
 - b. a symmetric function
 - c. a function ranging from -1 to 1
 - d. a function that integral over its support is 1
6. If we were to compute the distance between two real valued vectors and take the magnitude into account, which distance metric is the most appropriate?
- a. Spearman correlation
 - ☒ b. Euclidean distance
 - c. Pearson correlation

Word File Edit View Insert Format Tools Table Window Help

Home Insert Design Layout References Mailings Review View

Quiz_1_2018

d. Manhattan distance

7. Which of the following is true for hierarchical and k-means clustering?

- ☒ a. k-means clustering needs user specified parameter while hierarchical clustering does not
- ☒ b. hierarchical clustering force samples into certain clusters
- ☒ c. k-means clustering force samples into certain clusters
- ☒ d. hierarchical clustering is always faster than k-means clustering

8. Why is clustering called unsupervised learning?

- a. No feature information from the data are used
- b. The algorithm assigns samples randomly
- ☒ c. No class information from the data are used
- d. No sample information from the data are used

9. What is the difference between classification and regression?

- a. Classification uses response variable whereas regression does not
- ☒ b. Classification algorithms always give discrete output whereas regression algorithms always gives continues output
- c. Classification uses discrete class label whereas regression uses continues

- a. No feature information from the data are used
- b. The algorithm assigns samples randomly
- c. No class information from the data are used
- d. No sample information from the data are used

9. What is the difference between classification and regression?

- ☒ a. Classification uses response variable whereas regression does not
- ☒ b. Classification algorithms always give discrete output whereas regression algorithms always gives continues output
- ☒ c. Classification uses discrete class label whereas regression uses continues response variable
- ☒ d. Regression gives discrete prediction whereas classification gives continues prediction

10. Which of the following is true about k-nearest neighbour (kNN) classifier?

- ☒ a. kNN performs better than Logistic regression when the decision boundary is highly nonlinear
- ☒ b. kNN performs better than Logistic regression when data samples follow roughly normal distribution
- ☒ c. kNN requires a distance measure for calculating similarity among samples.
- ☒ d. kNN is a parametric classifier because user need to pre-defined k