



LIMITED SAMPLE OF MULTIPLE-CHOICE QUESTIONS

April 8 - April 22, 2024



Brought to you by:



This document contains a small sample of multiple-choice questions similar to those on the FDP exam. It serves as a useful tool for individuals who are considering taking the FDP exam and want to gain a better understanding of the types of questions they may encounter. Please note that this document is not a sample exam as it does not include the multi-part constructed response questions, which constitute a portion of the exam.

If you pursue the FDP Charter by registering for the exam, you will gain access to a comprehensive set of Practice Questions and a Sample Exam. For details about the exam structure, please refer to the FDP Study Guide, which is freely available on the FDP website.

- Hossein Kazemi

SAMPLE QUESTIONS FOR REGISTERED PROFILES

1. Frank plays pool on Simonis tablecloth 75% of the time. His overall potting success rate is 80%, which increases to 90% when he plays on Simonis tablecloth. What is the probability that Frank is playing on Simonis tablecloth if he has successfully potted a ball?

Options:

- A. 0.82
- B. 0.84
- C. 0.92
- D. 0.94

Answer: B

2. In the k-nearest neighbors approach to classification, what is the impact of increasing k on the decision boundary?

Options:

- A. It becomes wider.
- B. It becomes close to linear.
- C. It becomes highly nonlinear.

Answer: B

3. A Bayesian classifier is used to classify a binary variable Y such that if $P(Y = 1|X) > 0.5$, then $Y = 1$, else $Y = 0$. The following estimates for $P(Y = 1|X)$ are obtained from the training set:

| $P(Y=1 X)$ |
|------------|
| 0.15 |
| 0.33 |
| 0.54 |
| 0.86 |

What is the Bayes error rate?

Options:

- A. 0.27
- B. 0.45
- C. 1.88

Answer: A

4. A binary target variable, Y , can only assume values of 0 or 1. Its log-odds linear function, $\ln\left(\frac{p(Y=1)}{1-p(Y=1)}\right)$, is based on features $F1$ and $F2$, and is shown below.

$$f(F1, F2) = 0.8 + 0.7 * F1 - 0.2 * F2$$

What is the probability that the target variable Y belongs to class 0, given new values of $F1 = 1, F2 = 2$?

Options:

- A. 0.25
- B. 0.33
- C. 0.75
- D. 0.80

Answer: A

5. The Euclidean distance of an observation from a landmark is 2 and the scaling parameter for the observation is $1/2$. If the support vector machine approach to classification is being used, what is the value of the radial basis function (RBF) for the observation?

Options:

- A. 0.135
- B. 1.000
- C. 1.135

Answer: A

6. A linear regression of the form shown below is estimated using a small dataset shown in the following table. \bar{x} denotes the average of x and \bar{y} denotes the average of y .

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

| Observation | x_i | y_i | $x_i - \bar{x}$ | $y_i - \bar{y}$ | $(x_i - \bar{x})(y_i - \bar{y})$ | $(x_i - \bar{x})^2$ | $(y_i - \bar{y})^2$ | ε_i^2 |
|-------------|-------|-------|-----------------|-----------------|----------------------------------|---------------------|---------------------|-------------------|
| 1 | 3 | 4 | 5 | -9.6 | -48 | 25 | 92.16 | 3.72 |
| 2 | 2 | 6 | 4 | -7.6 | -30.4 | 16 | 57.76 | 2.63 |
| 3 | 0 | 8 | 2 | -5.6 | -11.2 | 4 | 31.36 | 0.98 |
| 4 | -2 | 10 | 0 | -3.6 | 0 | 0 | 12.96 | 12.96 |
| 5 | -2 | 12 | 0 | -1.6 | 0 | 0 | 2.56 | 2.56 |
| 6 | -3 | 13 | -1 | -0.6 | 0.6 | 1 | 0.36 | 8.44 |
| 7 | -4 | 16 | -2 | 2.4 | -4.8 | 4 | 5.76 | 4.89 |
| 8 | -4 | 20 | -2 | 6.4 | -12.8 | 4 | 40.96 | 3.20 |
| 9 | -5 | 22 | -3 | 8.4 | -25.2 | 9 | 70.56 | 2.20 |
| 10 | -5 | 25 | -3 | 11.4 | -34.2 | 9 | 129.96 | 20.10 |
| Sum | -20 | 136 | | | -166 | 72 | 444.40 | 61.68 |

What is the standard error of β ?

Options:

- A. 0.09
- B. 0.11
- C. 0.29
- D. 0.33

Answer: D

7. The following regression equation is estimated using a training data set:

$$Y_i = 1 + 4.5X_{1i} + 5.7X_{2i} - 8.2X_{3i} + e_i$$

If the means of X_1 , X_2 , and X_3 in the training dataset are 10, 5, and 7, respectively, what is the unconditional mean of Y ?

Options:

- A. 17.1
- B. 22.7
- C. 75.6

Answer: A

8. When interaction effects are included, the main effects also must be included even if the p-values associated with the coefficients of the main effects are not significant. What is the name of this principle?

Options:

- A. **Baseline principle**
- B. **Interactive principle**
- C. **Hierarchical principle**

Answer: C

9. Using a training dataset, two child nodes (c_1 and c_2) are created from a parent node. The entropy of the parent node is 0.7, and the entropies of c_1 and c_2 are 0.4 and 0.1, respectively. If the information gain from this split is 0.54, what is the probability of a training instance ($p(c_1)$) being in child node c_1 ?

Options:

- A. 0.20
- B. 0.30
- C. 0.40

Answer: A

10. While building a decision tree, a leaf with 15 positive and zero negative instances is reached. Using the Laplace correction, what is the probability of a new instance being negative?

Options:

- A. 0.00
- B. 0.01
- C. 0.06
- D. 0.10

Answer: C

11. A data set contains financial information on a group of firms. 60% of these firms have a debt-to-total asset ratio greater than or equal to 0.5, and the same ratio is less than 0.5 for the remaining firms. What is the Gini measure of this data set?

Options:

- A. 0.20
- B. 0.48
- C. 0.52
- D. 0.60

Answer: B

12. The table below provides feature information for eight observations and the target variable for seven of the eight observations. The last column provides the Euclidean distance from the first observation.

| Observation | Feature 1 | Feature 2 | Target | Distance from Observation 1 |
|-------------|-----------|-----------|--------|-----------------------------|
| 1 | 37 | 50 | ? | 0 |
| 2 | 22 | 50 | Yes | 15.0 |
| 3 | 63 | 200 | No | 152.2 |
| 4 | 59 | 170 | Yes | 122.0 |
| 5 | 25 | 40 | No | 15.6 |
| 6 | 30 | 55 | No | 8.6 |
| 7 | 42 | 80 | Yes | 30.4 |
| 8 | 35 | 60 | No | 10.2 |

Using the four nearest neighbors, what is the probability of the first observation having a positive (Yes) response?

Options:

- A. 25%
- B. 50%
- C. 75%
- D. 100%

Answer: A

13. Which of the following best explains the Naïve Bayes classifier for a new example?

Options:

- A. It estimates the probability that the example belongs to a class and reports the class with the lowest entropy.
- B. It estimates the probability that the example belongs to the class with the most members and reports that probability.
- C. It estimates the probability that the example belongs to each class and reports the class with the highest probability.

Answer: C

14. The following 8 observations have a mean of 60.8 and a variance of 691.2. What would be the scaled value of 21.6 after applying Z-score normalization?

| | | | | | | | |
|------|------|------|------|------|------|------|------|
| 21.6 | 84.1 | 74.1 | 39.9 | 90.3 | 68.5 | 30.3 | 77.7 |
|------|------|------|------|------|------|------|------|

Options:

- A. -1.491
- B. -0.057
- C. 0.057
- D. 1.491

Answer: A

15. Given two inputs $x_1 = 2$ and $x_2 = 3$ with two weights $w_1 = 0.3$ and $w_2 = 0.5$ and a bias of $a = 1$, what is the value of a neuron that is determined by a sigmoid function?

Options:

- A. 0.431
- B. 0.957
- C. 1.050
- D. 3.100

Answer: B

16. A gradient descent algorithm with a learning rate of 0.1 is used to minimize the following function.

$$f(x) = 2x^2 - 3x + 7$$

What would be the new value of x in the next iteration?

| Iteration | x | Gradient |
|-----------|-----|----------|
| 0 | 2 | 5.00 |

Options:

- A. -3.0
- B. 1.5
- C. 2.5
- D. 5.2

Answer: B

17. Which is a characteristic of a convolutional neural network?

Options:

- A. It connects the values in one layer to all the neurons in the previous layer.
- B. It connects the values in one layer to the latent variables of the previous layer.
- C. It connects the values in one layer to only a subset of the neurons in the previous layer.

Answer: C

18. A machine learning algorithm is used to predict customers who will purchase a product. It costs \$10 to target a customer. When a customer purchases the product, the net profit from that sale is \$40. A profit curve is to be constructed for the algorithm. The following confusion matrix is being considered for creating a particular point on the profit curve.

| | Actual Positive | Actual Negative | Sum |
|--------------------|-----------------|-----------------|-----|
| Predicted Positive | 90 | 20 | 110 |
| Predicted Negative | 40 | 150 | 190 |
| Sum | 130 | 170 | 300 |

What are the coordinates of the above confusion matrix in the profit curve?

Options:

- A. 0.30, \$3,200
- B. 0.37, \$3,400
- C. 0.43, \$3,600
- D. 0.49, \$3,800

Answer: B

19. Suppose classifier A plots in the top-left corner of a ROC space and classifier B plots in the bottom-right corner of a ROC space. Which is true about the characteristics of the two classifiers?

Options:

- A. Classifier A has a high true positive rate, and classifier B has a high true negative rate
- B. Classifier A has a high true negative rate, and classifier B has a high true positive rate
- C. Classifier A has a high true positive rate, and classifier B has a high false positive rate

Answer: C

20. What is one of the limitations of the Shapley values?

Options:

- A. They do not work well for ensemble models comprising several underlying models.
- B. They can affect the prediction in different ways even if two features are symmetric.
- C. Interactions between features can cause them to consider unrealistic combinations of feature values.

Answer: C

21. When is the “N-grams” approach more useful than the “Bag-of-Words” approach in representing a text document?

Options:

- A. When a smaller number of features are required.
- B. When it is necessary to model a set of topics in a corpus separately.
- C. When particular phrases are significant, but the words they contain are not.

Answer: C

22. In applying natural language processing to a text document, what process maps a word (such as “worse”), to its root (in this case, “bad”)?

Options:

- A. Stemming
- B. Tokenization
- C. Lemmatization
- D. Syntactical parsing

Answer: C

23. What is the key assumption when applying the Naïve Bayes Classifier to text document classification?

Options:

- A. All words have the same prior probabilities.
- B. Words must be unambiguous and singularly connoted.

- C. Words deemed positive must be considered separately from words deemed negative.
- D. The occurrence of a particular word is completely unrelated to the occurrence of any other words.

Answer: D

- 24.** According to “The 10 Reasons Most Machine Learning Funds Fail” by López de Prado, what is the best way of organizing a machine learning fund that works on several different strategies?

Options:

- A. Each strategy must be managed by a separate group, organized as silos.
- B. All strategies must be managed by the entire group, organized like an assembly line.
- C. A single team must manage the common tasks, and unique aspects of each strategy must be managed by separate groups.

Answer: B

- 25.** Twelve (12) documents are being analyzed. The word “intelligence” represents 5% of all the words that appear in document D1. The word “intelligence” appears in 6 of the 12 documents. What is the term frequency inverse document frequency (TFIDF) of the document D1 regarding the word “intelligence”?

Options:

- A. 0.0153
- B. 0.0500
- C. 0.0847
- D. 0.3000

Answer: C