

Databricks Certified Professional Data Scientist

1. Feature Hashing approach is "SGD-based classifiers avoid the need to predetermine vector size by simply picking a reasonable size and shoehorning the training data into vectors of that size" now with large vectors or with multiple locations per feature in Feature hashing?

A. Is a problem with accuracy

B. It is hard to understand what classifier is doing

C. It is easy to understand what classifier is doing

D. Is a problem with accuracy as well as hard to understand what classifier us doing

Answer(s): B

2. What are the advantages of the Hashing Features?

A. Requires the less memory

B. Less pass through the training data

C. Easily reverse engineer vectors to determine which original feature mapped to a vector location

Answer(s): A B

3. Question-3: In machine learning, feature hashing, also known as the hashing trick (by analogy to the kernel trick), is a fast and space-efficient way of vectorizing features (such as the words in a language), i.e., turning arbitrary features into indices in a vector or matrix. It works by applying a hash function to the features and using their hash values modulo the number of features as indices directly, rather than looking the indices up in an associative array. So what is the primary reason of the hashing trick for building classifiers?

A. It creates the smaller models

B. It requires the lesser memory to store the coefficients for the model

C. It reduces the non-significant features e.g. punctuations

D. Noisy features are removed

Answer(s): B

4. Suppose A, B, and C are events. The probability of A given B, relative to P(C), is the same as the probability of A given B and C (relative to P). That is,

A. $P(A,B|C) P(B|C) = P(A|B,C)$

B. $P(A,B|C) P(B|C) = P(B|A,C)$

C. $P(A,B|C) P(B|C) = P(C|B,C)$

D. $P(A,B|C) P(B|C) = P(A|C,B)$

Answer(s): A

5. What is the considerable difference between L1 and L2 regularization?

A. L1 regularization has more accuracy of the resulting model

B. Size of the model can be much smaller in L1 regularization than that produced by L2-regularization

C. L2-regularization can be of vital importance when the application is deployed in resource-tight environments such as cell-phones.

D. All of the above are correct

Answer(s): B

6. Regularization is a very important technique in machine learning to prevent overfitting. Mathematically speaking, it adds a regularization term in order to prevent the coefficients to fit so perfectly to overfit. The difference between the L1 and L2 is...

A. L2 is the sum of the square of the weights, while L1 is just the sum of the weights

B. L1 is the sum of the square of the weights, while L2 is just the sum of the weights

C. L1 gives Non-sparse output while L2 gives sparse outputs

D. None of the above

Answer(s): A

7. Select the correct option which applies to L2 regularization

A. Computational efficient due to having analytical solutions

B. Non-sparse outputs

C. No feature selection

Answer(s): A B C

8. Regularization is a very important technique in machine learning to prevent over fitting. And Optimizing with a L1 regularization term is harder than with an L2 regularization term because

A. The penalty term is not differentiate

B. The second derivative is not constant

C. The objective function is not convex

D. The constraints are quadratic

Answer(s): A

9. Logistic regression is a model used for prediction of the probability of occurrence of an event. It makes use of several variables that may be.....

A. Numerical

B. Categorical

C. Both 1 and 2 are correct

D. None of the 1 and 2 are correct

Answer(s): C

10. Spam filtering of the emails is an example of

A. Supervised learning

B. Unsupervised learning

C. Clustering

D. 1 and 3 are correct

E. 2 and 3 are correct

Answer(s): A

11. Select the choice where Regression algorithms are not best fit

A. When the dimension of the object given

B. Weight of the person is given

C. Temperature in the atmosphere

D. Employee status

Answer(s): D

12. Question-13.

Which of the following is not the Classification algorithm?

A. Logistic Regression

B. Support Vector Machine

C. Neural Network

D. Hidden Markov Models

E. None of the above

Answer(s): E

13. Suppose a man told you he had a nice conversation with someone on the train. Not knowing anything about this conversation, the probability that he was speaking to a woman is 50% (assuming the train had an equal number of men and women and the speaker was as likely to strike up a conversation with a man as with a woman). Now suppose he also told you that his conversational partner had long hair. It is now more likely he was speaking to a woman, since women are more likely to have long hair than men. _____ can be used to calculate the probability that the person was a woman.

A. SVM

B. MLE

C. Bayes' theorem

D. Logistic Regression

Answer(s): C

14. Which of the following could be features?

A. Words in the document

B. Symptoms of a diseases

C. Characteristics of an unidentified object

D. Only 1 and 2

E. All 1,2 and 3 are possible

Answer(s): E

15. Refer to image below

Given a class variable Y and a dependent feature vector X through x_n , Bayes' theorem states the following relationship:

$$P(y | x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n | y)}{P(x_1, \dots, x_n)}$$

Using the naive independence assumption that

$$P(x_i | y, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i | y),$$

for all i , this relationship is simplified to

$$P(y | x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^n P(x_i | y)}{P(x_1, \dots, x_n)}$$

Since $P(x_1, \dots, x_n)$ is constant given the input, we can use the following classification rule:

$$P(y | x_1, \dots, x_n) \propto P(y) \prod_{i=1}^n P(x_i | y)$$
$$\Downarrow$$
$$\hat{y} = \arg \max_y P(y) \prod_{i=1}^n P(x_i | y).$$

The different naive Bayes classifiers differ mainly by the assumptions they make regarding the distribution of

- A. $P(x_i | y)$
- B. $\arg \max_y P(y)$
- C. $\prod_{i=1}^n P(x_i | y)$
- D. $P(y)$

A. Option A

B. Option B

C. Option C

D. Option D

Answer(s): A

16. A fruit may be considered to be an apple if it is red, round, and about 3" in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of the

A. Presence of the other features.

B. Absence of the other features.

C. Presence or absence of the other features

D. None of the above

Answer(s): C

17. Select the correct statement regarding the naive Bayes classification

A. it only requires a small amount of training data to estimate the parameters

B. Independent variables can be assumed

C. only the variances of the variables for each class need to be determined

D. for each class entire covariance matrix need to be determined

Answer(s): A B C

18. In which of the following scenario we can use naive Bayes theorem for classification

A. Classify whether a given person is a male or a female based on the measured features. The features include height, weight and foot size.

B. To classify whether an email is spam or not spam

C. To identify whether a fruit is an orange or not based on features like diameter, color and shape

Answer(s): A B C

19. Which of the following are advantages of the Support Vector machines?

A. Effective in high dimensional spaces.

B. it is memory efficient

C. possible to specify custom kernels

D. Effective in cases where number of dimensions is greater than the number of samples

E. Number of features is much greater than the number of samples, the method still give good performances

F. SVMs directly provide probability estimates

Answer(s): A B C D

20. Support vector machines (SVMs) are a set of supervised learning methods used for

A. Linear classification

B. Non-linear classification

C. Regression

Answer(s): A B C

