Question 1   0 / 1 point

Consider a dataset with p features and n observations. Which of the following statements is true for Principal Component Analysis?

a ⃝ There are at most $max(n+1, p)$ principal components.

b ⃝ There are at least $min(n-1, p)$ principal components.

c ⃝ There are at most $max(n-1, p)$ principal components.

d ⃝ There are at most $min(n-1, p)$ principal components.

Question 2   0 / 1 point

Which of the following hyperparameters controls the size of the region of the input image that the convolutional filter processes at each step?

a ⃝ Batch size

b ⃝ Number of epochs

c ⃝ Kernel size

d ⃝ Learning rate

Question 3   0 / 1 point

Which of the following is a major limitation of the K-Means clustering algorithm?

a ⃝ K-Means cannot handle non-numeric data.

b ⃝ K-Means can only be used for clustering spherical clusters.

c ⃝ K-Means requires the user to specify the number of clusters after the model is trained.

d ⃝ K-Means is sensitive to the initial placement of centroids.

Question 4 0 / 1 point

What is the primary goal of Principal Component Analysis (PCA)?

- a ○ To increase the number of dimensions in the dataset to improve performance.

- b ○ To reduce the dimensionality of the data while preserving as much variance as possible.

- c ○ To remove outliers from the dataset.

- d ○ To classify the data into different categories based on features.
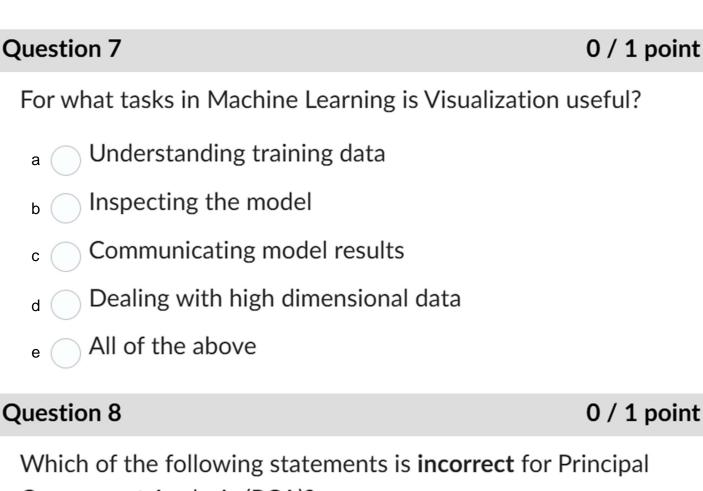
Question 5 0 / 1 point

Which of the following types of problems is Naive Bayes best suited for?

- a ○ Problems where classes are not linearly separable.

- b ○ Problems where features are strongly correlated.

- c ○ Text classification tasks such as spam email detection.

- d ○ Tasks involving a large number of continuous features with no categorical variables.

Which of the following statements best describes the primary difference between **generative** and **discriminative** algorithms?

a ◯ Discriminative models can create new data instances, while generative models focus on modeling different kinds of instances.

b ◯ Discriminative models require fewer features than generative models to achieve the same level of accuracy.

c ◯ Discriminative models learn the probability of the classes given the data, while generative models learn the joint probability of the data and the class.

d ◯ Generative models typically perform better on large datasets, while discriminative models are best for smaller datasets.

Question 7   0 / 1 point

For what tasks in Machine Learning is Visualization useful?

- a ◯ Understanding training data
- b ◯ Inspecting the model
- c ◯ Communicating model results
- d ◯ Dealing with high dimensional data
- e ◯ All of the above

Question 8   0 / 1 point

Which of the following statements is **incorrect** for Principal Component Analysis (PCA)?

- a ◯ PCA is used for feature selection.
- b ◯ Principal Component directions are computed using the order sequence of eigenvectors.
- c ◯ For multi-dimensional data, there are multiple principal components.
- d ◯ PCA considers linear combination of all features.

Question 9   0 / 1 point

Which of the following is **NOT** a core component of a Markov Decision Process (MDP)?

a ◯ States (S)

b ◯ Loss Function (L)

c ◯ Rewards (R)

d ◯ Actions (A)

Question 10   0 / 1 point

What is the key feature of a Recurrent Neural Network (RNN) that distinguishes it from traditional feedforward neural networks?

a ◯ RNNs process data sequentially, maintaining a memory of previous inputs through feedback connections.

b ◯ RNNs have no internal state and do not capture temporal dependencies.

c ◯ RNNs use a fixed-size architecture with no connections between layers.

d ◯ RNNs are only used for image classification tasks.

Question 11   0 / 1 point
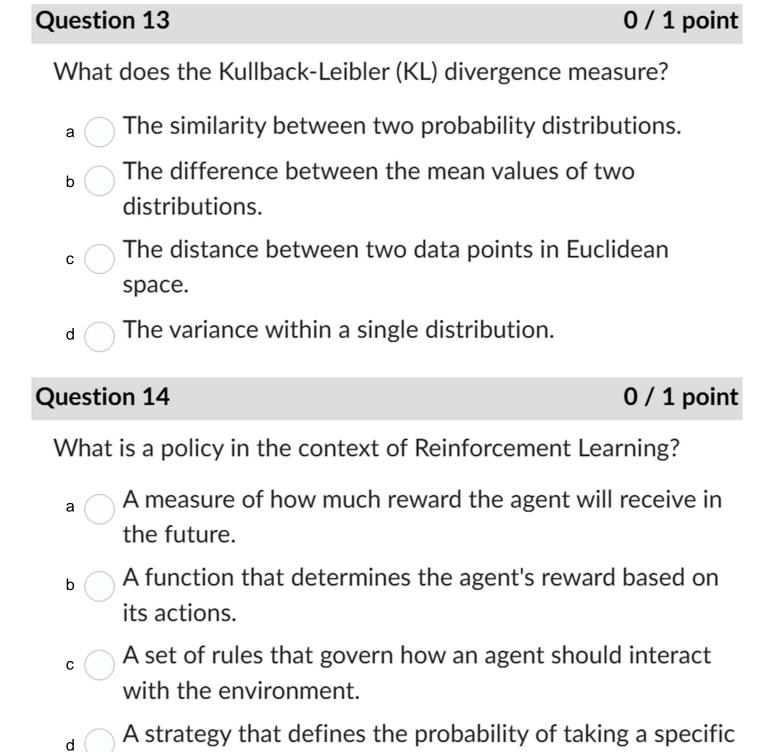
Which of the following is a key assumption made by the Naive Bayes algorithm?

- a ○ Features are conditionally independent given the class label.

- c ○ Features are highly correlated with each other.

- d ○ The class distribution follows a Gaussian distribution.

- e ○ The dataset contains only binary features.

Question 12   0 / 1 point

Which of the following is a variant of Recurrent Neural Networks (RNNs) designed to address the vanishing gradient problem?

- a ○ Long Short-Term Memory (LSTM)

- b ○ Convolutional Neural Network (CNN)

- c ○ Generative Adversarial Network (GAN)

- d ○ Autoencoder

Question 13   0 / 1 point

What does the Kullback-Leibler (KL) divergence measure?

- a ◯ The similarity between two probability distributions.

- b ◯ The difference between the mean values of two distributions.

- c ◯ The distance between two data points in Euclidean space.

- d ◯ The variance within a single distribution.

Question 14   0 / 1 point

What is a policy in the context of Reinforcement Learning?

- a ◯ A measure of how much reward the agent will receive in the future.

- b ◯ A function that determines the agent's reward based on its actions.

- c ◯ A set of rules that govern how an agent should interact with the environment.

- d ◯ A strategy that defines the probability of taking a specific action given a state.

What does the Bellman Equation represent in the context of Reinforcement Learning?

- a ○ A recursive relationship that helps calculate the optimal state-value functions for the agent.

- b ○ A method to calculate the transition probabilities between states.

- c ○ A function used to update the agent's position in the environment.

- d ○ A formula that calculates the reward for each action taken by an agent.

Question 16   **0 / 1 point**

Consider a 5x5 input image with the following values:

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

Consider you are applying a 3x3 filter (kernel) with the following values:

1	0	-1
1	0	-1
1	0	-1

Perform a **valid** convolution operation, which means the filter is applied only where it fits entirely inside the input image. What is the resulting value at the **top-left** corner of the output feature map?

- a ○ -12
- b ○ 6
- c ○ -6
- d ○ 12

Consider the below equation for reinforcement learning, where R is the reward at state t (T is the terminal state) and there are k states to sweep over. With G being the overall reward during learning, what is gamma (γ)?

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \ ... + \gamma^{k-1} R_{T+k+1}$$

a   ○   Number of Actions

b   ○   Discount Rate

c   ○   Learning Rate

d   ○   Iteration Number

Question 18   0 / 1 point

Which of the following is a key characteristic of hierarchical clustering?

a ○ It uses a predefined number of clusters to begin with and iterates to optimize them.

b ○ It requires the user to specify the number of clusters before starting the algorithm.

c ○ It starts with all data points as individual clusters and merges them iteratively.

d ○ It starts by treating all data points as one single cluster and splits them iteratively.

Question 19   0 / 1 point

Which of the following layers is typically used in Convolutional Neural Networks to reduce the spatial dimensions of feature maps, while keeping the most important information?

a ○ Convolutional Layer

b ○ Fully Connected Layer

c ○ ReLU Layer

d ○ Pooling Layer

Question 20   0 / 1 point

In Reinforcement Learning, what does an agent aim to maximize?

a ○ The cumulative reward over time.

b ○ The total number of actions taken.

c ○ The total number of states available.

d ○ The number of states visited.

Answers

1) d

2) c

3) d

4) b

5) c

6) c

7) e

8) a

9) b

10) a

11) a

12) a

13) a

14) d

15) a

16) c

17) b

18) c

19) d

20) a