Question Description	Answer Choice 1	Answer Choice 2	Answer Choice 3	Answer Choice 4	Correct Answer Choice
What is an advantage of density-based clustering algorithms?	a. They require a predefined number of clusters	b. They are not suitable for high-dimensional data	c. They can handle clusters of arbitrary shapes and sizes	d. They are computationally complex.	3
Which of the following is an application of unsupervised learning?	a. Sentiment analysis	b. Image classification	c. Anomaly detection	d. Handwriting recognition.	3
What is a limitation of K-means clustering?	a. It is sensitive to outliers	b. It requires a large amount of computational resources	c. It cannot handle non-linearly separable data	d. It always converges to the global optimum.	1
What is the convergence criterion in K-means clustering?	a. When the within-cluster sum of squares is minimized	b. When the between-cluster sum of squares is maximized	c. When the centroids stop changing significantly between iterations	d. When the number of clusters stabilizes.	3
Which technique is commonly used to initialize cluster centroids in K-means clustering?	a. Random initialization	b. K-nearest neighbors	c. Decision trees	d. Support Vector Machines (SVM).	1
What happens if K-means clustering fails to converge?	a. The algorithm restarts with different initial centroids	b. The algorithm assigns data points randomly to clusters	c. The algorithm terminates and returns an error	d. The algorithm stops and returns the current clusters.	3
What distinguishes clustering from classification?	a. Clustering involves grouping data points based on similarity while classification assigns predefined labels to data points.	b. Clustering requires labeled data while classification does not.	c. Clustering is a supervised learning technique while classification is unsupervised.	d. Clustering can only handle numerical data while classification can handle categorical data.	1
Which of the following is NOT a common application of clustering?	a. Customer segmentation	b. Image classification	c. Anomaly detection	d. Document clustering.	2
What is the main advantage of unsupervised learning techniques like clustering?	a. They require less computational resources compared to supervised learning	b. They do not require labeled data for training	c. They always produce accurate predictions	d. They are more interpretable than supervised learning techniques.	2
Which parameter of the KMeans class specifies the number of clusters to form?	a. n_clusters	b. n_init	c. max_iter	d. random_state.	1
What is the purpose of the fit() method in scikit-learn's KMeans class?	a. It initializes the cluster centroids	b. It assigns each data point to the nearest cluster	c. It updates the cluster centroids based on the data	d. It evaluates the quality of clustering.	2
Which attribute of the KMeans object in scikit-learn stores the cluster centroids?	a. cluster_centers_	b. labels_	c. inertia_	d. n_clusters.	1
Which method can be used to determine the optimal number of clusters when using KMeans in scikit-learn?	a. Elbow method	b. Silhouette method	c. Gap statistic method	d. All of the above.	4
What range of values does the Silhouette coefficient typically lie between?	a. [-1, 0]	b. [0, 1]	c. [0, ∞)	d. [-∞ ∞).	1
In the context of clustering evaluation what does a Silhouette coefficient close to 1 indicate?	a. Good clustering	b. Poor clustering	c. Random clustering	d. No clustering.	1
What is the Davies-Bouldin index used for in clustering evaluation?	a. Measuring the compactness and separation of clusters	b. Calculating the distance between clusters	c. Assessing the similarity between two clusterings	d. Evaluating the density and separation of clusters.	1
What does the term "principal components" refer to in PCA?	a. The features with the highest variance	b. The original features of the dataset	c. The eigenvectors of the covariance matrix	d. The target variables of the dataset.	3
In PCA how are principal components ordered?	a. In ascending order of their eigenvalues	b. In descending order of their eigenvalues	c. In random order	d. In the order they appear in the dataset.	2
What is the significance of the eigenvalues in PCA?	a. They represent the amount of variance explained by each principal component	b. They determine the number of features in the dataset	c. They indicate the number of clusters formed by the data	d. They represent the number of observations in the dataset.	1

Which technique aims to find a lower-dimensional representation of the data that preserves the pairwise distances between data points?	a. Principal Component Analysis (PCA)	b. Linear Discriminant Analysis (LDA)	c. t-Distributed Stochastic Neighbor Embedding (t-SNE)	d. Autoencoders.	3
What is the primary advantage of using PCA for dimensionality reduction?	a. It is computationally efficient	b. It preserves the class separability of the data	c. It is robust to outliers	d. It captures the maximum variance in the data.	4
Which dimensionality reduction technique is particularly useful for visualizing high-dimensional data in low-dimensional space?	a. PCA	b. LDA	c. t-SNE	d. Random Forest.	3
What is the purpose of the fit() method in scikit-learn's PCA class?	a. It transforms the data into principal components	b. It computes the mean and standard deviation of the data	c. It fits the PCA model to the data	d. It computes the explained variance ratio.	3
Which attribute of the PCA object in scikit-learn stores the principal axes in feature space?	a. components_	b. explained_variance_	c. explained_variance_ratio_	d. mean_.	1
What does the explained_variance_ratio_ attribute represent in scikit-learn's PCA object?	a. The variance of each principal component	b. The total variance explained by all the principal components	c. The eigenvalues of the covariance matrix	d. The mean of the data.	2
What does each point in a scatter plot represent when visualizing reduced dimensions using PCA?	a. A feature of the dataset	b. A dimension of the dataset	c. A data point in the high-dimensional space	d. An outlier in the dataset.	3
How can you interpret the distance between points in a scatter plot visualization of reduced dimensions?	a. It represents the variance explained by each component	b. It indicates the correlation between features	c. It shows the similarity between data points in the reduced space	d. It signifies the number of dimensions in the dataset.	3
What is the primary advantage of using scatter plots for dimensionality reduction visualization?	a. They can handle high-dimensional data efficiently	b. They allow easy identification of clusters and patterns in the data	c. They provide detailed statistical summaries of the dataset	d. They visualize the data distribution with histograms.	2
Which aspect of the data can be explored using dimensionality reduction visualization techniques?	a. Distribution of individual features	b. Relationships between pairs of features	c. Mean and standard deviation of the data	d. Number of data points in each cluster.	2
What does the df.plot.hist(bins=20) parameter do in Pandas?	a. Sets the color of the histogram bars to 20	b. Sets the transparency of the histogram bars to 20	c. Sets the number of bins in the histogram to 20	d. Sets the width of the histogram bars to 20	3
What is the main idea behind Principal Component Analysis (PCA) in unsupervised learning?	a. Dividing data into clusters	b. Reducing dimensionality by transforming features into a new set of uncorrelated variables	c. Classifying data points into categories	d. Predicting continuous outcomes	2
Which of the following is an application of unsupervised learning?	a. Image recognition	b. Predicting stock prices	c. Customer segmentation	d. Spam email detection	3
What does the term "clustering" refer to in the context of unsupervised learning?	a. Assigning a label to each data point	b. Dividing data points into groups with similar characteristics	c. Predicting a continuous numerical output	d. Reducing dimensionality of data	2
Which unsupervised learning technique is used to fill in missing values in a dataset?	a. Hierarchical clustering	b. Imputation	c. Dimensionality reduction	d. Anomaly detection	2

What is the role of the "elbow method" in K-Means clustering?	a. Identifying the optimal number of clusters by selecting the "elbow" point in the plot of the sum of squared distances	b. Assigning labels to data points	c. Determining outliers in the dataset	d. Reducing the dimensionality of the data	1
What is the primary limitation of hierarchical clustering in unsupervised learning?	a. Sensitivity to the initial choice of cluster centers	b. Difficulty in handling large datasets	c. Inability to form non-convex clusters	d. Lack of flexibility in the number of clusters	2
Which unsupervised learning technique is suitable for detecting anomalies or outliers in a dataset?	a. Principal Component Analysis (PCA)	b. Hierarchical clustering	c. K-Means clustering	d. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	4
What is the primary goal of t-Distributed Stochastic Neighbor Embedding (t-SNE) in unsupervised learning?	a. Dimensionality reduction	b. Clustering data points	c. Anomaly detection	d. Imputing missing values	1
Which unsupervised learning algorithm is used for density-based clustering?	a. K-Means	b. Hierarchical Clustering	c. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	d. PCA (Principal Component Analysis)	3
What is the primary objective of using t-Distributed Stochastic Neighbor Embedding (t-SNE) in unsupervised learning?	a. Classify data points into predefined categories	b. Dimensionality reduction for visualization	c. Predict continuous numerical output	d. Impute missing values in a dataset	2
Which type of neural network is commonly used for image recognition tasks?	a. Recurrent Neural Network (RNN)	b. Convolutional Neural Network (CNN)	c. Multilayer Perceptron (MLP)	d. Radial Basis Function Neural Network (RBFNN).	2
What is the purpose of pooling layers in a convolutional neural network (CNN)?	a. To reduce the dimensionality of feature maps	b. To increase the number of filters in the network	c. To introduce non-linearity into the network	d. To compute the dot product of weights and inputs.	1
Which of the following is a common optimization algorithm used for training neural networks?	a. Gradient Descent	b. K-Means	c. Decision Trees	d. Support Vector Machine.	1
What is the purpose of dropout regularization in neural networks?	a. To reduce the computational complexity of the network	b. To increase the learning rate during training	c. To prevent overfitting by randomly dropping neurons during training	d. To normalize the input data.	3
Which technique can be used to visualize the decision boundaries learned by a Multilayer Perceptron model trained using the MLPClassifier?	a. Confusion matrix	b. Gradient descent	c. Decision boundary plot	d. Activation maximization.	3
What is the primary advantage of using scikit-learn's MLPClassifier over implementing a neural network from scratch?	a. It provides a more flexible and customizable architecture	b. It is computationally more efficient for large datasets	c. It automatically handles backpropagation and optimization	d. It requires less memory for training.	3
Which parameter of the MLPClassifier class specifies the learning rate of the network?	a. learning_rate	b. learning_rate_init	c. learning_rate_schedule	d. learning_rate_decay.	2
What is the primary advantage of using mini-batch training over batch training for training a Multilayer Perceptron?	a. It reduces the computational complexity of the network	b. It converges faster to the optimal solution	c. It requires less memory for training	d. It introduces randomness into the optimization process.	4

Which regression evaluation metric measures the proportion of variance in the dependent variable that is predictable from the independent variables?	a. Mean Absolute Error (MAE)	b. Mean Squared Error (MSE)	c. Root Mean Squared Error (RMSE)	d. R-squared (R2).	4
What does the R-squared (R2) value indicate in the context of regression evaluation metrics?	a. The average absolute difference between the predicted and actual values of the target variable	b. The proportion of variance explained by the model	c. The square root of the average squared difference between the predicted and actual values of the target variable	d. The proportion of correctly classified instances out of all instances in the dataset.	2
In multiclass classification what does the confusion matrix represent?	a. The proportion of correctly classified instances out of all instances in the dataset	b. The average squared difference between the predicted and actual values of the target variable	c. The distribution of predicted classes compared to the actual classes	d. The proportion of variance explained by the model.	3
Which value in the confusion matrix represents the number of true negative predictions?	a. Top-left cell	b. Bottom-left cell	c. Top-right cell	d. Bottom-right cell.	1
Which step is involved in Principal Component Analysis (PCA)?	a. Randomly initializing cluster centroids	b. Assigning each data point to the nearest centroid	c. Identifying the principal components through eigenvector decomposition of the covariance matrix	d. All of the above.	3
What does the explained variance ratio represent in PCA?	a. The proportion of variance explained by each principal component	b. The average squared difference between the predicted and actual values of the target variable	c. The proportion of correctly classified instances out of all instances in the dataset	d. The trade-off between true positive rate and false positive rate.	1
Which unsupervised learning algorithm is used to detect outliers and cluster irregularly shaped datasets?	a. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	b. K-Means clustering	c. Hierarchical clustering	d. Mean Shift clustering.	1
What is the primary disadvantage of K-Means clustering?	a. It is sensitive to the order of data points	b. It cannot handle non-linearly separable data	c. It requires a predetermined number of clusters	d. It has a high computational complexity.	3
How does the choice of initial centroids affect K-Means clustering?	a. It affects the convergence of the algorithm	b. It determines the number of clusters	c. It influences the distance metric used	d. It has no impact on the clustering result.	1
What happens if the number of clusters (K) in K-Means clustering is set too low?	a. The algorithm converges faster	b. The clusters may be too coarse to capture meaningful patterns	c. The algorithm may not converge	d. The centroids become more stable.	2
Which method of the KMeans class is used to fit the K-Means clustering model to the data?	a. fit_predict	b. fit_transform	c. predict	d. transform.	1
What does the "max_iter" parameter in the KMeans class specify?	a. The number of clusters to form	b. The maximum number of iterations for convergence	c. The number of times the algorithm will be run with different centroid seeds	d. The number of features in the dataset.	2
Which method of the KMeans class is used to predict the closest cluster for each instance in the dataset?	a. fit	b. predict	c. fit_predict	d. transform.	2
What does a lower Davies-Bouldin Index value indicate about the clustering result?	a. Better separation between clusters and compactness within clusters	b. Poor separation between clusters and compactness within clusters	c. Better separation between clusters but poor compactness within clusters	d. Poor separation between clusters but good compactness within clusters.	1
Which metric is used to evaluate the similarity between two clusterings of the same dataset?	a. Davies-Bouldin Index	b. Silhouette score	c. Adjusted Rand Index	d. Calinski-Harabasz Index.	3
What does a higher Adjusted Rand Index value indicate about the similarity between two clusterings?	a. Higher similarity between the clusterings	b. Lower similarity between the clusterings	c. No similarity between the clusterings	d. Perfect agreement between the clusterings.	1

How does PCA achieve dimensionality reduction while preserving most of the variance?	a. By selecting a subset of features randomly	b. By eliminating features with low variance	c. By projecting the data onto a lower-dimensional subspace defined by the principal components	d. By increasing the dimensionality of the dataset.	3
What is the relationship between eigenvalues and principal components in PCA?	a. Eigenvalues represent the direction of the principal components	b. Eigenvalues represent the variance explained by each principal component	c. Eigenvalues represent the correlation between features	d. Eigenvalues represent the mean squared error of the model.	2
How are eigenvalues used to determine the importance of principal components in PCA?	a. Larger eigenvalues correspond to more important principal components	b. Smaller eigenvalues correspond to more important principal components	c. Eigenvalues are not used for this purpose	d. Eigenvalues are squared and then compared.	1
Which method of the PCA class is used to fit the PCA model to the data?	a. fit_transform	b. fit	c. transform	d. predict.	2
What does the "transform" method of the PCA class do?	a. Fits the PCA model to the data	b. Computes the covariance matrix	c. Projects the data onto the principal components	d. Selects the number of principal components.	3
Which parameter of the PCA class specifies whether or not to center the data before computing the covariance matrix?	a. n_components	b. whiten	c. copy	d. svd_solver.	2
How does PCA contribute to visualization of reduced dimensions?	a. By increasing the number of features	b. By reducing the dimensionality of the dataset while preserving most of its variance	c. By eliminating all dimensions except one	d. By introducing noise into the dataset for visualization purposes.	2
Which aspect of the data is typically visualized in reduced dimensions using PCA?	a. The covariance matrix	b. The correlation between features	c. The variance explained by each principal component	d. The number of observations in the dataset.	3
Which of the following statements about visualizing reduced dimensions is true?	a. It increases the interpretability of high-dimensional data	b. It preserves all dimensions of the original dataset	c. It introduces noise into the dataset for visualization purposes	d. It reduces the computational complexity of models.	1
How does scaling affect K-Means clustering?	a. It reduces the computational complexity	b. It makes the algorithm more sensitive to outliers	c. It has no impact on the clustering result	d. It improves the convergence of the algorithm.	2
In K-Means clustering what does the term "local optimum" refer to?	a. The global minimum of the objective function	b. The best clustering result achieved after multiple runs of the algorithm	c. A clustering solution that may not be the best possible globally	d. The distance between data points and centroids.	3
What is an advantage of density-based clustering algorithms?	a. They require a predefined number of clusters	b. They are not suitable for high-dimensional data	c. They can handle clusters of arbitrary shapes and sizes	d. They are computationally complex.	3
Which of the following is an application of unsupervised learning?	a. Sentiment analysis	b. Image classification	c. Anomaly detection	d. Handwriting recognition.	3
What is a limitation of K-means clustering?	a. It is sensitive to outliers	b. It requires a large amount of computational resources	c. It cannot handle non-linearly separable data	d. It always converges to the global optimum.	1
What is the convergence criterion in K-means clustering?	a. When the within-cluster sum of squares is minimized	b. When the between-cluster sum of squares is maximized	c. When the centroids stop changing significantly between iterations	d. When the number of clusters stabilizes.	3
Which technique is commonly used to initialize cluster centroids in K-means clustering?	a. Random initialization	b. K-nearest neighbors	c. Decision trees	d. Support Vector Machines (SVM).	1
What happens if K-means clustering fails to converge?	a. The algorithm restarts with different initial centroids	b. The algorithm assigns data points randomly to clusters	c. The algorithm terminates and returns an error	d. The algorithm stops and returns the current clusters.	3
What distinguishes clustering from classification?	a. Clustering involves grouping data points based on similarity while classification assigns predefined labels to data points.	b. Clustering requires labeled data while classification does not.	c. Clustering is a supervised learning technique while classification is unsupervised.	d. Clustering can only handle numerical data while classification can handle categorical data.	1
Which of the following is NOT a common application of clustering?	a. Customer segmentation	b. Image classification	c. Anomaly detection	d. Document clustering.	2

What is the main advantage of unsupervised learning techniques like clustering?	a. They require less computational resources compared to supervised learning	b. They do not require labeled data for training	c. They always produce accurate predictions	d. They are more interpretable than supervised learning techniques.	2
Which parameter of the KMeans class specifies the number of clusters to form?	a. n_clusters	b. n_init	c. max_iter	d. random_state.	1
What is the purpose of the fit() method in scikit-learn's KMeans class?	a. It initializes the cluster centroids	b. It assigns each data point to the nearest cluster	c. It updates the cluster centroids based on the data	d. It evaluates the quality of clustering.	2
Which attribute of the KMeans object in scikit-learn stores the cluster centroids?	a. cluster_centers_	b. labels_	c. inertia_	d. n_clusters.	1
Which method can be used to determine the optimal number of clusters when using KMeans in scikit-learn?	a. Elbow method	b. Silhouette method	c. Gap statistic method	d. All of the above.	4
What range of values does the Silhouette coefficient typically lie between?	a. [-1, 0]	b. [0, 1]	c. [0, ∞)	d. [-∞ ∞).	1
In the context of clustering evaluation what does a Silhouette coefficient close to 1 indicate?	a. Good clustering	b. Poor clustering	c. Random clustering	d. No clustering.	1
What is the Davies-Bouldin index used for in clustering evaluation?	a. Measuring the compactness and separation of clusters	b. Calculating the distance between clusters	c. Assessing the similarity between two clusterings	d. Evaluating the density and separation of clusters.	1
What does the term "principal components" refer to in PCA?	a. The features with the highest variance	b. The original features of the dataset	c. The eigenvectors of the covariance matrix	d. The target variables of the dataset.	3
In PCA how are principal components ordered?	a. In ascending order of their eigenvalues	b. In descending order of their eigenvalues	c. In random order	d. In the order they appear in the dataset.	2
What is the significance of the eigenvalues in PCA?	a. They represent the amount of variance explained by each principal component	b. They determine the number of features in the dataset	c. They indicate the number of clusters formed by the data	d. They represent the number of observations in the dataset.	1
Which technique aims to find a lower-dimensional representation of the data that preserves the pairwise distances between data points?	a. Principal Component Analysis (PCA)	b. Linear Discriminant Analysis (LDA)	c. t-Distributed Stochastic Neighbor Embedding (t-SNE)	d. Autoencoders.	3
What is the primary advantage of using PCA for dimensionality reduction?	a. It is computationally efficient	b. It preserves the class separability of the data	c. It is robust to outliers	d. It captures the maximum variance in the data.	4
Which dimensionality reduction technique is particularly useful for visualizing high-dimensional data in low-dimensional space?	a. PCA	b. LDA	c. t-SNE	d. Random Forest.	3
What is the purpose of the fit() method in scikit-learn's PCA class?	a. It transforms the data into principal components	b. It computes the mean and standard deviation of the data	c. It fits the PCA model to the data	d. It computes the explained variance ratio.	3
Which attribute of the PCA object in scikit-learn stores the principal axes in feature space?	a. components_	b. explained_variance_	c. explained_variance_ratio_	d. mean_.	1
What does the explained_variance_ratio_ attribute represent in scikit-learn's PCA object?	a. The variance of each principal component	b. The total variance explained by all the principal components	c. The eigenvalues of the covariance matrix	d. The mean of the data.	2
What does each point in a scatter plot represent when visualizing reduced dimensions using PCA?	a. A feature of the dataset	b. A dimension of the dataset	c. A data point in the high-dimensional space	d. An outlier in the dataset.	3

How can you interpret the distance between points in a scatter plot visualization of reduced dimensions?	a. It represents the variance explained by each component	b. It indicates the correlation between features	c. It shows the similarity between data points in the reduced space	d. It signifies the number of dimensions in the dataset.	3
What is the primary advantage of using scatter plots for dimensionality reduction visualization?	a. They can handle high-dimensional data efficiently	b. They allow easy identification of clusters and patterns in the data	c. They provide detailed statistical summaries of the dataset	d. They visualize the data distribution with histograms.	2
Which aspect of the data can be explored using dimensionality reduction visualization techniques?	a. Distribution of individual features	b. Relationships between pairs of features	c. Mean and standard deviation of the data	d. Number of data points in each cluster.	2
What does the df.plot.hist(bins=20) parameter do in Pandas?	a. Sets the color of the histogram bars to 20	b. Sets the transparency of the histogram bars to 20	c. Sets the number of bins in the histogram to 20	d. Sets the width of the histogram bars to 20	3
What is the main idea behind Principal Component Analysis (PCA) in unsupervised learning?	a. Dividing data into clusters	b. Reducing dimensionality by transforming features into a new set of uncorrelated variables	c. Classifying data points into categories	d. Predicting continuous outcomes	2
Which of the following is an application of unsupervised learning?	a. Image recognition	b. Predicting stock prices	c. Customer segmentation	d. Spam email detection	3
What does the term "clustering" refer to in the context of unsupervised learning?	a. Assigning a label to each data point	b. Dividing data points into groups with similar characteristics	c. Predicting a continuous numerical output	d. Reducing dimensionality of data	2
Which unsupervised learning technique is used to fill in missing values in a dataset?	a. Hierarchical clustering	b. Imputation	c. Dimensionality reduction	d. Anomaly detection	2
What is the role of the "elbow method" in K-Means clustering?	a. Identifying the optimal number of clusters by selecting the "elbow" point in the plot of the sum of squared distances	b. Assigning labels to data points	c. Determining outliers in the dataset	d. Reducing the dimensionality of the data	1
What is the primary limitation of hierarchical clustering in unsupervised learning?	a. Sensitivity to the initial choice of cluster centers	b. Difficulty in handling large datasets	c. Inability to form non-convex clusters	d. Lack of flexibility in the number of clusters	2
Which unsupervised learning technique is suitable for detecting anomalies or outliers in a dataset?	a. Principal Component Analysis (PCA)	b. Hierarchical clustering	c. K-Means clustering	d. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	4
What is the primary goal of t-Distributed Stochastic Neighbor Embedding (t-SNE) in unsupervised learning?	a. Dimensionality reduction	b. Clustering data points	c. Anomaly detection	d. Imputing missing values	1
Which unsupervised learning algorithm is used for density-based clustering?	a. K-Means	b. Hierarchical Clustering	c. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	d. PCA (Principal Component Analysis)	3

What is the primary objective of using t-Distributed Stochastic Neighbor Embedding (t-SNE) in unsupervised learning?	a. Classify data points into predefined categories	b. Dimensionality reduction for visualization	c. Predict continuous numerical output	d. Impute missing values in a dataset	2
Which type of neural network is commonly used for image recognition tasks?	a. Recurrent Neural Network (RNN)	b. Convolutional Neural Network (CNN)	c. Multilayer Perceptron (MLP)	d. Radial Basis Function Neural Network (RBFNN).	2
What is the purpose of pooling layers in a convolutional neural network (CNN)?	a. To reduce the dimensionality of feature maps	b. To increase the number of filters in the network	c. To introduce non-linearity into the network	d. To compute the dot product of weights and inputs.	1
Which of the following is a common optimization algorithm used for training neural networks?	a. Gradient Descent	b. K-Means	c. Decision Trees	d. Support Vector Machine.	1
What is the purpose of dropout regularization in neural networks?	a. To reduce the computational complexity of the network	b. To increase the learning rate during training	c. To prevent overfitting by randomly dropping neurons during training	d. To normalize the input data.	3
Which technique can be used to visualize the decision boundaries learned by a Multilayer Perceptron model trained using the MLPClassifier?	a. Confusion matrix	b. Gradient descent	c. Decision boundary plot	d. Activation maximization.	3
What is the primary advantage of using scikit-learn's MLPClassifier over implementing a neural network from scratch?	a. It provides a more flexible and customizable architecture	b. It is computationally more efficient for large datasets	c. It automatically handles backpropagation and optimization	d. It requires less memory for training.	3
Which parameter of the MLPClassifier class specifies the learning rate of the network?	a. learning_rate	b. learning_rate_init	c. learning_rate_schedule	d. learning_rate_decay.	2
What is the primary advantage of using mini-batch training over batch training for training a Multilayer Perceptron?	a. It reduces the computational complexity of the network	b. It converges faster to the optimal solution	c. It requires less memory for training	d. It introduces randomness into the optimization process.	4
Which regression evaluation metric measures the proportion of variance in the dependent variable that is predictable from the independent variables?	a. Mean Absolute Error (MAE)	b. Mean Squared Error (MSE)	c. Root Mean Squared Error (RMSE)	d. R-squared (R2).	4
What does the R-squared (R2) value indicate in the context of regression evaluation metrics?	a. The average absolute difference between the predicted and actual values of the target variable	b. The proportion of variance explained by the model	c. The square root of the average squared difference between the predicted and actual values of the target variable	d. The proportion of correctly classified instances out of all instances in the dataset.	2
In multiclass classification what does the confusion matrix represent?	a. The proportion of correctly classified instances out of all instances in the dataset	b. The average squared difference between the predicted and actual values of the target variable	c. The distribution of predicted classes compared to the actual classes	d. The proportion of variance explained by the model.	3
Which value in the confusion matrix represents the number of true negative predictions?	a. Top-left cell	b. Bottom-left cell	c. Top-right cell	d. Bottom-right cell.	1
Which step is involved in Principal Component Analysis (PCA)?	a. Randomly initializing cluster centroids	b. Assigning each data point to the nearest centroid	c. Identifying the principal components through eigenvector decomposition of the covariance matrix	d. All of the above.	3

What does the explained variance ratio represent in PCA?	a. The proportion of variance explained by each principal component	b. The average squared difference between the predicted and actual values of the target variable	c. The proportion of correctly classified instances out of all instances in the dataset	d. The trade-off between true positive rate and false positive rate.	1
Which unsupervised learning algorithm is used to detect outliers and cluster irregularly shaped datasets?	a. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	b. K-Means clustering	c. Hierarchical clustering	d. Mean Shift clustering.	1
What is the primary disadvantage of K-Means clustering?	a. It is sensitive to the order of data points	b. It cannot handle non-linearly separable data	c. It requires a predetermined number of clusters	d. It has a high computational complexity.	3
How does the choice of initial centroids affect K-Means clustering?	a. It affects the convergence of the algorithm	b. It determines the number of clusters	c. It influences the distance metric used	d. It has no impact on the clustering result.	1
What happens if the number of clusters (K) in K-Means clustering is set too low?	a. The algorithm converges faster	b. The clusters may be too coarse to capture meaningful patterns	c. The algorithm may not converge	d. The centroids become more stable.	2
Which method of the KMeans class is used to fit the K-Means clustering model to the data?	a. fit_predict	b. fit_transform	c. predict	d. transform.	1
What does the "max_iter" parameter in the KMeans class specify?	a. The number of clusters to form	b. The maximum number of iterations for convergence	c. The number of times the algorithm will be run with different centroid seeds	d. The number of features in the dataset.	2
Which method of the KMeans class is used to predict the closest cluster for each instance in the dataset?	a. fit	b. predict	c. fit_predict	d. transform.	2
What does a lower Davies-Bouldin Index value indicate about the clustering result?	a. Better separation between clusters and compactness within clusters	b. Poor separation between clusters and compactness within clusters	c. Better separation between clusters but poor compactness within clusters	d. Poor separation between clusters but good compactness within clusters.	1
Which metric is used to evaluate the similarity between two clusterings of the same dataset?	a. Davies-Bouldin Index	b. Silhouette score	c. Adjusted Rand Index	d. Calinski-Harabasz Index.	3
What does a higher Adjusted Rand Index value indicate about the similarity between two clusterings?	a. Higher similarity between the clusterings	b. Lower similarity between the clusterings	c. No similarity between the clusterings	d. Perfect agreement between the clusterings.	1
How does PCA achieve dimensionality reduction while preserving most of the variance?	a. By selecting a subset of features randomly	b. By eliminating features with low variance	c. By projecting the data onto a lower-dimensional subspace defined by the principal components	d. By increasing the dimensionality of the dataset.	3
What is the relationship between eigenvalues and principal components in PCA?	a. Eigenvalues represent the direction of the principal components	b. Eigenvalues represent the variance explained by each principal component	c. Eigenvalues represent the correlation between features	d. Eigenvalues represent the mean squared error of the model.	2
How are eigenvalues used to determine the importance of principal components in PCA?	a. Larger eigenvalues correspond to more important principal components	b. Smaller eigenvalues correspond to more important principal components	c. Eigenvalues are not used for this purpose	d. Eigenvalues are squared and then compared.	1
Which method of the PCA class is used to fit the PCA model to the data?	a. fit_transform	b. fit	c. transform	d. predict.	2
What does the "transform" method of the PCA class do?	a. Fits the PCA model to the data	b. Computes the covariance matrix	c. Projects the data onto the principal components	d. Selects the number of principal components.	3
Which parameter of the PCA class specifies whether or not to center the data before computing the covariance matrix?	a. n_components	b. whiten	c. copy	d. svd_solver.	2

How does PCA contribute to visualization of reduced dimensions?	a. By increasing the number of features	b. By reducing the dimensionality of the dataset while preserving most of its variance	c. By eliminating all dimensions except one	d. By introducing noise into the dataset for visualization purposes.	2
Which aspect of the data is typically visualized in reduced dimensions using PCA?	a. The covariance matrix	b. The correlation between features	c. The variance explained by each principal component	d. The number of observations in the dataset.	3
Which of the following statements about visualizing reduced dimensions is true?	a. It increases the interpretability of high-dimensional data	b. It preserves all dimensions of the original dataset	c. It introduces noise into the dataset for visualization purposes	d. It reduces the computational complexity of models.	1
How does scaling affect K-Means clustering?	a. It reduces the computational complexity	b. It makes the algorithm more sensitive to outliers	c. It has no impact on the clustering result	d. It improves the convergence of the algorithm.	2
In K-Means clustering what does the term "local optimum" refer to?	a. The global minimum of the objective function	b. The best clustering result achieved after multiple runs of the algorithm	c. A clustering solution that may not be the best possible globally	d. The distance between data points and centroids.	3
Which technique is commonly used for data preprocessing before applying unsupervised learning algorithms?	a. Feature scaling	b. Feature engineering	c. Regularization	d. Cross-validation.	1
What is a key advantage of unsupervised learning?	a. It requires less computational resources compared to supervised learning	b. It does not require labeled data	c. It always produces better results than supervised learning	d. It is more interpretable than supervised learning.	2
In which scenario would density-based clustering algorithms like DBSCAN perform well?	a. When the number of clusters is known beforehand	b. When the clusters are well-separated	c. When the clusters have irregular shapes	d. When the data is linearly separable.	3
Which technique is used to evaluate the quality of clustering in unsupervised learning?	a. F1 score	b. Silhouette coefficient	c. Area Under the ROC Curve (AUC)	d. Mean Squared Error (MSE).	2
What type of algorithm is K-means clustering?	a. Supervised learning	b. Unsupervised learning	c. Reinforcement learning	d. Semi-supervised learning.	2
What is the objective of K-means clustering?	a. To minimize the within-cluster sum of squares	b. To maximize the between-cluster sum of squares	c. To classify data into predefined categories	d. To predict an output variable based on input data alone.	1
How does K-means clustering initialize cluster centroids?	a. Randomly	b. Based on the median of the data	c. Through gradient descent	d. By predefined rules.	1
What is the primary advantage of partitioning clustering algorithms like K-means?	a. They can handle clusters of arbitrary shapes and sizes	b. They are computationally efficient	c. They do not require the number of clusters to be specified beforehand	d. They do not require distance or similarity measures.	2
What is the main purpose of dimensionality reduction techniques in clustering?	a. To increase the interpretability of the clustering results	b. To reduce the computational complexity of clustering algorithms	c. To visualize high-dimensional data in lower dimensions	d. To improve the accuracy of clustering.	3
How does density-based clustering differ from partitioning clustering?	a. Density-based clustering requires the number of clusters to be specified beforehand while partitioning clustering does not.	b. Density-based clustering forms clusters based on data density while partitioning clustering divides data into distinct groups.	c. Density-based clustering always produces spherical clusters while partitioning clustering does not.	d. Density-based clustering is computationally more efficient than partitioning clustering.	2
Which clustering algorithm is based on constructing a similarity graph of data points?	a. K-means clustering	b. Hierarchical clustering	c. DBSCAN	d. Spectral clustering.	4
What does the tol parameter control in scikit-learn's KMeans class?	a. The initialization of cluster centroids	b. The maximum number of iterations for the algorithm	c. The tolerance to declare convergence	d. The number of times the algorithm will be run with different centroid seeds.	3
How can you access the labels assigned to each data point after clustering using KMeans in scikit-learn?	a. By calling the fit() method	b. By accessing the labels_ attribute of the KMeans object	c. By calling the predict() method	d. By accessing the cluster_centers_ attribute.	2

What is the impact of selecting a different number of clusters (K) in K-means clustering?	a. It does not affect the clustering result	b. It affects the convergence of the algorithm	c. It affects the interpretability of the clusters	d. It affects the quality of the clustering result.	4
What is clustering in machine learning?	a. A technique used to classify data points into predefined categories	b. An unsupervised learning technique for grouping similar data points together	c. A method for predicting continuous output variables	d. A reinforcement learning algorithm.	2
What is the primary goal of clustering?	a. To minimize the within-cluster sum of squares	b. To maximize the between-cluster sum of squares	c. To predict an output variable based on input data alone	d. To discover underlying patterns and structures in data.	4
Which of the following statements about clustering is true?	a. Clustering requires labeled data for training	b. Clustering involves assigning data points to predefined categories	c. Clustering can be supervised or unsupervised	d. Clustering algorithms always produce accurate predictions.	3
Which library in Python is commonly used for implementing K-means clustering?	a. TensorFlow	b. Scikit-learn	c. PyTorch	d. Pandas.	2
How can you import the KMeans class from scikit-learn?	a. from sklearn.cluster import KMeans	b. from sklearn import KMeans	c. from sklearn.kmeans import KMeans	d. import KMeans from sklearn.cluster.	1
What is the first step in using KMeans in scikit-learn?	a. Initializing the cluster centroids	b. Fitting the KMeans model to the data	c. Importing the necessary libraries	d. Preprocessing the data.	3
What is the primary goal of evaluating clustering results?	a. To determine the computational complexity of the clustering algorithm	b. To assess the quality of the clusters formed	c. To visualize the data points in lower dimensions	d. To preprocess the data before clustering.	2
Which metric measures the compactness of clusters in clustering evaluation?	a. Silhouette coefficient	b. Davies-Bouldin index	c. Rand index	d. Adjusted Rand index.	1
What does the Silhouette coefficient measure?	a. The distance between clusters	b. The separation between clusters	c. The compactness and separation of clusters	d. The purity of clusters.	3
What is the primary objective of Principal Component Analysis (PCA)?	a. To maximize the variance of the data	b. To minimize the dimensionality of the data	c. To classify data points into predefined categories	d. To predict an output variable based on input data alone.	2
Which of the following statements about PCA is true?	a. PCA is a supervised learning technique	b. PCA can only handle categorical data	c. PCA finds orthogonal axes that best represent the data variance	d. PCA requires labeled data for training.	3
How does PCA achieve dimensionality reduction?	a. By removing outliers from the data	b. By selecting the most important features	c. By projecting the data onto a lower-dimensional subspace	d. By increasing the number of features.	3
Which clustering algorithm is particularly useful for identifying clusters of arbitrary shapes and sizes?	a. K-means clustering	b. Hierarchical clustering	c. DBSCAN	d. Spectral clustering.	3
What does the "elbow method" help determine in clustering?	a. The optimal number of clusters	b. The optimal distance metric	c. The optimal cluster initialization	d. The optimal convergence criterion.	1
Which parameter of the KMeans class specifies the maximum number of iterations for the algorithm?	a. max_iter	b. n_clusters	c. n_init	d. tol.	1
What does the Rand index measure in clustering evaluation?	a. The compactness of clusters	b. The similarity between two clusterings	c. The density of clusters	d. The separation between clusters.	2
How can you create a scatter plot in Pandas?	a. df.plot.scatter(x='column1', y='column2')	b. df.plot.line(x='column1', y='column2')	c. df.plot.bar(x='column1', y='column2')	d. df.plot.pie(x='column1', y='column2')	1
What is the purpose of the df.plot.scatter(c='column3') parameter in Pandas?	a. Sets the color of the scatter points based on values in 'column3'	b. Sets the size of the scatter points based on values in 'column3'	c. Connects the scatter points with lines based on values in 'column3'	d. Sets the transparency of the scatter points based on values in 'column3'	1
What is the output of print(3 * 'abc')?	a. abc	b. abcabc	c. 333	d. 9	2

What does the len() function do in Python?	a. Returns the length of a list or string	b. Performs mathematical operations	c. Defines a new variable	d. Executes a loop	1
What is the purpose of the input() function in Python?	a. Display output on the console	b. Read user input from the console	c. Perform mathematical calculations	d. Create a graphical user interface	2
How does PCA handle multicollinearity in the dataset?	a. By removing features with low variance	b. By transforming the original features into orthogonal components	c. By normalizing the data	d. By adding new features to the dataset.	2
What is a limitation of PCA?	a. It requires labeled data for training	b. It cannot handle non-linear relationships between variables	c. It always produces interpretable results	d. It is computationally expensive.	2
Which step is typically performed after applying PCA for dimensionality reduction?	a. Visualizing the data in lower dimensions	b. Reconstructing the original features	c. Applying clustering algorithms	d. Preprocessing the data.	2
What is a limitation of dimensionality reduction techniques?	a. They always improve the interpretability of the data	b. They may lead to information loss	c. They are only applicable to small datasets	d. They cannot handle categorical variables.	2
Which dimensionality reduction technique focuses on maximizing the separation between classes in the data?	a. Principal Component Analysis (PCA)	b. Linear Discriminant Analysis (LDA)	c. Independent Component Analysis (ICA)	d. Non-negative Matrix Factorization (NMF).	2
Which type of neural network architecture is commonly used for sequence generation tasks such as text generation?	a. Recurrent Neural Network (RNN)	b. Convolutional Neural Network (CNN)	c. Multilayer Perceptron (MLP)	d. Radial Basis Function Neural Network (RBFNN).	1
What is the primary objective of transfer learning in neural networks?	a. To transfer knowledge from one domain to another	b. To transfer weights from one layer to another	c. To transfer data between different neural networks	d. To transfer activations between layers.	1
Which technique can be used to visualize the learned features in the hidden layers of a neural network?	a. Activation maximization	b. Gradient descent	c. Principal Component Analysis (PCA)	d. K-Means clustering.	1
In neural networks what is the purpose of mini-batch training?	a. To speed up the convergence of the optimization algorithm	b. To reduce the computational complexity of the network	c. To introduce randomness into the network	d. To update the weights of the network using a subset of the training data at each iteration.	4
What does the solver parameter control in the MLPClassifier class?	a. The optimization algorithm used to train the network	b. The number of iterations for training	c. The activation function for hidden layers	d. The learning rate schedule.	1
Which optimization algorithm is commonly used as the default solver in the MLPClassifier class?	a. Stochastic Gradient Descent (SGD)	b. Adam	c. Levenberg-Marquardt	d. L-BFGS.	1
What is the purpose of the alpha parameter in the MLPClassifier class?	a. To control the number of hidden layers in the network	b. To adjust the regularization strength	c. To specify the learning rate of the network	d. To determine the batch size for training.	2
How can you specify the activation function for the output layer of a Multilayer Perceptron using the MLPClassifier?	a. By setting the output_activation parameter	b. By passing a string to the activation parameter	c. By adjusting the learning rate	d. By setting the output_function parameter.	2
What does the Mean Average Precision (MAP) measure in the context of ranking evaluation metrics?	a. The average precision over all documents retrieved for each query	b. The proportion of relevant documents retrieved out of all relevant documents in the dataset	c. The harmonic mean of precision and recall	d. The proportion of correctly classified instances out of all instances in the dataset.	1
Which evaluation metric is commonly used for regression tasks to measure the goodness of fit between predicted and actual values?	a. Mean Absolute Error (MAE)	b. Mean Squared Error (MSE)	c. Root Mean Squared Error (RMSE)	d. R-squared (R2).	4
Which unsupervised learning algorithm is used to partition a dataset into a predetermined number of clusters?	a. K-Means clustering	b. Decision Trees	c. Support Vector Machines	d. Random Forests.	1

What is the primary objective of K-Means clustering?	a. To maximize the variance within clusters	b. To minimize the number of clusters	c. To minimize the distance between data points and centroids	d. To maximize the silhouette coefficient.	3
Which unsupervised learning algorithm is used for community detection in networks or graphs?	a. Spectral clustering	b. K-Means clustering	c. Agglomerative clustering	d. DBSCAN.	1
What is the primary advantage of spectral clustering over K-Means clustering?	a. It can handle non-linearly separable data	b. It is less sensitive to the choice of initialization	c. It does not require a predetermined number of clusters	d. It can capture complex relationships between data points.	3
Which unsupervised learning algorithm is used for reducing the dimensionality of text data?	a. Latent Semantic Analysis (LSA)	b. K-Means clustering	c. DBSCAN	d. Hierarchical clustering.	1
What is the primary disadvantage of the silhouette method for determining the optimal number of clusters?	a. It requires a predetermined number of clusters	b. It is sensitive to the shape of the clusters	c. It may not always produce meaningful results	d. It has a high computational complexity.	3
In K-Means clustering how are outliers typically handled?	a. By assigning them to the nearest cluster	b. By removing them from the dataset	c. By creating a separate cluster for outliers	d. By initializing centroids closer to the outliers.	1
What is the primary advantage of K-Means clustering over hierarchical clustering?	a. It produces a dendrogram	b. It is less sensitive to the choice of distance metric	c. It has a lower computational complexity	d. It can handle non-linearly separable data.	3
What is the default value of the "n_init" parameter in the KMeans class?	a. 10	b. 100	c. 300	d. 1000.	1
How does the choice of initialization method affect the performance of K-Means clustering in scikit-learn?	a. It has no impact on the clustering result	b. It affects the convergence of the algorithm	c. It determines the number of clusters to form	d. It influences the shape of the clusters formed.	2
Which method of the KMeans class is used to fit the model to the data and predict the cluster labels in a single step?	a. fit	b. predict	c. fit_predict	d. transform.	3
What does a lower Davies-Bouldin Index value indicate about the compactness within clusters?	a. Better compactness within clusters	b. Poor compactness within clusters	c. No compactness within clusters	d. Perfect compactness within clusters.	1
Which metric is used to evaluate the separation between clusters in clustering evaluation?	a. Davies-Bouldin Index	b. Silhouette score	c. Calinski-Harabasz Index	d. Adjusted Rand Index.	1
What does a lower Davies-Bouldin Index value indicate about the separation between clusters?	a. Better separation between clusters	b. Poor separation between clusters	c. No separation between clusters	d. Perfect separation between clusters.	1
How does PCA handle multicollinearity in the dataset?	a. By increasing the correlation between features	b. By decreasing the correlation between features	c. By introducing noise into the dataset	d. By ignoring features with high correlation.	2
What is the primary disadvantage of PCA?	a. It increases the interpretability of models	b. It may lead to information loss	c. It reduces the computational complexity of models	d. It introduces noise into the dataset.	2
What does PCA stand for?	a. Principal Correlation Analysis	b. Primary Component Analysis	c. Principal Component Adjustment	d. Principal Component Analysis.	4
What is the purpose of setting the "iterated_power" parameter in the PCA class?	a. To specify the number of singular values to compute	b. To specify the method used for singular value decomposition	c. To control the maximum number of iterations	d. To determine the random number generator for initialization.	3
How does PCA handle missing values in the dataset?	a. By imputing the missing values	b. By removing the observations with missing values	c. By ignoring the missing values during computation	d. By replacing the missing values with zeros.	3
Which parameter of the PCA class specifies the method used for singular value decomposition?	a. n_components	b. whiten	c. copy	d. svd_solver.	4

Which visualization technique is suitable for visualizing reduced dimensions in a non-linear manner?	a. PCA	b. t-SNE	c. Scatter plot	d. Histogram.	2
How does PCA contribute to visualization of reduced dimensions?	a. By preserving local structures in the data	b. By introducing noise into the dataset	c. By reducing the dimensionality of the dataset while preserving most of its variance	d. By preserving global structures in the data.	3
What is the primary objective of visualizing reduced dimensions using histograms?	a. To visualize the distribution of a single variable	b. To visualize the relationship between two or more variables in reduced dimensions	c. To visualize the correlation matrix of the dataset	d. To introduce noise into the dataset for visualization purposes.	1
In PCA what does the covariance matrix represent?	a. The relationship between features in the dataset	b. The relationship between observations in the dataset	c. The variance of the dataset	d. The mean of the dataset.	1
What is the role of eigenvectors in PCA?	a. They represent the variance of the dataset	b. They define the directions of the principal components	c. They measure the correlation between features	d. They represent the original features of the dataset.	2
What is the main challenge of hierarchical clustering?	a. Determining the number of clusters	b. Handling high-dimensional data	c. Dealing with outliers	d. Converging to the global optimum.	1
Which unsupervised learning algorithm uses a distance-based approach to form clusters?	a. K-means clustering	b. PCA	c. DBSCAN	d. Decision trees.	3
What is the primary goal of clustering in unsupervised learning?	a. To minimize the within-cluster sum of squares	b. To maximize the between-cluster sum of squares	c. To classify data into predefined categories	d. To group similar data points together.	4
What does the "K" in K-means clustering represent?	a. The number of features in the dataset	b. The number of clusters	c. The number of iterations	d. The number of data points.	2
What is the main limitation of K-means clustering?	a. It is sensitive to outliers	b. It cannot handle high-dimensional data	c. It requires a large number of clusters	d. It always converges to the global optimum.	1
How is the number of clusters typically determined in K-means clustering?	a. By trial and error	b. By minimizing the within-cluster sum of squares	c. By maximizing the between-cluster sum of squares	d. By setting an arbitrary number.	1
What is the impact of selecting a different number of clusters (K) in K-means clustering?	a. It does not affect the clustering result	b. It affects the convergence of the algorithm	c. It affects the interpretability of the clusters	d. It affects the quality of the clustering result.	4
What is the main drawback of density-based clustering algorithms?	a. They cannot handle large datasets	b. They are sensitive to the initial placement of centroids	c. They require the number of clusters to be specified beforehand	d. They may struggle with clusters of varying densities.	4
In clustering what does the "silhouette coefficient" measure?	a. The compactness of clusters	b. The separation between clusters	c. The homogeneity of clusters	d. The quality of clusters.	4
Which clustering algorithm forms clusters by iteratively merging or splitting existing clusters?	a. K-means clustering	b. Hierarchical clustering	c. DBSCAN	d. Spectral clustering.	2
How can you calculate the sum of squared distances of samples to their closest cluster center in scikit-learn?	a. By accessing the inertia_ attribute of the KMeans object	b. By calling the fit() method	c. By using the predict() method	d. By accessing the cluster_centers_ attribute.	1
What is the purpose of the transform() method in scikit-learn's KMeans class?	a. To assign cluster labels to new data points	b. To compute the distance between data points and cluster centroids	c. To update the cluster centroids based on new data	d. To evaluate the quality of clustering.	2
Which of the following is NOT a valid parameter for the KMeans class in scikit-learn?	a. learning_rate	b. n_clusters	c. init	d. random_state.	1
What is the computational complexity of the KMeans algorithm in scikit-learn?	a. O(n)	b. O(n log n)	c. O(n^2)	d. O(kn) where k is the number of clusters and n is the number of data points.	4
How does the Adjusted Rand index handle chance agreement?	a. It does not consider chance agreement	b. It penalizes chance agreement by comparing it to random clusterings	c. It assumes no chance agreement in the data	d. It adds a correction term to account for chance agreement.	4

What is the range of values for the Adjusted Rand index?	a. [-1, 1]	b. [0, 1]	c. [0, ∞)	d. [-∞, ∞).	1
In the context of clustering evaluation what does an Adjusted Rand index close to 0 indicate?	a. Good agreement between two clusterings	b. Poor agreement between two clusterings	c. No agreement between two clusterings	d. No clustering.	3
What is the primary advantage of using PCA for feature extraction?	a. It reduces the computational complexity of the dataset	b. It retains the interpretability of the original features	c. It removes outliers from the data	d. It captures the maximum variance in the dataset.	4
What is the significance of the explained variance ratio in PCA?	a. It indicates the percentage of variance explained by each principal component	b. It determines the number of clusters formed by the data	c. It measures the purity of clusters	d. It represents the eigenvalues of the covariance matrix.	1
Which of the following is NOT a common application of PCA?	a. Image compression	b. Anomaly detection	c. Text classification	d. Signal processing.	3
Which dimensionality reduction technique is commonly used for feature extraction in classification tasks?	a. PCA	b. LDA	c. t-SNE	d. Isomap.	2
In which domain is dimensionality reduction NOT commonly used?	a. Computer vision	b. Natural language processing	c. Fraud detection	d. Time series forecasting.	4
Which of the following statements about autoencoders is true?	a. Autoencoders are unsupervised learning techniques	b. Autoencoders cannot handle non-linear data	c. Autoencoders do not perform dimensionality reduction	d. Autoencoders require labeled data for training.	1
What is the primary purpose of applying dimensionality reduction techniques like PCA before clustering?	a. To increase the computational complexity of clustering algorithms	b. To improve the interpretability of clustering results	c. To reduce the dimensionality of data and speed up clustering	d. To introduce noise into the data.	3
What does the svd_solver parameter control in scikit-learn's PCA class?	a. The method used for matrix decomposition	b. The number of components to keep	c. The algorithm used for centering the data	d. The initialization of the PCA algorithm.	1
How does PCA handle missing values in the dataset by default in scikit-learn?	a. It imputes missing values with the mean of the feature	b. It removes features with missing values	c. It imputes missing values with the median of the feature	d. It imputes missing values with the mode of the feature.	2
Which method can be used to determine the optimal number of components to retain in PCA?	a. Elbow method	b. Silhouette method	c. Scree plot	d. Gap statistic method.	3
What is the primary advantage of using PCA before applying machine learning algorithms?	a. It increases the computational complexity of the dataset	b. It reduces the interpretability of the machine learning models	c. It speeds up the training of machine learning models	d. It reduces the dimensionality of the dataset.	4
How can you interpret the colors in a Dimensionality Reduction map?	a. They represent the frequency of data points in each cluster	b. They indicate the distance between data points	c. They signify the variance explained by each principal component	d. They depict the similarity of data points in the reduced space.	1
Which visualization technique is suitable for exploring the structure of clusters formed by the data?	a. Heatmap	b. 3D Scatter plot	c. Parallel Coordinates plot	d. Clustering map.	4
What is Python?	a. A high-level programming language	b. A type of snake	c. A character in a video game	d. A piece of hardware	1
How do you denote a single-line comment in Python?	a. //	b. #	c. /* */	d. ---	2
Which of the following is the correct way to declare a variable in Python?	a. variable x;	b. x := 5	c. x = 5	d. int x = 5	3
. What is the correct way to create a function in Python?	a. def function_name():	b. function function_name():	c. function_name def():	d. function_name() {	1

How do you open a file named "example.txt" in Python for writing?	a. file = open("example.txt", "r")	b. file = open("example.txt", "w")	c. file = open("example.txt", "a")	d. file = open("example.txt", "x")	2
What is the purpose of the break statement in Python?	a. Ends the current loop or iteration	b. Skips the next iteration of the loop	c. Exits the entire program	d. Continues to the next loop	1
What is the purpose of the else clause in a Python if-else statement?	a. It handles exceptions	b. It defines the condition to check	c. It is executed if the if condition is true	d. It is executed if the if condition is false	4
How do you define an empty list in Python?	a. list = []	b. list = {}	c. list = ()	d. list = [None]	1
What does the continue statement do in Python?	a. Exits the current loop	b. Skips the next iteration of the loop	c. Ends the entire program	d. Restarts the loop from the beginning	2
Which of the following is a correct way to iterate through a list in Python?	a. for i in list:	b. for i = 0 to len(list):	c. for i from 0 to len(list):	d. for i = 0; i < len(list); i++:	1
Which type of neural network architecture is commonly used for time series forecasting?	a. Recurrent Neural Network (RNN)	b. Convolutional Neural Network (CNN)	c. Multilayer Perceptron (MLP)	d. Radial Basis Function Neural Network (RBFNN).	1
In reinforcement learning what is the purpose of the reward signal?	a. To define the loss function of the network	b. To determine the number of layers in the network	c. To provide feedback to the agent about the quality of its actions	d. To adjust the learning rate during training.	3
Which layer of a neural network is responsible for learning feature representations from the input data?	a. Input layer	b. Hidden layer	c. Output layer	d. Dropout layer.	2
What is the role of the learning rate in training neural networks using gradient descent optimization?	a. To control the speed at which weights are updated during training	b. To adjust the number of epochs during training	c. To determine the size of each layer in the network	d. To introduce randomness into the network.	1
What is the default activation function used in the output layer of the MLPClassifier for binary classification tasks?	a. Sigmoid	b. Tanh (Hyperbolic Tangent)	c. ReLU (Rectified Linear Activation)	d. Linear.	1
How can you specify the learning rate schedule for training a Multilayer Perceptron using the MLPClassifier?	a. By setting the learning_rate parameter	b. By adjusting the momentum parameter	c. By passing a tuple to the learning_rate_init parameter	d. By setting the learning_rate_schedule parameter.	3
What is the purpose of the momentum parameter in the MLPClassifier class?	a. To control the number of hidden layers in the network	b. To adjust the regularization strength	c. To specify the learning rate of the network	d. To stabilize the convergence of the optimization algorithm.	4
Which technique can be used to prevent overfitting in a Multilayer Perceptron model trained using the MLPClassifier?	a. Dropout regularization	b. Increasing the number of hidden layers	c. Removing the bias terms from the network	d. Decreasing the number of iterations for training.	1
What does the Balanced Accuracy metric measure in the context of imbalanced classification evaluation?	a. The proportion of correctly classified instances out of all instances in the dataset	b. The average of sensitivity and specificity	c. The harmonic mean of precision and recall	d. The trade-off between true positive rate and false positive rate.	2
Which evaluation metric is sensitive to the class distribution in imbalanced classification tasks?	a. Accuracy	b. Precision	c. Recall	d. F1-score.	4

Which evaluation metric penalizes false negatives more heavily in imbalanced classification tasks?	a. Accuracy	b. Precision	c. Recall	d. F1-score.	3
Which evaluation metric is commonly used for ranking evaluation in information retrieval tasks?	a. Precision	b. Recall	c. F1-score	d. Mean Average Precision (MAP).	4
What is the primary disadvantage of K-Means clustering?	a. It requires a predetermined number of clusters	b. It is sensitive to outliers	c. It cannot handle non-linearly separable data	d. It has a high computational complexity.	1
Which unsupervised learning algorithm is used for dimensionality reduction and feature extraction in high-dimensional data?	a. t-Distributed Stochastic Neighbor Embedding (t-SNE)	b. K-Means clustering	c. DBSCAN	d. Hierarchical clustering.	1
What is the primary objective of t-Distributed Stochastic Neighbor Embedding (t-SNE)?	a. To minimize the distance between data points and centroids	b. To identify the most influential features in the dataset	c. To transform the data into a lower-dimensional space while preserving the local structure of the data	d. To partition the dataset into a predetermined number of clusters.	3
What is the primary objective of K-Means++ an improvement over the original K-Means algorithm?	a. To reduce the computational complexity	b. To avoid initialization sensitivity	c. To increase the number of iterations	d. To eliminate the need for specifying K.	2
Which technique can be used to determine the optimal number of clusters in K-Means clustering?	a. Elbow method	b. Silhouette method	c. Gap statistic	d. All of the above.	4
What does the "elbow" in the elbow method of determining the optimal number of clusters represent?	a. The number of data points in the dataset	b. The sum of squared distances within clusters	c. The optimal number of clusters	d. The distance between data points.	3
Which method of the KMeans class is used to predict the cluster labels for new data points?	a. fit_predict	b. predict	c. fit_transform	d. transform.	2
What is the primary purpose of the "fit_predict" method in the KMeans class?	a. To fit the model to the data and return the cluster labels	b. To transform the data into a lower-dimensional space	c. To visualize the clusters	d. To initialize the cluster centroids.	1
Which parameter of the KMeans class determines the initialization method for cluster centroids?	a. n_clusters	b. n_init	c. init	d. random_state.	3
Which metric is used to evaluate the stability of clustering by measuring the similarity of clusters across different runs or subsets of the data?	a. Davies-Bouldin Index	b. Silhouette score	c. Jaccard Index	d. Adjusted Rand Index.	3
What does a higher Jaccard Index value indicate about the similarity of clusters across different runs or subsets of the data?	a. Higher similarity of clusters	b. Lower similarity of clusters	c. No similarity of clusters	d. Perfect similarity of clusters.	1
Which metric is used to evaluate the compactness within clusters in clustering evaluation?	a. Davies-Bouldin Index	b. Silhouette score	c. Calinski-Harabasz Index	d. Adjusted Rand Index.	1
How does PCA help in reducing overfitting in machine learning models?	a. By increasing the number of features	b. By reducing the computational complexity	c. By reducing the dimensionality of the dataset	d. By introducing noise into the dataset.	3
What is the computational complexity of PCA?	a. Linear with respect to the number of features	b. Quadratic with respect to the number of features	c. Exponential with respect to the number of features	d. Constant with respect to the number of features.	1

Which of the following statements about PCA is true?	a. PCA always leads to better model performance	b. PCA preserves all information in the dataset	c. PCA may lead to information loss	d. PCA is not suitable for datasets with high dimensionality.	3
Which parameter of the PCA class specifies the number of iterations for computing the principal components?	a. n_components	b. whiten	c. copy	d. svd_solver.	4
What is the default value of the "copy" parameter in the PCA class?	a. True	b. False	c. None	d. Depends on the dataset.	1
Which method of the PCA class is used to fit the model to the data and transform the data into the principal components space in a single step?	a. fit	b. fit_transform	c. transform	d. predict.	2
What is the primary objective of visualizing reduced dimensions using line plots?	a. To visualize the distribution of a single variable	b. To visualize the relationship between two or more variables in reduced dimensions	c. To visualize the correlation matrix of the dataset	d. To introduce noise into the dataset for visualization purposes.	2
How does UMAP differ from PCA and t-SNE in terms of preserving structures in the data during visualization?	a. UMAP preserves only local structures	b. UMAP preserves both local and global structures	c. UMAP preserves only global structures	d. UMAP does not preserve any structures.	2
What is the primary disadvantage of visualizing reduced dimensions?	a. It increases the interpretability of high-dimensional data	b. It may distort or lose some information present in the original high-dimensional data	c. It preserves all dimensions of the original dataset	d. It reduces the computational complexity of models.	2
How are eigenvalues used in PCA?	a. They represent the variance explained by each feature	b. They determine the number of principal components to retain	c. They represent the original features of the dataset	d. They define the directions of the principal components.	2
What is the primary use of the "explained variance ratio" in PCA?	a. To determine the number of principal components to retain	b. To calculate the correlation between features	c. To standardize the features	d. To compute the covariance matrix.	1
1. What is Artificial Intelligence (AI)?	A type of computer hardware	Human-like thinking in machines	. A programming language	Advanced robotics	2
What is labeled data in machine learning?	Data with clear instructions on how to use it	Data that has been preprocessed	Data with associated output labels	Data with encryption labels	3
What is the purpose of a validation set in machine learning?	To train the model	To evaluate the model's performance	To test the model on unseen data	To fine-tune hyperparameters	2
What is the primary use of Speech Recognition in Machine Learning applications?	Translating languages	. Identifying emotions in text	c. Converting spoken language into text	d. Predicting future events	3
How can you install Pandas in Python?	pip install pandas	pip install pandas --upgrade	conda install pandas	All of the above	4
Which function is used to display the first few rows of a DataFrame in Pandas?	display()	show()	head()	top()	3
In Pandas, what does a missing value (NaN) represent?	Zero	An error	A placeholder for a missing or undefined value	Infinity	3

What is a Pandas Series?	A one-dimensional labeled array	. A two-dimensional labeled array	A data visualization tool	A machine learning model	1
How can you create a Pandas Series from a Python list?	pd.Series(list)	pd.Series.from_list(list)	pd.create_series(list)	All of the above	1
What is the purpose of the loc and iloc attributes in a Pandas DataFrame?	To locate specific elements in a DataFrame	To access rows and columns by label (loc) or integer index (iloc)	To filter data based on conditions	Both a and b	2
How can you convert a Pandas DataFrame to a NumPy array?	to_array()	as_numpy()	numpy()	to_numpy()	4
What does the str.contains() function in Pandas do?	Checks if a string contains a specified substring	Converts string values to uppercase	Removes leading and trailing whitespaces	Concatenates two strings	1
What is the purpose of the last() function in Pandas?	Retrieves the last row of a DataFrame	Finds the last occurrence of a value in a Series	Calculates the last element of a numeric column	None of the above	1
How can you read a CSV file into a Pandas DataFrame?	pd.load_csv("file.csv")	pd.read_csv("file.csv")	pd.import_csv("file.csv")	pd.read("file.csv")	2
What is the purpose of the to_csv() method in Pandas?	Reads a CSV file into a DataFrame	Converts a DataFrame to a CSV file	Combines multiple CSV files	Checks for the presence of CSV files	2
How can you filter rows in a Pandas DataFrame based on a condition?	df.filter(condition)	df.query(condition)	df.select(condition)	df.filter_rows(condition)	2
Which method is used to drop missing values from a Pandas DataFrame?	df.drop_null()	df.drop_missing()	df.remove_null()	df.dropna()	4
How can you handle missing values by filling them with a specific value in Pandas?	df.fill_null(value)	df.replace_null(value)	df.fillna(value)	df.null_fill(value)	3
What does the df.groupby() method in Pandas allow you to do?	Group data by rows	Group data by columns	Group data based on a condition	Group data for aggregation based on one or more columns	4
How can you perform aggregation functions like sum, mean, or count on grouped data in Pandas?	df.grouped_agg()	. df.group_by().agg()	df.aggregate(grouped)	df.agg_by_group()	2
What does the df.plot(rot=45) parameter do in Pandas?	Rotates the x-axis labels by 45 degrees	Rotates the y-axis labels by 45 degrees	Rotates both x and y-axis labels by 45 degrees	Sets the rotation angle of the entire plot to 45 degrees	1
How can you set the title of a line plot in Pandas?	df.plot(title='My Plot')	df.plot.line(title='My Plot')	df.plot().set_title('My Plot')	df.plot().title('My Plot')	1

What is the purpose of the df.plot(legend=True) parameter in Pandas?	a. Adds legends to the plot	b. Removes legends from the plot	c. Adjusts the position of the legends	d. Changes the color of the legends	1
. How can you create a vertical bar plot in Pandas?	a. df.plot.bar()	b. df.plot.line()	c. df.plot.scatter()	d. df.plot.pie()	1
What does the df.plot.barh() method do in Pandas?	a. Creates a horizontal bar plot	b. Creates a vertical bar plot	c. Creates a scatter plot	d. Creates a line plot	1
How can you specify the column to be used as the bar labels in Pandas?	a. df.plot.bar(x='column_labels')	b. df.plot.bar(y='column_labels')	c. df.plot.bar().set_x('column_labels')	d. df.plot.bar().set_labels('column_labels')	1
What is the purpose of the df.plot.bar(stacked=True) parameter in Pandas?	a. Creates separate bars for each category	b. Stacks the bars for each category	c. Adds legends to the plot	d. Sets the color of the bars	2
How can you customize the color of the bars in a Pandas bar plot?	a. df.plot.bar(color='red')	b. df.plot.bar().set_color('red')	c. df.plot.bar().color('red')	d. df.plot.bar().set_colors(['red', 'blue', 'green'])	1
What does the df.plot.bar(width=0.8) parameter do in Pandas?	a. Adjusts the width of the entire plot	b. Adjusts the width of the bars	c. Sets the space between bars	d. Sets the transparency of the bars	2
How can you add grid lines to a Pandas bar plot?	a. df.plot.bar(grid=True)	b. df.plot.bar().grid()	c. df.plot.bar().set_grid(True)	d. df.plot.bar().grid(True)	1
What does the df.plot.bar(logy=True) parameter do in Pandas?	a. Sets the y-axis to logarithmic scale	b. Sets the x-axis to logarithmic scale	c. Sets both x and y axes to logarithmic scale	d. Converts the data to logarithmic format	1
How can you set the title of a bar plot in Pandas?	a. df.plot.bar(title='My Plot')	b. df.plot(title='My Plot')	c. df.plot().set_title('My Plot')	d. df.plot.bar().set_title('My Plot')	1
What does the df.plot.bar(rot=45) parameter do in Pandas?	a. Rotates the x-axis labels by 45 degrees	b. Rotates the y-axis labels by 45 degrees	c. Rotates both x and y-axis labels by 45 degrees	d. Sets the rotation angle of the entire plot to 45 degrees	1
How can you create a scatter plot in Pandas?	a. df.plot.scatter(x='column1', y='column2')	b. df.plot.line(x='column1', y='column2')	c. df.plot.bar(x='column1', y='column2')	d. df.plot.pie(x='column1', y='column2')	1
What is the purpose of the df.plot.scatter(c='column3') parameter in Pandas?	a. Sets the color of the scatter points based on values in 'column3'	b. Sets the size of the scatter points based on values in 'column3'	c. Connects the scatter points with lines based on values in 'column3'	d. Sets the transparency of the scatter points based on values in 'column3'	1
What is the output of print(3 * 'abc')?	a. abc	b. abcabc	c. 333	d. 9	2

What does the len() function do in Python?	a. Returns the length of a list or string	b. Performs mathematical operations	c. Defines a new variable	d. Executes a loop	1
What is the purpose of the input() function in Python?	a. Display output on the console	b. Read user input from the console	c. Perform mathematical calculations	d. Create a graphical user interface	2
Which of the following is a correct way to check if two variables, a and b, are equal in value?	a. a == b	b. a = b	c. a != b	d. a < b	1
What is the output of print("Hello" + "World")?	a. Hello World l d	b. Hello + World	c. HelloWorld	d. H e l l o W o r	3
Which of the following is a subfield of AI focused on language understanding?	a. Robotics	b. Natural Language Processing (NLP)	c. Computer Vision	d. Expert Systems \	2
What is overfitting in machine learning?	a. Model fits the training data too closely	b. Model performs well on the test set	c. Model generalizes well to new data	d. Model underperforms on both training and test sets	1
Which algorithm is commonly used for classification tasks in machine learning?	a. Regression	b. Decision Trees	c. Clustering	d. Association Rule Mining	2
Which industry commonly utilizes Time Series Forecasting using Machine Learning?	a. Entertainment	b. Transportation	c. Weather forecasting	d. Real estate	3
Which method is used to check for missing values in a Pandas DataFrame?	a. find_missing()	b. check_nan()	c. isnull()	d. missing_values()	3
What is the default axis value when applying a function like sum() or mean() to a Pandas DataFrame?	a. 0	b. 1	c. Both 0 and 1	d. None	1
How can you create a Pandas DataFrame from a Python dictionary?	a. pd.create_dataframe(dictionary)	b. pd.DataFrame.from_dict(dictionary)	c. pd.DataFrame(dictionary)	d. All of the above	4
What is the default behavior of the read_csv() function in Pandas when loading a CSV file?	a. Loads the entire file into memory	b. Skips the first row	c. Assumes the first row is the header	d. Discards rows with missing values	3
How can you drop a column in a Pandas DataFrame?	a. drop(column_name)	b. remove_column(column_name)	c. drop(columns=column_name)	d. delete_column(column_name)	3
Which method is used to filter rows in a Pandas DataFrame based on a condition?	a. filter()	b. select()	c. where()	d. query()	3

How can you fill missing values in a Pandas DataFrame?	a. fill_null()	b. impute()	c. fillna()	d. replace_null()	3
What does the to_datetime() function in Pandas do?	a. Converts a Series to a DataFrame	b. Converts a column of strings to datetime objects	c. Extracts the time component from a datetime column	d. Checks if a value is a valid datetime	2
Which method is used to read an Excel file into a Pandas DataFrame?	a. pd.load_excel("file.xlsx")	b. pd.read_excel("file.xlsx")	c. pd.import_excel("file.xlsx")	d. pd.read("file.xlsx")	2
How can you write a Pandas DataFrame to an Excel file?	a. df.save_excel("output.xlsx")	b. df.to_excel("output.xlsx")	c. df.write_excel("output.xlsx")	d. df.export_excel("output.xlsx")	2
Which function is used to read data from an SQL database into a Pandas DataFrame?	a. pd.sql("SELECT * FROM table")	b. pd.read_sql_query("SELECT * FROM table", connection)	c. pd.load_sql("SELECT * FROM table")	d. pd.import_sql("SELECT * FROM table")	2
Which method is used to pivot a Pandas DataFrame?	a. df.pivot_table()	b. df.pivote()	c. df.pivot()	d. df.transpose()	3
How can you visualize data using histograms in Pandas?	a. df.plot.scatter()	b. df.plot.hist()	c. df.plot.line()	d. df.plot.box()	2
What is Matplotlib?	a. A Pandas module for data manipulation	b. A machine learning library	c. A data visualization library	d. A statistical analysis tool	3
How can you install Matplotlib?	a. pip install matplotlib	b. conda install matplotlib	c. Both a and b	d. None of the above	3
Which of the following is true about Matplotlib's pyplot module?	a. It is the main plotting library in Matplotlib	b. It is used only for 3D plotting	c. It is deprecated in the latest version of Matplotlib	d. It is not compatible with Pandas	1
How can you customize the marker style in a Pandas scatter plot?	a. df.plot.scatter(marker='circle')	b. df.plot.scatter().set_marker('circle')	c. df.plot.scatter().marker('circle')	d. df.plot.scatter(marker='o')	4
What does the df.plot.scatter(s='column4') parameter do in Pandas?	a. Sets the color of the scatter points based on values in 'column4'	b. Sets the size of the scatter points based on values in 'column4'	c. Connects the scatter points with lines based on values in 'column4'	d. Sets the transparency of the scatter points based on values in 'column4'	2
How can you add grid lines to a Pandas scatter plot?	a. df.plot.scatter(grid=True)	b. df.plot.scatter().grid()	c. df.plot.scatter().set_grid(True)	d. df.plot.scatter().grid(True)	1
What does the df.plot.scatter(loglog=True) parameter do in Pandas?	a. Sets both x and y axes to logarithmic scale	b. Sets the x-axis to logarithmic scale	c. Sets the y-axis to logarithmic scale	d. Converts the data to logarithmic format	1
How can you customize the color of the scatter points in a Pandas scatter plot?	a. df.plot.scatter(color='red')	b. df.plot.scatter().set_color('red')	c. df.plot.scatter().color('red')	d. df.plot.scatter(c='red')	4

What is the purpose of the df.plot.scatter(alpha=0.5) parameter in Pandas?	a. Sets the transparency of the scatter points to 0.5	b. Sets the size of the scatter points to 0.5	c. Connects the scatter points with lines of 0.5 thickness	d. Sets the color of the scatter points to 0.5	1
How can you set the title of a scatter plot in Pandas?	a. df.plot.scatter(title='My Plot')	b. df.plot(title='My Plot')	c. df.plot().set_title('My Plot')	d. df.plot.scatter().set_title('My Plot')	1
What does the df.plot.scatter(rot=45) parameter do in Pandas?	a. Rotates the x-axis labels by 45 degrees	b. Rotates the y-axis labels by 45 degrees	c. Rotates both x and y-axis labels by 45 degrees	d. Sets the rotation angle of the entire plot to 45 degrees	1
How can you create a histogram in Pandas?	a. df.plot.scatter()	b. df.plot.line()	c. df.plot.bar()	d. df.plot.hist()	4
What does the df.plot.hist(bins=20) parameter do in Pandas?	a. Sets the color of the histogram bars to 20	b. Sets the transparency of the histogram bars to 20	c. Sets the number of bins in the histogram to 20	d. Sets the width of the histogram bars to 20	3
What is the main idea behind Principal Component Analysis (PCA) in unsupervised learning?	a. Dividing data into clusters	b. Reducing dimensionality by transforming features into a new set of uncorrelated variables	c. Classifying data points into categories	d. Predicting continuous outcomes	2
Which of the following is an application of unsupervised learning?	a. Image recognition	b. Predicting stock prices	c. Customer segmentation	d. Spam email detection	3
What does the term "clustering" refer to in the context of unsupervised learning?	a. Assigning a label to each data point	b. Dividing data points into groups with similar characteristics	c. Predicting a continuous numerical output	d. Reducing dimensionality of data	2
Which unsupervised learning technique is used to fill in missing values in a dataset?	a. Hierarchical clustering	b. Imputation	c. Dimensionality reduction	d. Anomaly detection	2
What is the role of the "elbow method" in K-Means clustering?	a. Identifying the optimal number of clusters by selecting the "elbow" point in the plot of the sum of squared distances	b. Assigning labels to data points	c. Determining outliers in the dataset	d. Reducing the dimensionality of the data	1
What is the primary limitation of hierarchical clustering in unsupervised learning?	a. Sensitivity to the initial choice of cluster centers	b. Difficulty in handling large datasets	c. Inability to form non-convex clusters	d. Lack of flexibility in the number of clusters	2
Which unsupervised learning technique is suitable for detecting anomalies or outliers in a dataset?	a. Principal Component Analysis (PCA)	b. Hierarchical clustering	c. K-Means clustering	d. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	4
What is the primary goal of t-Distributed Stochastic Neighbor Embedding (t-SNE) in unsupervised learning?	a. Dimensionality reduction	b. Clustering data points	c. Anomaly detection	d. Imputing missing values	1

Which unsupervised learning algorithm is used for density-based clustering?	a. K-Means	b. Hierarchical Clustering	c. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	d. PCA (Principal Component Analysis)	3
What is the primary objective of using t-Distributed Stochastic Neighbor Embedding (t-SNE) in unsupervised learning?	a. Classify data points into predefined categories	b. Dimensionality reduction for visualization	c. Predict continuous numerical output	d. Impute missing values in a dataset	2
What does the Turing Test assess in AI?	a. Memory capacity	b. Cognitive abilities	c. Conversational abilities	d. Visual perception	3
What is the primary application of Natural Language Processing (NLP) in Machine Learning?	a. Image recognition	b. Speech synthesis	c. Text analysis and understanding	d. Predictive modeling	3
In which industry is Fraud Detection commonly applied using Machine Learning?	a. Healthcare	b. Finance	c. Retail	d. Manufacturing	2
What is Pandas?	a. A programming language	b. A data manipulation library	c. An operating system	d. A machine learning algorithm	2
What is the purpose of the groupby() function in Pandas?	a. Grouping rows based on a specified column	b. Sorting rows in alphabetical order	c. Combining multiple DataFrames	d. Reshaping the DataFrame	1
How can you concatenate two Pandas DataFrames vertically?	a. concat(df1, df2, axis=0)	b. merge(df1, df2, how='inner')	c. join(df1, df2, on='key')	d. append(df1, df2)	1
What is the purpose of the pivot_table() function in Pandas?	a. Creating a new DataFrame from existing data	b. Reshaping and summarizing data based on specified criteria	c. Sorting data in ascending or descending order	d. Combining multiple DataFrames	2
How can you handle missing values in a Pandas DataFrame?	a. Drop rows with missing values using dropna()	b. Replace missing values with a specified value using fillna()	c. Interpolate missing values using interpolate()	d. All of the above	4
What is the purpose of the merge() function in Pandas?	a. Combining DataFrames based on common columns or indices	b. Sorting data in ascending or descending order	c. Reshaping the DataFrame	d. Filtering rows based on a condition	1
What does the stack() function do in Pandas?	a. Unstacks a DataFrame	b. Converts a Series to a DataFrame	c. Stacks a DataFrame from a multi-level index	d. None of the above	3
How can you drop duplicate rows in a Pandas DataFrame?	a. drop_duplicates()	b. remove_duplicates()	c. delete_duplicates()	d. deduplicate()	1
What is the purpose of the sample() function in Pandas?	a. Randomly samples rows from a DataFrame	b. Extracts a random subset of columns	c. Sorts rows in random order	d. None of the above	1

How can you connect Pandas to an SQL database?	a. pd.connect_sql()	b. pd.create_connection()	c. pd.sql.connect()	d. pd.read_sql()	3
What does the pd.read_sql_query() function do?	a. Reads an entire SQL database	b. Reads the result of a SQL query into a Pandas DataFrame	c. Executes a SQL query without returning any results	d. Creates a new table in the SQL database	2
How can you execute a parameterized SQL query with Pandas?	a. pd.execute_sql_query(query, params)	b. pd.read_sql_query(query, connection, params)	c. pd.execute_query_with_params(query, connection, params)	d. pd.query_sql(query, params)	2
How can you create a simple line plot using Matplotlib in Pandas?	a. df.plot.scatter(x='column1', y='column2')	b. df.plot.line(x='column1', y='column2')	c. df.plot.bar(x='column1', y='column2')	d. df.plot.pie(x='column1', y='column2')	2
What does the plt.show() function do in Matplotlib?	a. Displays the plot	b. Saves the plot to a file	c. Closes the plot	d. Exports the plot to a different format	1
How can you customize the color of a plot in Matplotlib?	a. df.plot(color='red')	b. df.plot.line(color='red')	c. df.plot(color='red', style='line')	d. df.plot.line(style='red')	2
What is the purpose of the plt.title() function in Matplotlib?	a. Adds a title to the plot	b. Sets the color of the plot	c. Changes the style of the plot	d. Adjusts the size of the plot	1
How can you add labels to the x-axis and y-axis in a Matplotlib plot?	a. plt.xlabel() and plt.ylabel()	b. plt.x_axis() and plt.y_axis()	c. plt.axis_x() and plt.axis_y()	d. plt.x_label() and plt.y_label()	1
How can you add grid lines to a Pandas histogram?)	a. df.plot.hist(grid=True)	b. df.plot.hist().grid()	c. df.plot.hist().set_grid(True)	d. df.plot.hist().grid(True)	1
What is the purpose of the df.plot.hist(orientation='horizontal') parameter in Pandas?	a. Rotates the x-axis labels by 45 degrees	b. Rotates the y-axis labels by 45 degrees	c. Creates a horizontal bar plot instead of a vertical histogram	d. Adjusts the size of the histogram	3
How can you customize the color of the histogram bars in Pandas?	a. df.plot.hist(color='red')	b. df.plot.hist().set_color('red')	c. df.plot.hist().color('red')	d. df.plot.hist(facecolor='red')	4
What does the df.plot.hist(logy=True) parameter do in Pandas?	a. Sets the y-axis to logarithmic scale	b. Sets the x-axis to logarithmic scale	c. Sets both x and y axes to logarithmic scale	d. Converts the data to logarithmic format	1
How can you set the title of a histogram plot in Pandas?	a. df.plot.hist(title='My Plot')	b. df.plot(title='My Plot')	c. df.plot().set_title('My Plot')	d. df.plot.hist().set_title('My Plot')	1
What does the df.plot.hist(alpha=0.5) parameter do in Pandas?	a. Sets the transparency of the histogram bars to 0.5	b. Sets the width of the histogram bars to 0.5	c. Sets the color of the histogram bars to 0.5	d. Sets the size of the histogram bars to 0.5	1
How can you customize the edge color of the histogram bars in Pandas?	a. df.plot.hist(edgecolor='blue')	b. df.plot.hist().set_edgecolor('blue')	c. df.plot.hist().edgecolor('blue')	d. df.plot.hist(color='blue', edgecolor='black')	1

What is the purpose of the df.plot.hist(cumulative=True) parameter in Pandas?	a. Creates a stacked histogram	b. Sets the histogram to a cumulative distribution	c. Creates a 3D histogram	d. Adds legends to the plot	2
How can you create a box plot in Pandas?	a. df.plot.scatter()	b. df.plot.line()	c. df.plot.bar()	d. df.plot.box()	4
What is the purpose of the df.plot.box(vert=False) parameter in Pandas?	a. Creates a horizontal box plot instead of a vertical box plot	b. Sets the color of the box plot to False	c. Vertically stacks multiple box plots	d. Adds a vertical line to the box plot	1
Which unsupervised learning method is suitable for discovering hidden patterns in data?	a. Clustering	b. Imputation	c. Regression	d. Dimensionality reduction	1
What is the key idea behind the Expectation-Maximization (EM) algorithm in unsupervised learning?	a. Minimizing prediction errors	b. Maximizing likelihood by iteratively estimating hidden variables	c. Reducing dimensionality of data	d. Assigning labels to data points	2
Which of the following is an application of unsupervised learning in natural language processing (NLP)?	a. Sentiment analysis	b. Named entity recognition	c. Part-of-speech tagging	d. Document classification	2
In unsupervised learning, what is the primary challenge associated with hierarchical clustering?	a. Sensitivity to the initial choice of cluster centers	b. Determining the optimal number of clusters	c. Difficulty in handling large datasets	d. Inability to form non-convex clusters	2
What is the role of imputation in unsupervised learning?	a. Dividing data points into clusters	b. Filling in missing values in a dataset	c. Reducing dimensionality of data	d. Detecting anomalies in the dataset	2
Which unsupervised learning technique is used for reducing the dimensionality of data while preserving the variance?	a. Principal Component Analysis (PCA)	b. K-Means clustering	c. Hierarchical Clustering	d. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	1
What is the primary goal of hierarchical clustering in unsupervised learning?	a. Assigning labels to data points	b. Minimizing the sum of squared distances within clusters	c. Creating a hierarchy of clusters	d. Predicting continuous numerical output	3
Which unsupervised learning technique is commonly used for anomaly detection in network security?	a. K-Means clustering	b. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	c. Isolation Forest	d. Principal Component Analysis (PCA)	3
What is the primary purpose of K-Means clustering in unsupervised learning?	a. Density-based clustering	b. Hierarchical clustering	c. Partitioning data points into k clusters	d. Dimensionality reduction	3
What does the acronym DBSCAN stand for in unsupervised learning?	a. Distance-Based Spatial Clustering of Applications with Noise	b. Density-Based Spatial Clustering of Applications with Noise	c. Dimension-Based Spatial Clustering of Applications with Noise	d. Discrete-Based Spatial Clustering of Applications with Noise	2

Which programming language is widely used in AI development?	a. Java	b. C++	c. Python	d. HTML	3
Which Machine Learning application is used for recommending products or content based on user preferences?	a. Sentiment analysis	b. Collaborative filtering	c. Object detection	d. Regression analysis	2
What is Computer Vision primarily focused on in Machine Learning applications?	a. Speech recognition	b. Image and video analysis	c. Text translation	d. Predictive modeling	2
Which of the following data structures is not provided by Pandas?	a. Series	b. DataFrame	c. Array	d. Panel	3
How can you rename columns in a Pandas DataFrame?	a. rename_columns()	b. columns.rename()	c. rename()	d. change_columns()	3
What is the purpose of the to_csv() function in Pandas?	a. Saves a DataFrame to a CSV file	b. Converts a CSV file to a DataFrame	c. Combines multiple CSV files into a single file	d. Reads data from a CSV file	3
Which method is used to calculate summary statistics of a Pandas DataFrame?	a. stats()	b. summary()	c. describe()	d. statistics()	3
What does the astype() function do in Pandas?	a. Adds a new column to the DataFrame	b. Changes the data type of a column	c. Drops a column from the DataFrame	d. Concatenates two DataFrames	2
How can you sort a Pandas DataFrame based on a specific column?	a. sort_by(column)	b. order_by(column)	c. sort_values(by=column)	d. arrange(column)	3
What is the purpose of the cut() function in Pandas?	a. Cuts a DataFrame into smaller pieces	b. Divides data into intervals and assigns labels	c. Removes duplicate rows from a DataFrame	d. Merges two DataFrames based on a common column	2
What is the purpose of the pd.to_sql() function in Pandas?	a. Converts a Pandas DataFrame to an SQL database	b. Writes data from a Pandas DataFrame to an existing SQL table	c. Executes a SQL query	d. Connects to an SQL database	2
How can you specify the index column when writing a Pandas DataFrame to an SQL table?	a. pd.to_sql("table", connection, index_column="index")	b. pd.to_sql("table", connection, index_col="index")	c. pd.write_sql("table", connection, index_column="index")	d. pd.write_sql("table", connection, index_col="index")	2
What is the purpose of the if_exists parameter in pd.to_sql()?	a. Specifies whether to create a new table or append data to an existing one	b. Specifies the SQL database connection	c. Sets the index column of the SQL table	d. Determines the SQL query to be executed	1
How can you execute a raw SQL query with Pandas?	a. pd.execute_sql(raw_query)	b. pd.read_sql(raw_query, connection)	c. pd.sql.execute(raw_query)	d. pd.raw_sql(raw_query)	2

What is the purpose of the chunksize parameter in pd.read_sql_query()?	a. Specifies the number of rows to read at a time from the SQL database	b. Sets the maximum number of chunks to be retrieved	c. Filters the query results based on chunksize	d. Determines the size of each chunk in the SQL database	1
What does the plt.legend() function do in Matplotlib?	a. Adds legends to the plot	b. Sets the legend's position	c. Removes legends from the plot	d. Adjusts the size of the legend	1
How can you save a Matplotlib plot to a file?	a. plt.save('plot.png')	b. plt.export('plot.png')	c. plt.savefig('plot.png')	d. plt.saveplot('plot.png')	3
What is Seaborn?	a. A statistical analysis library	b. A machine learning framework	c. A data manipulation tool	d. A data visualization library	4
How can you install Seaborn?	a. pip install seaborn	b. conda install seaborn	c. pip install pandas seaborn	d. conda install pandas seaborn	1
Which function in Seaborn is used to create a scatter plot?	a. seaborn.scatterplot()	b. seaborn.scatter()	c. seaborn.plot.scatter()	d. seaborn.scatter_plot()	1
How can you add grid lines to a Pandas box plot?	a. df.plot.box(grid=True)	b. df.plot.box().grid()	c. df.plot.box().set_grid(True)	d. df.plot.box().grid(True)	1
What is the purpose of the df.plot.box(showfliers=False) parameter in Pandas?	a. Removes outliers from the box plot	b. Displays only the outliers in the box plot	c. Sets the color of the outliers to False	d. Adjusts the size of the box plot	1
How can you customize the color of the box plot elements in Pandas?	a. df.plot.box(color='red')	b. df.plot.box().set_color('red')	c. df.plot.box().color('red')	d. df.plot.box(patch_artist=True, boxprops=dict(facecolor='red'))	4
What does the df.plot.box(sym='b+') parameter do in Pandas?	a. Sets the color of the box plot to 'b+'	b. Sets the marker style for the outliers to 'b+'	c. Sets the transparency of the box plot to 'b+'	d. Adds a legend to the box plot with the label 'b+'	2
How can you set the title of a box plot in Pandas?	a. df.plot.box(title='My Plot')	b. df.plot(title='My Plot')	c. df.plot().set_title('My Plot')	d. df.plot.box().set_title('My Plot')	1
What does the df.plot.box(logy=True) parameter do in Pandas?	a. Sets the y-axis to logarithmic scale	b. Sets the x-axis to logarithmic scale	c. Sets both x and y axes to logarithmic scale	d. Converts the data to logarithmic format	1
How can you customize the width of the box plot in Pandas?	a. df.plot.box(width=0.5)	b. df.plot.box().set_width(0.5)	c. df.plot.box().width(0.5)	d. df.plot.box(boxprops=dict(linewidth=0.5))	4

What does the df.plot.box(patch_artist=True) parameter do in Pandas?	a. Sets the color of the box plot to True	b. Adds patches to the box plot	c. Adjusts the size of the box plot	d. Adds legends to the box plot	2
What is the main goal of supervised learning?	a. Clustering data points	b. Predicting an output variable based on input variables	c. Reducing dimensionality of data	d. Identifying outliers in the dataset	2
In supervised learning, what is the role of the input features?	a. They are the outputs predicted by the model	b. They are the independent variables used to predict the output	c. They are not used in supervised learning	d. They are used only in unsupervised learning	2
In unsupervised learning, what is the role of the "silhouette score"?	a. Evaluating the performance of a regression model	b. Measuring the compactness and separation of clusters	c. Imputing missing values in a dataset	d. Detecting outliers in the dataset	2
Which unsupervised learning technique is used for reducing the dimensionality of data by projecting it onto a lower-dimensional space?	a. K-Means clustering	b. Principal Component Analysis (PCA)	c. Hierarchical clustering	d. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	2
What is the primary limitation of K-Means clustering in unsupervised learning?	a. Sensitivity to outliers	b. Difficulty in handling large datasets	c. Inability to form non-convex clusters	d. Requirement of specifying the number of clusters in advance	1
Which unsupervised learning method is suitable for detecting patterns or structures in high-dimensional data?	a. K-Means clustering	b. Hierarchical clustering	c. t-Distributed Stochastic Neighbor Embedding (t-SNE)	d. Imputation	3
What is the primary challenge associated with the curse of dimensionality in unsupervised learning?	a. Difficulty in handling large datasets	b. Sensitivity to outliers	c. Increased computational complexity	d. Requirement of labeled training data	3
Which unsupervised learning technique is used for imputing missing values based on the relationships among variables?	a. Hierarchical clustering	b. K-Means clustering	c. Expectation-Maximization (EM) algorithm	d. Isolation Forest	3
In unsupervised learning, what is the primary goal of hierarchical clustering?	a. Assigning labels to data points	b. Minimizing the sum of squared distances within clusters	c. Creating a hierarchy of clusters	d. Predicting continuous numerical output	3
Which unsupervised learning technique is suitable for identifying clusters of varying shapes and sizes?	a. K-Means clustering	b. Hierarchical clustering	c. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	d. Principal Component Analysis (PCA)	3
How do you write a multiline string in Python?	a. string = 'This is a multiline string'	b. string = "This is a multiline string"	c. string = '''This is a multiline string'''	d. string = """This is a multiline string""	4

What is the purpose of the range() function in Python?	a. Generate a sequence of numbers	b. Find the square root of a number	c. Calculate the factorial of a number	d. Create a random list of elements	1
What is the primary goal of AI?	a. Replicating human emotions	b. Mimicking human behavior	c. Solving complex problems intelligently	d. Creating autonomous machines	3
In which industry is predictive maintenance using Machine Learning commonly employed?	a. Agriculture	b. Healthcare	c. Manufacturing	d. Education	3
What is the primary purpose of Recommender Systems in Machine Learning?	a. Identifying outliers in data	b. Predicting future stock prices	c. Filtering and suggesting items based on user preferences	d. Classifying images	3
What is the primary purpose of Pandas in data analysis?	a. Image processing	b. Statistical modeling	c. Data manipulation and analysis	d. Speech recognition	3
How can you reset the index of a Pandas DataFrame?	a. reset_index()	b. reindex()	c. set_index()	d. index_reset()	1
What is the purpose of the agg() function in Pandas?	a. Aggregates data using a custom function	b. Adds a new column to the DataFrame	c. Filters rows based on a condition	d. Aggregates data using predefined statistical functions	1
How can you check for duplicates in a Pandas DataFrame?	a. find_duplicates()	b. duplicated()	c. check_duplicates()	d. has_duplicates()	2
What is the purpose of the cumsum() function in Pandas?	a. Computes the cumulative sum of a column	b. Counts the occurrences of each unique value	c. Calculates the cumulative product of a column	d. Finds the cumulative maximum value	1
How can you apply a function to each element of a Pandas DataFrame?	a. map()	b. apply()	c. transform()	d. All of the above	4
What does the nunique() function in Pandas do?	a. Counts the number of missing values	b. Counts the number of unique values in a column	c. Calculates the cumulative number of occurrences	d. Finds the total number of elements	2
How can you handle SQL NULL values when writing a Pandas DataFrame to an SQL table?	a. Use the fillna() method to replace NULL values	b. Set the na_values parameter in pd.to_sql()	c. Specify the if_exists parameter as replace	d. All of the above	2

Which method is used to close an SQL connection in Pandas?	a. pd.close_connection()	b. pd.disconnect_sql()	c. connection.close()	d. pd.sql.close()	3
What does the numpy.mean() function calculate?	a. Median	b. Mean	c. Mode	d. Standard Deviation	3
In Python, which library is commonly used to perform statistical calculations?	a. Matplotlib	b. Pandas	c. NumPy	d. SciPy	3
What is the purpose of the numpy.median() function?	a. Finds the average	b. Finds the middle value in a dataset	c. Identifies the most common value	d. Measures the spread of data	2
Which function in Seaborn is used to create a scatter plot?	a. seaborn.scatterplot()	b. seaborn.scatter()	c. seaborn.plot.scatter()	d. seaborn.scatter_plot()	1
What is the purpose of the seaborn.set() function in Seaborn?	a. Sets the style of the plots	b. Sets the color palette	c. Sets the size of the plots	d. Sets the title of the plots	1
5. How can you create a boxplot in Seaborn?	a. seaborn.box()	b. seaborn.plot.box()	c. seaborn.boxplot()	d. seaborn.create_boxplot()	3
6. What does the hue parameter do in Seaborn?	a. Sets the background color of the plot	b. Adds a title to the plot	c. Specifies the variable to map to color	d. Adjusts the size of the plot	3
7. How can you create a heatmap in Seaborn?	a. seaborn.plot.heatmap()	b. seaborn.heatmap()	c. seaborn.create_heatmap()	d. seaborn.heat_map()	2
Which term is used for the output variable in a supervised learning problem?	a. Target variable	b. Feature variable	c. Independent variable	d. Covariate	1
What is the primary task in a regression problem in supervised learning?	a. Classifying data points into categories	b. Predicting a continuous numerical output	c. Identifying outliers in the dataset	d. Clustering similar data points	2
Which algorithm is commonly used for binary classification problems in supervised learning?	a. K-Means	b. Decision Trees	c. K-Nearest Neighbors	d. Support Vector Machines	4
What is the purpose of the training phase in supervised learning?	a. To evaluate the model's performance	b. To predict new, unseen data points	c. To adjust the model parameters based on labeled training data	d. To visualize the data distribution	3

Which metric is commonly used to evaluate the performance of a classification model in supervised learning?	a. Mean Squared Error (MSE)	b. F1 Score	c. Silhouette Score	d. Explained Variance	2
In a supervised learning problem, what does the test dataset typically represent?	a. The labeled data used for training	b. Unseen data points used to evaluate the model's performance	c. A subset of the training data used for cross-validation	d. The output variables predicted by the model	2
Which of the following is an example of a supervised learning task?	a. Customer segmentation	b. Anomaly detection	c. Image classification	d. Market basket analysis	3
What is the purpose of cross-validation in supervised learning?	a. To split the dataset into training and test sets	b. To validate the model's performance on unseen data	c. To adjust hyperparameters based on training data	d. To identify outliers in the dataset	2
1. What is the primary goal of unsupervised learning?	a. Predicting an output variable based on input features	b. Reducing dimensionality of data	c. Identifying outliers in the dataset	d. Clustering data points without labeled output	4
In unsupervised learning, what is the role of the input features?	a. They are the outputs predicted by the model	b. They are the independent variables used to predict the output	c. They are not used in unsupervised learning	d. They are used only in supervised learning	2
What is Python?	a. A high-level programming language	b. A type of snake	c. A character in a video game	d. A piece of hardware	1
How do you denote a single-line comment in Python?	a. //	b. #	c. /* */	d. ---	2
Which of the following is the correct way to declare a variable in Python?	a. variable x;	b. x := 5	c. x = 5	d. int x = 5	3
. What is the correct way to create a function in Python?	a. def function_name():	b. function function_name():	c. function_name def():	d. function_name() {	1
How do you open a file named "example.txt" in Python for writing?	a. file = open("example.txt", "r")	b. file = open("example.txt", "w")	c. file = open("example.txt", "a")	d. file = open("example.txt", "x")	2
What is the purpose of the break statement in Python?	a. Ends the current loop or iteration	b. Skips the next iteration of the loop	c. Exits the entire program	d. Continues to the next loop	1
What is the purpose of the else clause in a Python if-else statement?	a. It handles exceptions	b. It defines the condition to check	c. It is executed if the if condition is true	d. It is executed if the if condition is false	4
How do you define an empty list in Python?	a. list = []	b. list = {}	c. list = ()	d. list = [None]	1
What does the continue statement do in Python?	a. Exits the current loop	b. Skips the next iteration of the loop	c. Ends the entire program	d. Restarts the loop from the beginning	2

Which of the following is a correct way to iterate through a list in Python?	a. for i in list:	b. for i = 0 to len(list):	c. for i from 0 to len(list):	d. for i = 0; i < len(list); i++:	1
What is Machine Learning (ML)?	a. A type of computer virus	b. A branch of AI focused on systems learning from data	c. Human learning through computers	d. Machine programming language	2
Which Machine Learning application involves categorizing data into predefined classes?	a. Regression analysis	b. Clustering	c. Classification	d. Dimensionality reduction	3
In which field is Machine Learning often used for medical diagnosis and prognosis?	a. Astrophysics	b. Geology	c. Medicine	d. Environmental Science	3
Which Python library is the foundation for Pandas?	a. NumPy	b. Matplotlib	c. Scikit-learn	d. TensorFlow	1
What is the purpose of the shape attribute in a Pandas DataFrame?	a. Returns the number of elements in the DataFrame	b. Returns the number of dimensions (axes) of the DataFrame	c. Returns the dimensions (rows, columns) of the DataFrame	d. Returns the data types of the DataFrame	3
How can you extract a specific column from a Pandas DataFrame?	a. get_column()	b. extract_column()	c. column()	d. []	4
What does the pivot() function do in Pandas?	a. Reshapes a DataFrame based on specified criteria	b. Combines multiple DataFrames horizontally	c. Sorts a DataFrame based on a column	d. Filters rows based on a condition	1
How can you calculate the correlation matrix in Pandas?	a. correlation_matrix()	b. calculate_correlation()	c. corr()	d. correlate()	3
What is the purpose of the pivot_table() function in Pandas?	a. Reshapes a DataFrame based on specified criteria	b. Combines multiple DataFrames horizontally	c. Calculates summary statistics by grouping data	d. Sorts a DataFrame based on a column	3
How can you check if a specific value exists in a Pandas DataFrame?	a. contains()	b. exists()	c. is_present()	d. isin()	4
Which function is used to calculate the mode in Python?	a. numpy.mean()	b. numpy.median()	c. statistics.mode()	d. numpy.mode()	3
What does the term "variance" represent in statistics?	a. Measure of central tendency	b. Measure of data spread or dispersion	c. Measure of skewness	d. Measure of kurtosis	2
Which function is used to calculate the variance in Python?	a. numpy.var()	b. numpy.mean()	c. statistics.variance()	d. numpy.std()	1

What does the numpy.std() function calculate?	a. Median	b. Mean	c. Standard deviation	d. Mode	1
Which of the following is a measure of the spread of a dataset?	a. Mean	b. Median	c. Mode	d. Range	4
How can you calculate the 75th percentile of a dataset in Python?	a. numpy.percentile(data, 75)	b. numpy.median(data, 0.75)	c. statistics.percentile(data, 75)	d. numpy.percentile(data, 0.75)	1
What is the purpose of the numpy.histogram() function?	a. Calculates the mean of a datase	b. Computes the frequency distribution of a dataset	c. Measures the skewness of a dataset	d. Calculates the mode of a dataset	2
What does the df.head() method in Pandas do?	a. Displays the last 5 rows of the DataFrame	b. Displays the first 5 rows of the DataFrame	c. Displays a summary of the DataFrame	d. Displays the shape of the DataFrame	2
How can you check the data types and non-null counts of each column in a Pandas DataFrame?	a. df.describe()	b. df.info()	c. df.types()	d. df.data_types()	2
What is the purpose of the df.describe() method in Pandas?	a. Displays basic statistics of the DataFrame	b. Describes the structure of the DataFrame	c. Counts the unique values in each column	d. Displays the first 5 rows of the DataFrame	1
What is the purpose of the seaborn.pairplot() function?	a. Creates a pair plot between two variables	b. Creates a scatter plot matrix	c. Plots the distribution of a single variable	d. Plots a line plot between two variables	4
How can you create a bar plot in Seaborn?	a. seaborn.bar()	b. seaborn.plot.bar()	c. seaborn.create_barplot()	d. seaborn.barplot()	4
What does the seaborn.countplot() function do?	a. Creates a count plot of a categorical variable	b. Counts the number of non-null values in a DataFrame	c. Counts the number of unique values in a DataFrame	d. Creates a count plot of a numerical variable	1
1. How can you create a line plot in Pandas?	a. df.plot.scatter(x='column1', y='column2')	b. df.plot.line(x='column1', y='column2')	c. df.plot.bar(x='column1', y='column2')	d. df.plot.pie(x='column1', y='column2')	2
What does the df.plot() method do by default in Pandas?	a. Creates a bar plot	b. Creates a line plot	c. Creates a scatter plot	d. Creates a pie chart	2
. How can you customize the style of a line plot in Pandas?	a. df.plot(style='line')	b. df.plot.line(style='line')	c. df.plot.line().style('line')d. df.plot.line().set_style('line')	d. df.plot.line().set_style('line')	2

What does the df.plot(subplots=True) do in Pandas?	a. Creates a single plot for all columns	b. Creates a separate plot for each column	c. Creates subplots for specific columns	d. Creates a 3D plot	2
How can you add grid lines to a line plot in Pandas?	a. df.plot(grid=True)	b. df.plot.line(grid=True)	c. df.plot().grid()	d. df.plot.line().set_grid(True)	2
What is the purpose of the df.plot(logy=True) parameter in Pandas?	a. Sets the y-axis to logarithmic scale	b. Sets the x-axis to logarithmic scale	c. Sets both x and y axes to logarithmic scale	d. Converts the data to logarithmic format	1
How can you specify the color of the line in a Pandas line plot?	a. df.plot(color='red')	b. df.plot.line(color='red')	c. df.plot.line().set_color('red')	d. df.plot.line().color('red')	2
Which technique is commonly used for dimensionality reduction in unsupervised learning?	a. Principal Component Analysis (PCA)	b. Support Vector Machines	c. Decision Trees	d. K-Means Clustering	1
What is the primary task in a clustering problem in unsupervised learning?	a. Predicting a continuous numerical output	b. Identifying outliers in the dataset	c. Classifying data points into groups with similar characteristics	d. Predicting a categorical output variable	3
Which algorithm is commonly used for clustering in unsupervised learning?	a. Decision Trees	b. Random Forest	c. K-Means	d. Linear Regression	3
What does the term "anomaly detection" refer to in unsupervised learning?	a. Identifying outliers or unusual patterns in data	b. Predicting future outcomes based on historical data	c. Classifying data points into predefined categories	d. Reducing the dimensionality of data	1
Which of the following is an example of an unsupervised learning task?	a. Image classification	b. Spam email detection	c. Customer segmentation	d. Predicting stock prices	3
What is the purpose of the training phase in unsupervised learning?	a. To evaluate the model's performance	b. To predict new, unseen data points	c. To adjust the model parameters based on labeled training data	d. Unsupervised learning doesn't have a traditional training phase	4
Which metric is commonly used to evaluate the performance of a clustering model in unsupervised learning?	a. F1 Score	b. Mean Squared Error (MSE)	c. Silhouette Score	d. Explained Variance	3
What is the purpose of the test dataset in unsupervised learning?	a. To train the model on labeled data	b. To evaluate the model's performance on unseen data	c. To adjust hyperparameters based on training data	d. To predict the output variables	2
In unsupervised learning, what is the primary objective of dimensionality reduction techniques?	a. Increase the number of features	b. Minimize the number of data points	c. Improve model interpretability	d. Reduce the number of input features while preserving relevant information	4
Which algorithm is commonly used for hierarchical clustering in unsupervised learning?	a. K-Means	b. DBSCAN	c. Agglomerative Hierarchical Clustering	d. PCA (Principal Component Analysis)	3

How do you check if a key is present in a dictionary?	a. if key in dictionary:	b. if key = dictionary:	c. if key == dictionary:	d. if key exists dictionary:	1