Practice Questions for Final Assessment

NumPy Questions

1. What is the primary data structure of NumPy? a) `list` b) `Series` c) `ndarray` d) `DataFrame`

   Answer: c) `ndarray`

2. What is the output of `type(np.array([1, 2, 3]))`? a) `list` b) `tuple` c) `numpy.ndarray` d) `int`

   Answer: c) `numpy.ndarray`

3. Given an array `arr = np.array([[1, 2], [3, 4]])`, what does `arr.shape` return? a) `4` b) `(2, 2)` c) `[2, 2]` d) `(2, 4)`

   Answer: b) `(2, 2)`

4. How do you create a 1D array of 10 zeros? a) `np.zeros(10)` b) `np.zeros([10])` c) `np.array(10, 0)` d) `np.zeroes(10)`

   Answer: a) `np.zeros(10)`

5. What is the correct way to select the element `4` from the array `arr = np.array([[1, 2, 3], [4, 5, 6]])`? a) `arr[1, 0]` b) `arr[0, 1]` c) `arr[1][0]` d) `arr.get(1, 0)`

   Answer: a) `arr[1, 0]`

6. Which function creates an array of evenly spaced values over a specified interval? a) `np.arange()` b) `np.linspace()` c) `np.array_range()` d) `np.random.rand()`

   Answer: b) `np.linspace()`

7. What is the result of `np.array([1, 2, 3]) + 10`? a) `np.array([11, 12, 13])` b) `np.array([1, 2, 3, 10])` c) `np.array([1, 2, 3])` d) An error

   Answer: a) `np.array([11, 12, 13])`

8. What does the `reshape()` method do? a) It changes the data type of the array. b) It changes the dimensions of the array without changing its data. c) It adds new elements to the array. d) It flattens the array into a 1D array.

   Answer: b) It changes the dimensions of the array without changing its data.

9. Given `arr = np.array([1, 2, 3, 4, 5])`, what does `arr[arr > 3]` return? a) `np.array([4, 5])` b) `np.array([True, True, True, False, False])` c) `np.array([1, 2, 3])` d) An error

   Answer: a) `np.array([4, 5])`

10. What is the main advantage of using NumPy arrays over Python lists for numerical operations? a) They are easier to create. b) They are more memory-efficient and faster. c) They can store elements of different data types. d) They support object-oriented programming.

    Answer: b) They are more memory-efficient and faster.

Pandas Questions

1. What are the two primary data structures in the Pandas library? a) `list` and `dictionary` b) `ndarray` and `list` c) `Series` and `DataFrame` d) `tuple` and `set`

   Answer: c) `Series` and `DataFrame`

2. How would you read a CSV file named `data.csv` into a Pandas DataFrame? a) `pd.open_csv('data.csv')` b) `pd.read_csv('data.csv')` c) `pd.load_csv('data.csv')` d) `pd.import_csv('data.csv')`

   Answer: b) `pd.read_csv('data.csv')`

3. Given a DataFrame `df`, which method is used to display the first 5 rows? a) `df.tail()` b) `df.info()` c) `df.describe()` d) `df.head()`

   Answer: d) `df.head()`

4. How do you select a single column named `'Age'` from a DataFrame `df`? a) `df.Age` b) `df['Age']` c) Both a and b d) `df.get('Age')`

   Answer: c) Both a and b

5. What is the purpose of the `dropna()` method? a) To drop all rows that contain duplicate values. b) To drop all columns from the DataFrame. c) To drop all rows or columns that contain missing (NaN) values. d) To drop all rows where a specific condition is not met.

   Answer: c) To drop all rows or columns that contain missing (NaN) values.

6. Which method is used to fill in missing (NaN) values in a DataFrame? a) `df.replace_na()` b) `df.fill_na()` c) `df.fillna()` d) `df.impute()`

Answer: c) `df.fillna()`

7. How would you get a summary of the count, mean, standard deviation, and quartiles for the numeric columns in a DataFrame `df`? a) `df.summary()` b) `df.statistics()` c) `df.info()` d) `df.describe()`

   Answer: d) `df.describe()`

8. What is the result of `df.groupby('Category').mean()`? a) It returns the mean of all columns for each group defined by the `'Category'` column. b) It filters the DataFrame to show only the `'Category'` column. c) It returns the mean of the `'Category'` column. d) It sorts the DataFrame by the `'Category'` column.

   Answer: a) It returns the mean of all columns for each group defined by the `'Category'` column.

9. Which method is used to combine two DataFrames based on a common key, similar to a SQL JOIN operation? a) `pd.concat()` b) `pd.join()` c) `pd.merge()` d) All of the above, depending on the use case.

   Answer: c) `pd.merge()`

10. What is the output of `df.shape` on a DataFrame with 10 rows and 3 columns? a) `10, 3` b) `(10, 3)` c) `13` d) `[10, 3]`

    Answer: b) `(10, 3)`

Seaborn Questions

1. Seaborn is a Python data visualization library based on which other library? a) Pandas b) NumPy c) Matplotlib d) SciPy

   Answer: c) Matplotlib

2. What is the primary strength of Seaborn compared to Matplotlib? a) It is faster for creating plots. b) It provides a high-level interface for drawing attractive and informative statistical graphics. c) It is the only library that can create scatter plots. d) It is better for creating 3D plots.

   Answer: b) It provides a high-level interface for drawing attractive and informative statistical graphics.

3. Which of the following plots is best suited for visualizing the distribution of a single numeric variable using Seaborn? a) `sns.heatmap()` b) `sns.boxplot()` c) `sns.scatterplot()` d) `sns.histplot()`

   Answer: d) `sns.histplot()`

4. Which Seaborn function is commonly used to visualize the correlation matrix of a DataFrame? a) `sns.lineplot()` b) `sns.pairplot()` c) `sns.boxplot()` d) `sns.heatmap()`

   Answer: d) `sns.heatmap()`

5. The `hue` parameter in many Seaborn functions is used to: a) Change the color of the entire plot. b) Map a categorical variable to the color of plot elements. c) Set the title of the plot. d) Adjust the brightness of the colors.

   Answer: b) Map a categorical variable to the color of plot elements.

6. Which function creates a plot that shows the relationship between two variables, with the option to draw a linear regression model fit? a) `sns.jointplot()` b) `sns.regplot()` c) `sns.pairplot()` d) `sns.barplot()`

   Answer: b) `sns.regplot()`

7. What does `sns.pairplot()` do? a) It creates a plot showing the distribution of a single variable. b) It creates a grid of plots, showing the relationships between all pairs of variables in a DataFrame. c) It plots a regression line for a pair of variables. d) It displays a matrix of correlation coefficients.

   Answer: b) It creates a grid of plots, showing the relationships between all pairs of variables in a DataFrame.

8. Which of the following plots is used to visualize the relationship between a numeric variable and a categorical variable, showing the distribution of the numeric variable for each category? a) `sns.countplot()` b) `sns.violinplot()` c) `sns.lineplot()` d) `sns.heatmap()`

   Answer: b) `sns.violinplot()`

9. What is the purpose of the `sns.set_theme()` function? a) To save a plot to a file. b) To reset the default Matplotlib parameters. c) To set the aesthetic style of the plots. d) To display a plot.

   Answer: c) To set the aesthetic style of the plots.

10. A `countplot` is a good visualization choice for: a) The relationship between two continuous variables. b) The distribution of a single continuous variable. c) The relationship between a continuous variable and a categorical variable. d) The number of occurrences of each category in a categorical variable.

    Answer: d) The number of occurrences of each category in a categorical variable.

Scikit-learn Questions

1. What is the main purpose of the Scikit-learn library? a) To perform complex mathematical calculations. b) To create interactive data visualizations. c) To provide simple and efficient tools for data mining and data analysis. d) To manage large datasets in memory.

   Answer: c) To provide simple and efficient tools for data mining and data analysis.

2. Which of the following is not a standard step in a typical Scikit-learn workflow? a) Importing the model. b) Instantiating the model. c) Training the model with `model.fit()`. d) Creating a new programming language.

   Answer: d) Creating a new programming language.

3. What is the purpose of the `fit()` method in Scikit-learn? a) To make predictions on new data. b) To evaluate the model's performance. c) To train the model using training data. d) To transform the data into a different format.

   Answer: c) To train the model using training data.

4. Which module is primarily used for splitting data into training and testing sets? a) `sklearn.linear_model` b) `sklearn.model_selection` c) `sklearn.preprocessing` d) `sklearn.metrics`

   Answer: b) `sklearn.model_selection`

5. What is the main purpose of the `predict()` method? a) To evaluate the model's accuracy. b) To apply the trained model to new data to make predictions. c) To train the model with new data. d) To normalize the feature values.

   Answer: b) To apply the trained model to new data to make predictions.

6. Which of these is a supervised learning algorithm? a) K-Means Clustering b) Principal Component Analysis (PCA) c) Linear Regression d) DBSCAN

Answer: c) Linear Regression

7. Which of these is an unsupervised learning algorithm? a) Logistic Regression b) Support Vector Machines (SVM) c) Random Forest d) K-Means Clustering

   Answer: d) K-Means Clustering

8. What is the purpose of `StandardScaler` in `sklearn.preprocessing`? a) To encode categorical features into numeric form. b) To scale features by removing the mean and scaling to unit variance. c) To handle missing values in the dataset. d) To reduce the number of features in a dataset.

   Answer: b) To scale features by removing the mean and scaling to unit variance.

9. What is the output of the `score()` method for a classification model? a) The model's loss value. b) The model's accuracy on the provided data. c) A list of predictions. d) The model's training time.

   Answer: b) The model's accuracy on the provided data.

10. In a `Pipeline` object, what is the correct order of operations? a) Estimators followed by transformers. b) Transformers followed by an estimator. c) Estimators only. d) Transformers only.

    Answer: b) Transformers followed by an estimator.

TensorFlow/Keras Questions

1. What is the primary purpose of Keras as a high-level API for TensorFlow? a) To perform low-level hardware optimizations. b) To provide an easy-to-use interface for building and training neural networks. c) To manage and version datasets. d) To visualize complex model architectures.

   Answer: b) To provide an easy-to-use interface for building and training neural networks.

2. In a Keras Sequential model, what does the `Dense` layer represent? a) A convolutional layer. b) A recurrent neural network layer. c) A fully connected layer. d) A dropout layer.

   Answer: c) A fully connected layer.

3. What is the purpose of the `model.compile()` method in Keras? a) To train the model on the data. b) To define the model's architecture. c) To configure the learning process by specifying the optimizer, loss function, and metrics. d) To make predictions on new data.

Answer: c) To configure the learning process by specifying the optimizer, loss function, and metrics.

4. Which of the following is an example of an optimizer used in Keras? a) `categorical_crossentropy` b) `sigmoid` c) `Adam` d) `accuracy`

   Answer: c) `Adam`

5. What does the `loss` function measure during the training of a neural network? a) The time it takes for the model to train. b) The number of parameters in the model. c) The performance of the model on the training data. d) The difference between the predicted output and the true output.

   Answer: d) The difference between the predicted output and the true output.

6. How would you add a layer to a Keras Sequential model named `model`? a) `model.add(Dense(32, activation='relu'))` b) `model.insert(Dense(32, activation='relu'))` c) `model.layer(Dense(32, activation='relu'))` d) `model.append(Dense(32, activation='relu'))`

   Answer: a) `model.add(Dense(32, activation='relu'))`

7. What is the purpose of the `activation` function in a neural network layer? a) To determine the learning rate of the model. b) To introduce non-linearity into the model's output. c) To split the data into training and testing sets. d) To regularize the model and prevent overfitting.

   Answer: b) To introduce non-linearity into the model's output.

8. Which of the following is a common `loss` function for binary classification problems in Keras? a) `mean_squared_error` b) `categorical_crossentropy` c) `sparse_categorical_crossentropy` d) `binary_crossentropy`

   Answer: d) `binary_crossentropy`

9. What is the correct method to train a Keras model? a) `model.fit(x_train, y_train, epochs=10)` b) `model.train(x_train, y_train, epochs=10)` c) `model.run(x_train, y_train, epochs=10)` d) `model.execute(x_train, y_train, epochs=10)`

   Answer: a) `model.fit(x_train, y_train, epochs=10)`

10. What does the `Dropout` layer in Keras do? a) It randomly sets a fraction of input units to zero during training to prevent overfitting. b) It removes layers from the model during training. c) It increases the number of connections between layers. d) It adds noise to the input data.

    Answer: a) It randomly sets a fraction of input units to zero during training to prevent overfitting.