Python Data Science Examination: 100 Multiple-Choice Questions

Python Lists, Tuples, Sets, and Dictionaries

1. What is the main difference between a list and a tuple in Python?

   a) Lists are immutable, while tuples are mutable.

   b) Lists can hold different data types, while tuples can only hold a single data type.

   c) Lists are mutable, while tuples are immutable.

   d) Lists are defined with parentheses `()`, while tuples are defined with square brackets `[]`.

   Correct Answer: c) Lists are mutable, while tuples are immutable.

2. Which of the following is an immutable data structure?

   a) List

   b) Dictionary

   c) Set

   d) Tuple

   Correct Answer: d) Tuple

3. How do you create an empty set in Python?

   a) `my_set = {}`

   b) `my_set = set()`

   c) `my_set = []`

   d) `my_set = tuple()`

   Correct Answer: b) `my_set = set()`

4. What is the output of `list(set([1, 2, 2, 3, 4, 4]))`?

   a) `[1, 2, 2, 3, 4, 4]`

   b) `[1, 2, 3, 4]`

   c) `(1, 2, 3, 4)`

   d) `{1, 2, 3, 4}`

   Correct Answer: b) `[1, 2, 3, 4]`

5. Which method is used to add an element to a set?

   a) `append()`

   b) `add()`

   c) `insert()`

   d) `extend()`

   Correct Answer: b) `add()`

6. What is the time complexity of searching for an element in a set?

   a) O(1)

   b) O(n)

   c) O(log n)

   d) $O(n^2)$

   Correct Answer: a) O(1)

7. Which of the following is a valid way to access the value 'banana' from `my_dict = {'fruit': 'banana', 'color': 'yellow'}`?

   a) `my_dict[0]`

   b) `my_dict.value('banana')`

   c) `my_dict['fruit']`

   d) `my_dict.get('banana')`

   Correct Answer: c) `my_dict['fruit']`

8. What happens if you try to access a non-existent key in a dictionary using square brackets `[]`?

   a) A `KeyError` is raised.

   b) It returns `None`.

   c) It creates a new key with a `None` value.

   d) A `ValueError` is raised.

   Correct Answer: a) A `KeyError` is raised.

9. What is the output of `[1, 2] + [3, 4]`?

   a) `[1, 2, 3, 4]`

   b) `[4, 6]`

   c) `[1, 2, [3, 4]]`

   d) `(1, 2, 3, 4)`

   Correct Answer: a) `[1, 2, 3, 4]`

10. What is the correct syntax for a list comprehension to create a list of squares for numbers 0 to 4?

    a) `[x**2 for x in range(5)]`

    b) `(x**2 for x in range(5))`

    c) `{x**2 for x in range(5)}`

    d) `[x^2 for x in range(5)]`

    Correct Answer: a) `[x**2 for x in range(5)]`

11. How do you remove the element `3` from the list `my_list = [1, 2, 3, 4]`?

    a) `my_list.remove(3)`

    b) `my_list.pop(3)`

    c) `my_list.delete(3)`

    d) `del my_list[3]`

    Correct Answer: a) `my_list.remove(3)`

12. Which of the following statements about sets is true?

a) Sets are ordered collections.

b) Sets can contain duplicate elements.

c) Sets are mutable.

d) Sets use keys and values.

Correct Answer: c) Sets are mutable.

13. What is the output of `(1, 2, 3) * 2`?

a) `(1, 2, 3, 1, 2, 3)`

b) `(2, 4, 6)`

c) `[1, 2, 3, 1, 2, 3]`

d) `(1, 2, 3) (1, 2, 3)`

Correct Answer: a) `(1, 2, 3, 1, 2, 3)`

14. How can you get all the keys from a dictionary `my_dict = {'a': 1, 'b': 2}`?

a) `my_dict.keys()`

b) `my_dict.values()`

c) `my_dict.items()`

d) `my_dict.get_keys()`

Correct Answer: a) `my_dict.keys()`

15. What is the output of `'a' in {'a': 1, 'b': 2}`?

a) `True`

b) `False`

c) `KeyError`

d) `ValueError`

Correct Answer: a) `True`

16. How do you find the number of elements in a tuple `my_tuple = (10, 20, 30)`?

a) `my_tuple.size()`

b) `len(my_tuple)`

c) `my_tuple.count()`

d) `my_tuple.length()`

Correct Answer: b) `len(my_tuple)`

17. Which of the following is NOT a valid way to create a dictionary?

a) `dict()`

b) `{'key1': 'value1'}`

c) `dict(key1='value1')`

d) `{'key1': 'value1', 'key1': 'value2'}`

Correct Answer: d) `{'key1': 'value1', 'key1': 'value2'}`

18. What is the purpose of a tuple?

a) To store a collection of mutable items.

b) To store a collection of unique items.

c) To store a fixed collection of items that should not be changed.

d) To store key-value pairs.

Correct Answer: c) To store a fixed collection of items that should not be changed.

19. Which operation is more efficient for checking membership: `in` on a list or `in` on a set?

a) `in` on a list is more efficient.

b) `in` on a set is more efficient.

c) They have the same efficiency.

d) It depends on the size of the collection.

Correct Answer: b) `in` on a set is more efficient.

20. What is the output of `[1, 2, 3][1:3]`?

a) `[1, 2]`

b) `[2, 3]`

c) `[2]`

d) `[1, 2, 3]`

Correct Answer: b) `[2, 3]`

21. How do you merge two dictionaries `d1 = {'a': 1}` and `d2 = {'b': 2}`?

a) `d1.update(d2)`

b) `d1 + d2`

c) `d1.merge(d2)`

d) `d1.append(d2)`

Correct Answer: a) `d1.update(d2)`

22. Which of the following is a key characteristic of a dictionary?

a) It is ordered and mutable.

b) It stores a collection of unique, unordered items.

c) It is a collection of key-value pairs.

d) It is immutable.

Correct Answer: c) It is a collection of key-value pairs.

23. What is the output of `set([1, 2, 3]) | set([3, 4, 5])`?

a) `{1, 2, 3, 4, 5}`

b) `{3}`

c) `{1, 2, 4, 5}`

d) `{1, 2, 3, 3, 4, 5}`

Correct Answer: a) `{1, 2, 3, 4, 5}`

24. What is the output of `set([1, 2, 3]) & set([3, 4, 5])`?

a) `{1, 2, 3, 4, 5}`

b) `{3}`

c) `{1, 2, 4, 5}`

d) `{1, 2, 3, 3, 4, 5}`

Correct Answer: b) `{3}`

25. What is the output of `list('hello')`?

a) `['h', 'e', 'l', 'l', 'o']`

b) `'hello'`

c) `('h', 'e', 'l', 'l', 'o')`

d) `{'h', 'e', 'l', 'l', 'o'}`

Correct Answer: a) `['h', 'e', 'l', 'l', 'o']`

***

NumPy

26. What is the main purpose of NumPy in data science?

a) To create interactive visualizations.

b) To perform high-performance numerical operations, especially on arrays.

c) To manage databases.

d) To build web applications.

Correct Answer: b) To perform high-performance numerical operations, especially on arrays.

27. How do you create a 3x3 NumPy array filled with zeros?

a) `np.zeros(3, 3)`

b) `np.zeros([3, 3])`

c) `np.zeros((3, 3))`

d) `np.array(0, 3, 3)`

Correct Answer: c) `np.zeros((3, 3))`

28. What is the key advantage of NumPy arrays over Python lists for numerical operations?

a) NumPy arrays are smaller in size.

b) NumPy arrays are faster and more memory-efficient.

c) NumPy arrays can store different data types.

d) NumPy arrays are mutable.

Correct Answer: b) NumPy arrays are faster and more memory-efficient.

29. How do you find the shape of a NumPy array `my_array`?

a) `my_array.shape()`

b) `shape(my_array)`

c) `my_array.size`

d) `my_array.shape`

Correct Answer: d) `my_array.shape`

30. What is the result of `np.array([1, 2, 3]) * 2`?

a) `[2, 4, 6]`

b) `[1, 2, 3, 1, 2, 3]`

c) `[1, 2, 3, 2]`

d) A `TypeError` is raised.

Correct Answer: a) `[2, 4, 6]`

31. How do you create a NumPy array with values from 0 to 9?

a) `np.range(10)`

b) `np.arange(10)`

c) `np.list(10)`

d) `np.array(10)`

Correct Answer: b) `np.arange(10)`

32. What is broadcasting in NumPy?

a) The process of converting an array to a list.

b) A mechanism for performing arithmetic operations on arrays of different shapes.

c) A method for saving an array to a file.

d) The process of creating a deep copy of an array.

Correct Answer: b) A mechanism for performing arithmetic operations on arrays of different shapes.

33. How do you perform matrix multiplication between two NumPy arrays `a` and `b`?

a) `a * b`

b) `np.dot(a, b)`

c) `a + b`

d) `a.multiply(b)`

Correct Answer: b) `np.dot(a, b)`

34. What is the output of `np.sum(np.array([1, 2, 3]))`?

a) `[1, 2, 3]`

b) `6`

c) `3`

d) `1`

Correct Answer: b) `6`

35. How do you reshape a 1D array `[1, 2, 3, 4, 5, 6]` into a 2x3 array?

a) `my_array.reshape(2, 3)`

b) `my_array.reshape([2, 3])`

c) `my_array.shape = (2, 3)`

d) `np.reshape(my_array, 2, 3)`

Correct Answer: a) `my_array.reshape(2, 3)`

36. What does the `axis=0` argument in a NumPy function (like `np.sum()`) typically refer to?

a) The columns of the array.

b) The rows of the array.

c) The depth of the array.

d) The total sum of all elements.

Correct Answer: b) The rows of the array.

37. How do you select all elements greater than 5 from a NumPy array `my_array`?

a) `my_array > 5`

b) `my_array[my_array > 5]`

c) `my_array.filter(> 5)`

d) `my_array.where(my_array > 5)`

Correct Answer: b) `my_array[my_array > 5]`

38. What is the data type of a NumPy array?

a) `list`

b) `ndarray`

c) `array`

d) `vector`

Correct Answer: b) `ndarray`

39. How do you create an array of 5 random integers between 0 and 10?

a) `np.random.rand(5, 0, 10)`

b) `np.random.randint(0, 10, 5)`

c) `np.random.choice(10, 5)`

d) `np.random.random(5)`

Correct Answer: b) `np.random.randint(0, 10, 5)`

40. What is the output of `np.array([1, 2, 3]) + np.array([4, 5, 6])`?

a) `[5, 7, 9]`

b) `[1, 2, 3, 4, 5, 6]`

c) `[[1, 4], [2, 5], [3, 6]]`

d) A `TypeError` is raised.

Correct Answer: a) `[5, 7, 9]`

***

Pandas

41. What is the primary data structure in Pandas?

a) `ndarray`

b) `list`

c) `DataFrame`

d) `Dictionary`

Correct Answer: c) `DataFrame`

42. How do you read a CSV file named `data.csv` into a Pandas DataFrame?

a) `pd.load_csv('data.csv')`

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

c) `pd.open_csv('data.csv')`

d) `pd.import_csv('data.csv')`

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

43. How do you select a single column named `Age` from a DataFrame `df`?

a) `df.Age`

b) `df['Age']`

c) `df.loc[:, 'Age']`

d) All of the above.

Correct Answer: d) All of the above.

44. Which method is used to get the first 5 rows of a DataFrame `df`?

a) `df.tail(5)`

b) `df.head()`

c) `df.head(5)`

d) `df.first(5)`

Correct Answer: c) `df.head(5)`

45. How do you check for missing values in a DataFrame `df`?

a) `df.isnull()`

b) `df.has_null()`

c) `df.check_na()`

d) `df.na()`

Correct Answer: a) `df.isnull()`

46. What is the purpose of the `groupby()` method in Pandas?

a) To group rows in a DataFrame based on a column's values.

b) To merge two DataFrames together.

c) To apply a function to every element in a DataFrame.

d) To sort the DataFrame by a specific column.

Correct Answer: a) To group rows in a DataFrame based on a column's values.

47. How do you drop a column named `Age` from a DataFrame `df` in-place?

a) `df.drop('Age', axis=1, inplace=True)`

b) `df.delete('Age')`

c) `df.drop_column('Age')`

d) `df.drop('Age', axis=0)`

Correct Answer: a) `df.drop('Age', axis=1, inplace=True)`

48. How do you get descriptive statistics (like mean, std, etc.) for a DataFrame `df`?

a) `df.stats()`

b) `df.describe()`

c) `df.info()`

d) `df.summary()`

Correct Answer: b) `df.describe()`

49. What is the primary difference between `.loc[]` and `.iloc[]`?

a) `.loc[]` is for integer-based indexing, while `.iloc[]` is for label-based indexing.

b) `.loc[]` is for label-based indexing, while `.iloc[]` is for integer-based indexing.

c) They are interchangeable and do the same thing.

d) `.loc[]` is used for rows, and `.iloc[]` is used for columns.

Correct Answer: b) `.loc[]` is for label-based indexing, while `.iloc[]` is for integer-based indexing.

50. How do you create a new column `Total` in a DataFrame `df` that is the sum of columns `A` and `B`?

a) `df['Total'] = df['A'] + df['B']`

b) `df.add('Total', df['A'], df['B'])`

c) `df.Total = df.A + df.B`

d) `df['Total'] = np.sum(df[['A', 'B']], axis=1)`

Correct Answer: a) `df['Total'] = df['A'] + df['B']`

51. How do you fill missing values (NaN) in a DataFrame `df` with the mean of each column?

a) `df.fillna(df.mean())`

b) `df.replace_na(df.mean())`

c) `df.fillna('mean')`

d) `df.fill_na_with_mean()`

Correct Answer: a) `df.fillna(df.mean())`

52. What is the default `axis` for most Pandas operations like `drop()` or `mean()`?

a) `axis=0` (row-wise)

b) `axis=1` (column-wise)

c) `axis=None`

d) `axis='rows'`

Correct Answer: a) `axis=0` (row-wise)

53. How do you perform a left merge on two DataFrames `df1` and `df2` on a common column `id`?

a) `pd.merge(df1, df2, on='id', how='left')`

b) `df1.merge(df2, on='id', how='left')`

c) Both a and b are correct.

d) Neither a nor b are correct.

Correct Answer: c) Both a and b are correct.

54. What is a Series in Pandas?

a) A 2-dimensional labeled data structure.

b) A 1-dimensional labeled array.

c) A collection of dictionaries.

d) A list of lists.

Correct Answer: b) A 1-dimensional labeled array.

55. How do you change the data type of a column `col` to an integer in a DataFrame `df`?

a) `df['col'] = df['col'].to_int()`

b) `df['col'] = df['col'].astype('int')`

c) `df.col.int()`

d) `df.int_col('col')`

Correct Answer: b) `df['col'] = df['col'].astype('int')`

***

scikit-learn (Sklearn)

56. What is the primary goal of scikit-learn?

    a) To perform deep learning.

    b) To provide a simple and efficient tool for data analysis and machine learning.

    c) To manage big data.

    d) To create complex visualizations.

    Correct Answer: b) To provide a simple and efficient tool for data analysis and machine learning.

57. Which module in scikit-learn is used for splitting data into training and testing sets?

    a) `sklearn.model_selection`

    b) `sklearn.preprocessing`

    c) `sklearn.datasets`

    d) `sklearn.metrics`

    Correct Answer: a) `sklearn.model_selection`

58. What is the purpose of `StandardScaler`?

    a) To normalize data by scaling features to a unit norm.

    b) To transform data by removing the mean and scaling to unit variance.

    c) To encode categorical features into numerical ones.

    d) To fill missing values in a dataset.

    Correct Answer: b) To transform data by removing the mean and scaling to unit variance.

59. What is the typical workflow for using a scikit-learn model?

    a) `fit()`, `predict()`, `transform()`

    b) `fit()`, `predict()`, `score()`

c) `instantiate()`, `fit()`, `predict()`

d) `train()`, `test()`, `evaluate()`

Correct Answer: c) `instantiate()`, `fit()`, `predict()`

60. Which of the following is a supervised learning algorithm?

a) K-Means Clustering

b) Principal Component Analysis (PCA)

c) Linear Regression

d) Hierarchical Clustering

Correct Answer: c) Linear Regression

61. What is a common metric for evaluating a classification model's performance?

a) Mean Squared Error (MSE)

b) R-squared

c) Accuracy

d) Root Mean Squared Error (RMSE)

Correct Answer: c) Accuracy

62. What is the purpose of `GridSearchCV`?

a) To automatically find the best hyperparameters for a model.

b) To perform cross-validation on a dataset.

c) To plot a grid of data.

d) To select the best features for a model.

Correct Answer: a) To automatically find the best hyperparameters for a model.

63. Which scikit-learn module is used for dimensionality reduction?

a) `sklearn.linear_model`

b) `sklearn.decomposition`

c) `sklearn.ensemble`

d) `sklearn.tree`

Correct Answer: b) `sklearn.decomposition` (e.g., PCA)

64. What does the `fit()` method do in a scikit-learn model?

a) It makes predictions on new data.

b) It trains the model on the provided data.

c) It evaluates the model's performance.

d) It transforms the data.

Correct Answer: b) It trains the model on the provided data.

65. What is the output of `train_test_split(X, y, test_size=0.2)`?

a) `X_train`, `X_test`, `y_train`, `y_test`

b) `X_train`, `y_train`

c) `X_test`, `y_test`

d) A single tuple containing all four arrays.

Correct Answer: a) `X_train`, `X_test`, `y_train`, `y_test`

66. What is the `random_state` parameter used for in `train_test_split`?

a) To shuffle the data randomly.

b) To set a seed for the random number generator to ensure reproducibility.

c) To select a random test size.

d) To randomly select features.

Correct Answer: b) To set a seed for the random number generator to ensure reproducibility.

67. What is the purpose of the `predict()` method?

a) To train the model.

b) To generate predictions for new, unseen data.

c) To evaluate the model's performance on the training data.

d) To preprocess the data.

Correct Answer: b) To generate predictions for new, unseen data.

68. Which of the following is a key component of the scikit-learn API?

a) All models have a `.fit()` and `.predict()` method.

b) All models are optimized for GPU usage.

c) All models are based on neural networks.

d) All models require data to be in a Pandas DataFrame.

Correct Answer: a) All models have a `.fit()` and `.predict()` method.

69. What is a common unsupervised learning task?

a) Classification

b) Regression

c) Clustering

d) Anomaly detection

Correct Answer: c) Clustering

70. Which of the following is used to handle categorical data?

a) `StandardScaler`

b) `OneHotEncoder`

c) `MinMaxScaler`

d) `PolynomialFeatures`

Correct Answer: b) `OneHotEncoder`

***

TensorFlow and Keras

71. What is TensorFlow primarily used for?

a) Statistical analysis

b) Web development

c) High-performance numerical computation and large-scale machine learning

d) Data visualization

Correct Answer: c) High-performance numerical computation and large-scale machine learning

72. What is Keras?

a) A deep learning library that is a high-level API for TensorFlow.

b) A low-level programming language.

c) A data visualization tool.

d) A machine learning library for traditional algorithms.

Correct Answer: a) A deep learning library that is a high-level API for TensorFlow.

73. What is the purpose of an optimizer in a neural network?

a) To compile the model.

b) To define the network architecture.

c) To adjust the model's weights and biases to minimize the loss function.

d) To evaluate the model's performance.

Correct Answer: c) To adjust the model's weights and biases to minimize the loss function.

74. What is an `activation function` in a neural network?

a) A function that calculates the loss.

b) A function applied to the output of a neuron to introduce non-linearity.

c) A function that initializes the weights.

d) A function that splits the data.

Correct Answer: b) A function applied to the output of a neuron to introduce non-linearity.

75. What is the role of the `Dense` layer in Keras?

a) It is a convolutional layer for image processing.

b) It is a recurrent layer for sequence data.

c) It is a fully connected layer where every neuron is connected to all neurons in the previous layer.

d) It is an output layer that performs classification.

Correct Answer: c) It is a fully connected layer where every neuron is connected to all neurons in the previous layer.

76. How do you compile a Keras model?

a) `model.fit()`

b) `model.compile()`

c) `model.summary()`

d) `model.evaluate()`

Correct Answer: b) `model.compile()`

77. What is the purpose of the `loss function`?

a) To predict the output.

b) To measure how well the model is performing.

c) To define the model's architecture.

d) To update the model's weights.

Correct Answer: b) To measure how well the model is performing.

78. Which of the following is a common loss function for a binary classification problem?

a) Mean Squared Error

b) Categorical Crossentropy

c) Binary Crossentropy

d) Mean Absolute Error

Correct Answer: c) Binary Crossentropy

79. What does the `fit()` method do in Keras?

a) It compiles the model.

b) It trains the model for a fixed number of epochs.

c) It makes predictions on new data.

d) It saves the trained model to a file.

Correct Answer: b) It trains the model for a fixed number of epochs.

80. What is an epoch?

a) A single pass of the entire training dataset through the model.

b) A single pass of a mini-batch of data.

c) The number of layers in the network.

d) The number of neurons in a layer.

Correct Answer: a) A single pass of the entire training dataset through the model.

81. What is the purpose of the `learning rate` in an optimizer?

a) It determines how quickly the model's weights are updated.

b) It defines the number of epochs.

c) It specifies the batch size.

d) It controls the number of layers.

Correct Answer: a) It determines how quickly the model's weights are updated.

82. Which of the following is an activation function often used in the output layer of a multi-class classification model?

a) Sigmoid

b) ReLU

c) Tanh

d) Softmax

Correct Answer: d) Softmax

83. What is the `Sequential` model in Keras?

a) A model for recurrent neural networks.

b) A way to create a model layer by layer.

c) A pre-trained model.

d) A model with multiple inputs and outputs.

Correct Answer: b) A way to create a model layer by layer.

84. How do you save a trained Keras model?

a) `model.save('my_model.h5')`

b) `model.export('my_model.h5')`

c) `model.write('my_model.h5')`

d) `model.dump('my_model.h5')`

Correct Answer: a) `model.save('my_model.h5')`

85. What is the purpose of the `validation_split` argument in the `fit()` method?

a) To split the data into training and testing sets before training.

b) To reserve a portion of the training data to evaluate the model's performance during training.

c) To perform a grid search for hyperparameters.

d) To apply a specific transformation to the data.

Correct Answer: b) To reserve a portion of the training data to evaluate the model's performance during training.

***

General Data Science

86. What is a key characteristic of supervised learning?

a) The algorithm learns from unlabeled data.

b) The algorithm learns from labeled data with input features and corresponding output labels.

c) The algorithm finds patterns without any guidance.

d) The algorithm is used for dimensionality reduction.

Correct Answer: b) The algorithm learns from labeled data with input features and corresponding output labels.

87. Which of the following is an example of a supervised learning task?

a) Clustering a dataset of customer purchase histories.

b) Predicting house prices based on features like size and location.

c) Grouping similar images together.

d) Detecting outliers in a dataset.

Correct Answer: b) Predicting house prices based on features like size and location.

88. What is the goal of unsupervised learning?

a) To make predictions on new data.

b) To find hidden patterns or intrinsic structures in unlabeled data.

c) To classify data into predefined categories.

d) To calculate the loss of a model.

Correct Answer: b) To find hidden patterns or intrinsic structures in unlabeled data.

89. What is a common technique for handling imbalanced datasets in classification?

a) Using `StandardScaler`.

b) Undersampling the majority class or oversampling the minority class.

c) Removing all outliers.

d) Using a different optimizer.

Correct Answer: b) Undersampling the majority class or oversampling the minority class.

90. What is a confusion matrix used for?

a) To visualize a scatter plot.

b) To evaluate the performance of a classification model.

c) To find the correlation between features.

d) To measure the mean squared error.

Correct Answer: b) To evaluate the performance of a classification model.

91. What is the `bias-variance trade-off`?

a) The trade-off between model accuracy and training time.

b) The trade-off between model simplicity and complexity.

c) The trade-off between a model's ability to fit training data well (low bias) and its ability to generalize to new data (low variance).

d) The trade-off between using a large dataset and a small dataset.

Correct Answer: c) The trade-off between a model's ability to fit training data well (low bias) and its ability to generalize to new data (low variance).

92. What is cross-validation?

a) A technique to split the data into training and testing sets once.

b) A technique to evaluate a model by training it on different subsets of the data and validating it on the remaining subset.

c) A method for hyperparameter tuning.

d) A way to normalize the data.

Correct Answer: b) A technique to evaluate a model by training it on different subsets of the data and validating it on the remaining subset.

93. What is overfitting?

a) When a model is too simple and cannot capture the underlying patterns in the data.

b) When a model performs well on the training data but poorly on unseen data.

c) When a model performs poorly on both training and test data.

d) When a model is trained on a small dataset.

Correct Answer: b) When a model performs well on the training data but poorly on unseen data.

94. What is a key purpose of feature engineering?

a) To reduce the number of features.

b) To create new features from existing ones to improve model performance.

c) To plot the features.

d) To normalize the data.

Correct Answer: b) To create new features from existing ones to improve model performance.

95. What is the purpose of `matplotlib`?

a) To perform numerical operations.

b) To create static, animated, and interactive visualizations in Python.

c) To manage dataframes.

d) To build machine learning models.

Correct Answer: b) To create static, animated, and interactive visualizations in Python.

96. Which library is often used for creating statistical data visualizations?

a) `Numpy`

b) `Scipy`

c) `Seaborn`

d) `Keras`

Correct Answer: c) `Seaborn`

97. What is a `hyperparameter`?

a) A parameter learned by the model during training.

b) A parameter whose value is set before the training process begins.

c) The input features of the dataset.

d) The output label of the dataset.

Correct Answer: b) A parameter whose value is set before the training process begins.

98. What is the difference between supervised and unsupervised learning?

a) Supervised learning uses labeled data, while unsupervised learning uses unlabeled data.

b) Supervised learning is for regression, and unsupervised learning is for classification.

c) Supervised learning is used for clustering, while unsupervised learning is used for prediction.

d) Supervised learning is faster than unsupervised learning.

Correct Answer: a) Supervised learning uses labeled data, while unsupervised learning uses unlabeled data.

99. What does the term "regularization" refer to in machine learning?

a) A technique to prevent underfitting.

b) A technique to normalize the data.

c) A technique to prevent overfitting by adding a penalty to the model's loss function.

d) A method for feature selection.

Correct Answer: c) A technique to prevent overfitting by adding a penalty to the model's loss function.

100. What is a common way to handle multicollinearity in a dataset?

a) Use `StandardScaler`.

b) Remove one of the highly correlated features.

c) Add more features.

d) Increase the number of epochs.

Correct Answer: b) Remove one of the highly correlated features.