Question 1

0.5 pts

Which machine learning technique is commonly used for binary classification tasks?

- ○ Linear Regression
- ○ K-Means Clustering
- ○ PCA
- ○ Logistic Regression

Question 2

1 pts

If the initial centroids for K-Means Clustering are [(10, 10), (0, 0)] and the points are [(8, 8), (2, 2), (12, 12), (4, 4)], what are the centroids after the first iteration?

- ○ [(2, 2), (12, 12)]
- ○ [(9, 9), (3, 3)]
- ○ [(8, 8), (4, 4)]
- ○ [(7, 7), (5, 5)]
- ○ [(10, 10), (3, 3)]

Question 3

1 pts

In a multi-class classification problem with four classes (A, B, C, D), the confusion matrix of a model is as follows:

	Predicted A	Predicted B	Predicted C	Predicted D
Actual A	80	10	5	5
Actual B	15	70	10	5
Actual C	10	8	70	12
Actual D	6	4	10	80

What is the accuracy of the model?

- ○ 0.80
- ○ 0.85
- ○ 0.75
- ○ 0.70

Question 4 — 0.5 pts

Consider a dataset containing a categorical feature "City" with three unique values: "New York," "Los Angeles," and "Chicago." After applying OneHotEncoding to this feature, how many new binary columns will be created?

- ○ 1
- ○ 2
- ○ 3
- ○ 4

Question 5 — 0.5 pts

Which of the following has its own address space?

- ○ thread
- ○ CPU
- ○ process
- ○ program

Question 6 — 0.5 pts

If you want to produce a dendrogram, what should you use?

- ○ LinearRegression
- ○ LogisticRegression
- ○ KMeans
- ○ AgglomerativeClustering
- ○ PCA

Question 7 — 1 pts

If you want to randomly split your data into train and test, but you don't want your results to change if you re-run your notebook, what should you pass to train_test_split?

- ○ test_size=0.75
- ○ train_size=320
- ○ random_state=50
- ○ stratify=False

Question 8

0.5 pts

Which scenario is most appropriate for choosing **KMeans** clustering over **Agglomerative Clustering**?

- ○ When you need to assign new data points to existing clusters.
- ○ When your dataset is small.
- ○ When your computer has limited memory.
- ○ When you want to show the step-by-step cluster formation.

Question 9

1 pts

Given *matrix = numpy.array([[1, 2, 4], [2, 4, 1]])*, what is *matrix.argmax(axis = 0)*?

- ○ [0, 0, 0]
- ○ [1, 1, 0]
- ○ [2, 1]
- ○ [2, 4, 4]
- ○ [1, 2]

Question 10

0.5 pts

Which of the following is not an sklearn pipeline **estimator**?

- ○ SVC
- ○ LinearRegression
- ○ LogisticRegression
- ○ PolynomialFeatures

Question 11

1 pts

The shape of A is (3, 2), the shape of B is (2, 3), and the shape of C is (3, 4). What is the shape of A @ B @ C?

- ○ (3, 3)
- ○ (3, 4)
- ○ (Error)
- ○ (4, 2)

Question 12

1 pts

```
sum = 0
for i in A[:100]:
    for j in B[:-100]:
        sum += i * j
```

Given A is a list of M elements and B is a list of N elements, what's the time complexity of the above code?

- ○ O(M)
- ○ O(N)
- ○ O(MN)
- ○ O(1)

Question 13

0.5 pts

Assume *element* is a Selenium WebElement given by *link*. Which of the following returns "page.html"? Choose the best option.

- ○ element.text
- ○ element.href
- ○ element.get_attribute("href")
- ○ element.get_attribute("text")

Question 14

1 pts

If a dataframe *df* has 100 columns and 10 rows. After applying *p = PCA(5)* and *p.fit_transform(df)*, what is the shape of *p.components_*?

- ○ (95, 100)
- ○ (100, 95)
- ○ (5, 100)
- ○ (5, 95)
- ○ (95, 5)

Question 15

1 pts

If A = np.array([[2, 3], [1, 4]]) and b = np.array([[5], [2]]), what is b * A?

- ○ np.array([18, 17])
- ○ np.array([[18], [17]])
- ○ np.array([[10, 8], [15, 2]])
- ○ np.array([[10, 15], [2, 8]])

Question 16

0.5 pts

What call makes predictions using a computation similar to **X @ c**? Choose the best option.

- ○ LinearRegression.predict
- ○ LogisticRegression.predict
- ○ LinearRegression.predict_proba
- ○ LogisticRegression.predict_proba

Question 17

1 pts

Consider the confusion matrix of a model:

	apples	**oranges**	**bananas**
apples	4	4	2
oranges	0	2	5
bananas	0	4	1

What is the *recall* for apples?

- ○ 10
- ○ 0.1
- ○ 1
- ○ 0.4
- ○ 4
- ○ 7

Question 18

0.5 pts

Consider the confusion matrix of a model:

	apples	oranges	bananas
apples	1	2	2
oranges	5	4	1
bananas	4	4	7

What is the *precision* for bananas?

- ○ 1
- ○ 10
- ○ 0.1
- ○ 7
- ○ 0.7
- ○ 4

Question 19

0.5 pts

When evaluating models using cross-validation, which set of characteristics in the cross-validation scores suggests a model's performance is more consistent and accurate?

- ○ Large mean, large standard deviation
- ○ Large mean, small standard deviation
- ○ Small mean, large standard deviation
- ○ Small mean, small standard deviation