Week 7 Quiz

True or False

1/1 point (graded)

In machine learning, algorithms and programs directly aim to learn a given task.

- True
- False ✔

Answer

Correct: Video: "Introduction to Machine Learning"

Submit

Multiple Choice

1/1 point (graded)

What is the definition of 'data mining'?

- Activities related to finding patterns in databases and data warehouses.

- Process of inspecting, cleansing, transforming, and engineering a particular dataset.
- Query processing and statistical analysis to summarize a dataset.

✔

Answer
Correct: Video: "Introduction to Machine Learning"

Submit

True or False

1/1 point (graded)
When you search an incorrectly spelled term online, suggested words is an example of machine learning.

- True
- False

✔

Answer
Correct: Video: "Introduction to Machine Learning"

Submit

Multiple Choice

1/1 point (graded)

When would you use the machine learning technique ‘regression’?

- When your model has to predict a categorical value.
- When your model has to predict a numerical value. ✔
- When you want to organize similar items in your dataset into groups.
- When you want to capture associations between items.

Answer
Correct: Video: "Categories of Machine Learning"

Submit

Multiple Choice

1/1 point (graded)
As an example, you have a dataset containing numerical values of subjects’ heart rates during exercise and categorical values describing how much they smoke. You want to determine whether smoking and heart rate are related. What machine learning category would this fall under?

- Classification
- Regression
- Cluster analysis
- Association analysis ✔

Answer
Correct: Video: "Categories of Machine Learning"

Submit

Multiple Choice

1/1 point (graded)
In general, are classification and regression often supervised or unsupervised approaches?

- (●) Supervised
- ( ) Unsupervised

✔

Answer
Correct: Video: "Categories of Machine Learning"

Submit

Multiple Choice

1/1 point (graded)
What is true between supervised and unsupervised approaches?

- ( ) In supervised approaches, the target is unavailable. In unsupervised approaches, the target is unavailable.
- ( ) In supervised approaches, the target is provided. In unsupervised approaches, the target is provided.

- In supervised approaches, the target is unavailable. In unsupervised approaches, the target is provided.
- In supervised approaches, the target is provided. In unsupervised approaches, the target is unavailable. ✔

Answer
Correct: Video: "Categories of Machine Learning"

Submit

Multiple Choice

1/1 point (graded)
What is the correct word to describe an instance of an entity in your data?

- Sample ✔
- Feature
- Attribute
- Field

Answer
Correct: Video: "Terminology Related to Machine Learning"

Submit

Multiple Choice

1/1 point (graded)
Is age group a numeric or a categorical variable?

- ○ Numeric
- ✔ Categorical

Answer
Correct: Video: "Terminology Related to Machine Learning"

Submit

Checkboxes

1/1 point (graded)
For a classification problem, if you want to predict the letter grade that a student would receive, what are 2 examples of reasonable input data to consider?

- ☑ Amount of time spent studying
- ☑ Percentage grade these students received in the previous semester
- ☐ Letter grade different students received in another class
- ☐ The students' ID numbers

Answer
Correct:
Video: "Classification"
Video: "Classification"
Video: "Classification"
Video: "Classification"

Submit

Checkboxes

1/1 point (graded)
What 2 statements describe classification in the context of machine learning?

- [x] Predict the category of the target given input data
- [x] Supervised task
- [ ] Unsupervised task
- [ ] Numerical target variable

✔

Answer
Correct:
Video: "Classification"
Video: "Classification"
Video: "Classification"
Video: "Classification"

Submit

True or False

1/1 point (graded)
The target variable is always categorical in classification.

- (•) True

False

Answer
Correct: Video: "Classification"

Submit

Multiple Choice

1/1 point (graded)
In building a machine learning model, why do we want to adjust the parameters?

To reduce the model's error

To compare different model variations

To provide the best graph of the model outputs

Answer
Correct: Video: "Building and Applying a Classification Model"

Submit

Multiple Choice

1/1 point (graded)
What is the next step in building a classification model after the model is constructed and parameters are adjusted?

Apply model to new data

Train the data

Minimize errors

✔

Answer
Correct: Video: "Building and Applying a Classification Model"

Submit

True or False

1/1 point (graded)
Test data is the same dataset as training data in classification models.

True

False

✔

Answer
Correct: Video: "Building and Applying a Classification Model"

Submit

Multiple Choice

1/1 point (graded)
Which algorithm to build classification models relies on the notion that samples with similar characteristics likely belong to the same class?

kNN

- Decision Tree
- Naive Bayes

✔

Answer
Correct: Video: "Building and Applying a Classification Model"

Submit

Multiple Choice

1/1 point (graded)
In a decision tree, which nodes do NOT have test conditions?

- Root nodes
- Internal nodes
- Leaf nodes

✔

Answer
Correct: Video: "Decision Trees"

Submit

Multiple Choice

1/1 point (graded)

How do you determine the size of a decision tree?

- The number of edges from the root node to that node.
- The number of edges in the longest path from the root node to the leaf node
- ✔ The number of nodes in the tree

Answer
Correct: Video: "Decision Trees"

Submit

Multiple Choice

1/1 point (graded)
What is the first step in constructing a decision tree?

- ✔ Start with all samples at a node.
- Partition the samples into subsets based on the input variables.
- Repeatedly partition data into successively purer subsets until stopping criteria are satisfied.

Answer
Correct: Video: "Decision Trees"

Submit

True or False

1/1 point (graded)

It works out better mathematically to measure the impurity of a split in a decision tree, rather than the purity.

- (●) True
- ( ) False

✔

Answer

Correct: Video: "Decision Trees"

Submit

Multiple Choice

1/1 point (graded)

Why are decision boundaries of a decision tree parallel to the axes formed by the variables?

- (●) Each split considers only a single variable
- ( ) Each subset should be as homogenous as possible
- ( ) The induction algorithm eventually stops expanding

✔

Answer

Correct: Video: "Decision Trees"

Submit

Multiple Choice

1/1 point (graded)

What is the command to get the number of rows in a data set titled “data”?

- (●) data.shape[0]
- ( ) data.shape[1]
- ( ) data.size()
- ( ) data.length()

✔

Answer

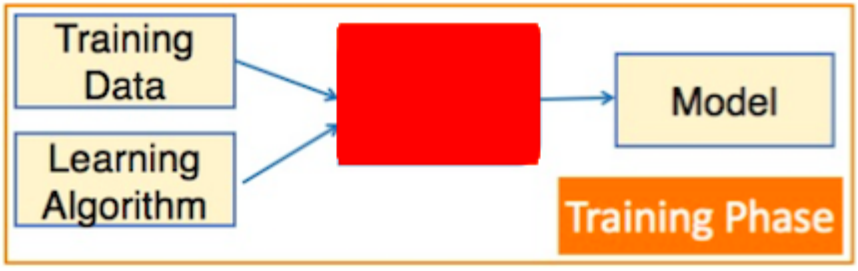
Correct: Video: "Decision Trees"

Submit

Multiple Choice

1/1 point (graded)

Which word goes in the red box as part of the training phase?



- (●) Build model
- ( ) Test data

Apply model

Results

✔

Answer
Correct: Video: "Decision Trees"

Submit

Multiple Choice

1/1 point (graded)
What does the following method call return?

```
accuracy_score(data_true = data_test, data_pred = predictions)
```

The fraction of correctly classified samples.

The number of correctly classified samples.

✔

Answer
Correct: Video: "Decision Trees"

Submit

Checkboxes

1/1 point (graded)
To use scikit-learn: DecisionTreeRegressor, train_test_split, and mean_squared_error, which of the following libraries are necessary? (Choose the

best two)

- [x] pandas
- [ ] sklearn.metrics
- [ ] sklearn.model_selection
- [ ] sklearn.tree
- [x] scikitlearn

✔

Answer
Correct:
Video: "Decision Trees"
Video: "Decision Trees"
Video: "Decision Trees"
Video: "Decision Trees"
Video: "Decision Trees"

Submit

Multiple Choice

1/1 point (graded)
What is the function call to output the name of columns of a dataframe named x?

- ( ) x.columns(0)
- (•) x.columns
- ( ) columns(x)

✔

Answer
Correct: Video: "Decision Trees"

Submit

True or False

1/1 point (graded)
True or False: The function call train_test_split(a, b) where a and b are dataframes will always output the same result.

- ○ True
- ● False

✔

Answer
Correct: Video: "Decision Trees"

Submit

Multiple Choice

1/1 point (graded)
Which is NOT mentioned in the course as a common similarity measure in cluster analysis?

- ○ Euclidean distance
- ○ Manhattan distance

Cosine similarity

Sine similarity

Answer
Correct: Video: "Clustering"

Submit

True or False

1/1 point (graded)
Cluster analysis is a supervised task.

True

False

Answer
Correct: Video: "Clustering"

Submit

Multiple Choice

1/1 point (graded)
How would you initially handle an anomaly (apparent outlier) in cluster analysis?

Throw it out of the dataset

- Disregard in further analysis
- Provide further analysis on the anomaly ✔

Answer
Correct: Video: "Clustering"

Submit

Multiple Choice

1/1 point (graded)
How do you assign each sample in a dataset to a centroid using the k-means algorithm?

- Assign the sample to the cluster with the closest centroid. ✔
- Assign the sample to the cluster with the furthest centroid.
- Assign the sample to a random cluster.

Answer
Correct: Video: "k-Means Clustering"

Submit

Multiple Choice

1/1 point (graded)

How do you determine the new centroid of a cluster?

- (●) Calculate the mean of the cluster
- ( ) Calculate the max of the cluster
- ( ) Calculate the mode of the cluster
- ( ) Calculate the min of the cluster

✔

Answer
Correct: Video: "k-Means Clustering"

Submit

Multiple Choice

1/1 point (graded)
What does the “within-cluster sum of squared error” provide?

- (●) A mathematical measure of the variation within a cluster.
- ( ) An error measurement for a specific sample in relation to the centroid of a particular cluster.
- ( ) An answer to which cluster is the most ‘correct.’

✔

Answer
Correct: Video: "k-Means Clustering"

Submit

True or False

1/1 point (graded)
Final clusters are sensitive to initial centroids.

- (•) True
- ( ) False

✔

Answer
Correct: Video: "k-Means Clustering"

Submit

Multiple Choice

1/1 point (graded)
Which parameter in the KMeans clustering algorithm do you have to specify for the number of clusters you want?

- (•) n_clusters
- ( ) clusters
- ( ) tot
- ( ) cluster_centers

✔

Answer
Correct: Video: "Clustering"

Submit

Multiple Choice

1/1 point (graded)

In the parallel_plot function, what was represented on the y-axis of the resulting plot?

- Each of the features
- Values of each cluster center
- Location of each cluster center
- Min and max values in each cluster

Answer

Correct: Video: "Clustering"

Submit

Multiple Choice

1/1 point (graded)

You are given a dataframe labeled x where the column 'number' indicates the index of a record. Which function call would create a new dataframe y that takes more than 10 samples x if x has 100 records?

- y = x[(x['number']%5)==0]

y = x[(x['number']%10)==0]

y = x[(x['number']%15)==0]

✔

Answer
Correct: Video: "Clustering"

Submit

Multiple Choice

1/1 point (graded)
What type of object does the function Kmeans output?

kmeans

dataframe

integer

series

✔

Answer
Correct: Video: "Clustering"

Submit

Multiple Choice

1/1 point (graded)
What is the difference between regression and classification for machine learning in

What is the difference between regression and classification for machine learning in Python?

- Regression transforms categorical values to numeric and then follows the same as classification.
- Regression is used to predict a numeric value while classification is used to predict a categorical value.
- Classification is used when the input data is categorical and regression is used when the input data is numeric.

✔

Answer
Correct: Video: "Regression Analysis"

Submit

Multiple Choice

1/1 point (graded)
For example, you want to predict the number of kids someone will have: either 0, 1, 2, or 3+. Is this an example of regression or classification?

- Regression
- Classification

✔

Answer
Correct: Video: "Regression Analysis"

Submit

True or False

1/1 point (graded)

Regression is an unsupervised task.

- True
- False ✔

Answer

Correct: Video: "Regression Analysis"

Submit

Multiple Choice

1/1 point (graded)

Which of the following is true about a model?

- built using test data
- evaluated on training data
- trained by the training data set ✔

Answer

Correct: Video: "Regression Analysis"

Submit

Multiple Choice

1/1 point (graded)

When is a prediction task referred to as simple linear regression?

- (•) When there is only one input variable.
- ( ) When there is more than one input variable.
- ( ) When there are two input variables.

✔

Answer

Correct: Video: "Linear Regression"

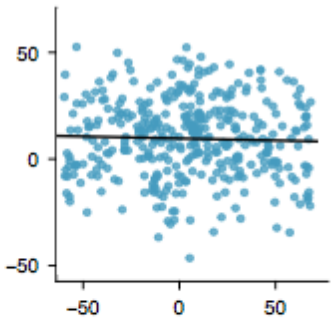
Submit

Multiple Choice

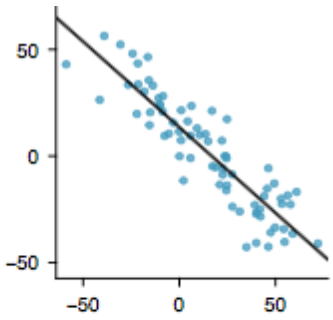
1/1 point (graded)

Which of the following graphs do you think is most appropriate for a simple linear regression model if you wanted to predict future values?

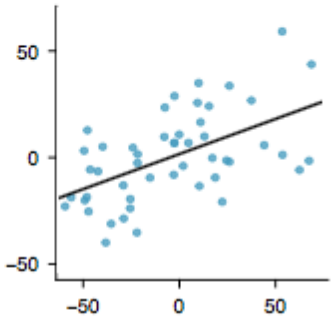
Graph 1:



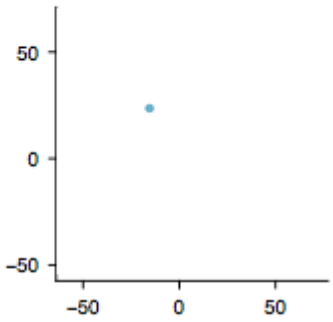
Graph 3:



Graph 2:



Graph 4:



- ○ Graph 1
- ○ Graph 2
- ◉ Graph 3
- ○ Graph 4

✔

Answer
Correct: Video: "Linear Regression"

Submit

Multiple Choice

1/1 point (graded)
What is the appropriate input for the following line of code to make a linear regression prediction?

```
y_prediction = regressor.predict(___)
```

- (●) x_test
- ( ) x_train
- ( ) y_train
- ( ) y_test

✔

Answer
Correct: Video: "Linear Regression"

Submit

Multiple Choice

1/1 point (graded)
Which Root Mean Square Error (RMSE) would represent a perfect prediction with no errors in regression?

- (●) 0
- ( ) NaN
- ( ) 1
- ( ) -1

✔

Answer
Correct: Video: "Linear Regression"

Submit

True or False

1/1 point (graded)
A Root Mean Square Error (RMSE) higher than our mean value would be too high.

- (•) True
- ( ) False

✔

Answer
Correct: Video: "Linear Regression"

Submit