Quiz 1

Saturday, April 8, 2023 2:23 AM

Back **Module 1 Quiz**
Graded Quiz • 20 min • 11 total points

1. Select the option that correctly completes the sentence: **1 point**

Training a model using labeled data and using this model to predict the labels for new data is known as _____.

- ○ Unsupervised Learning
- ◉ Supervised Learning
- ○ Density Estimation
- ○ Clustering

2. Select the option that correctly completes the sentence: **1 point**

Modeling the features of an unlabeled dataset to find hidden structure is known as _____.

- ○ Supervised Learning
- ○ Classification
- ○ Regression
- ◉ Unsupervised Learning

3. Select the option that correctly completes the sentence: **1 point**

Training a model using categorically labelled data to predict labels for new data is known as _____.

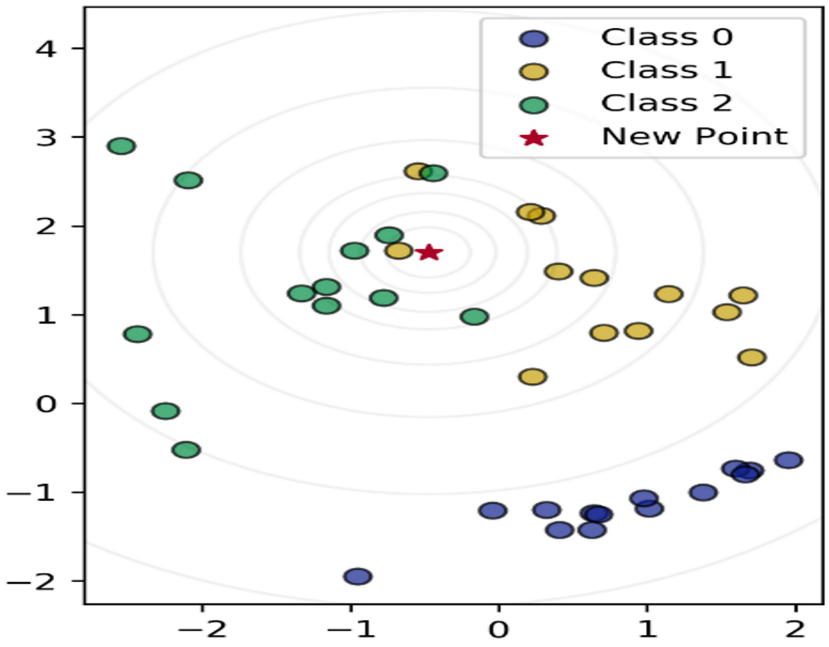
- ○ Clustering
- ◉ Classification
- ○ Regression
- ○ Feature Extraction

4. Select the option that correctly completes the sentence: **1 point**

Training a model using labelled data where the labels are continuous quantities to predict labels for new data is known as _____.

- ◉ Regression
- ○ Feature Extraction
- ○ Clustering
- ○ Classification

5. Using the data for classes 0, 1, and 2 plotted below, what class would a KNeighborsClassifier classify the new point as for k = 1 and k = 3? 1 point



- ○ • k=1: Class 0
  • k=3: Class 1
- ○ • k=1: Class 0
  • k=3: Class 2
- ◉ • k=1: Class 1
  • k=3: Class 2
- ○ • k=1: Class 1
  • k=3: Class 0
- ○ • k=1: Class 2
  • k=3: Class 1

6. Which of the following would *not* be something you specify for a nearest neighbor classifier algorithm? 1 point

- ☐ A method for pooling the classes of neighbor points to make a final classification decision
- ☐ An optional weighting function on the neighbor points
- ☑ The number of initial clusters to create from the training set
- ☐ How many nearest neighbors to examine
- ☐ A distance metric that finds neighbor points in the training set

7. Why is it important to examine your dataset as a first step in applying machine learning? (Select all that apply): 1 point

- ☑ See what type of cleaning or preprocessing still needs to be done
- ☑ You might notice missing data
- ☑ Gain insight on what machine learning model might be appropriate, if any
- ☑ Get a sense for how difficult the problem might be
- ☐ It is not important

8. The key purpose of splitting the dataset into training and test sets is: **1 point**

- (●) To estimate how well the learned model will generalize to new data
- ( ) To reduce the amount of labelled data needed for evaluating classifier accuracy
- ( ) To speed up the training process
- ( ) To reduce the number of features we need to consider as input to the learning algorithm

10. Given a dataset with 10,000 observations and 50 features plus one label, what would be the dimensions of X_train, y_train, X_test, and y_test? Assume a train/test split of 75%/25%. **1 point**

- ( )
  - X_train: (10000, 50)
  - y_train: (10000,)
  - X_test: (10000, 50)
  - y_test: (10000,)
- (●)
  - X_train: (7500, 50)
  - y_train: (7500,)
  - X_test: (2500, 50)
  - y_test: (2500,)
- ( )
  - X_train: (10000, 28)
  - y_train: (10000,)
  - X_test: (10000, 12)
  - y_test: (10000,)
- ( )
  - X_train: (2500, 50)
  - y_train: (2500,)
  - X_test: (7500, 50)
  - y_test: (7500,)
- ( )
  - X_train: (2500,)
  - y_train: (2500, 50)
  - X_test: (7500,)
  - y_test: (7500, 50)