### Back

# **Module 1 Quiz**

Quiz, 10 questions

Supervised Learning

```
Regression
Question 3
point
3. Question 3
Select the option that correctly completes the sentence:
Training a model using categorically labelled data to predict labels for new data is known
Classification
Regression
Feature Extraction
Clustering
Question 4
point
4. Question 4
Select the option that correctly completes the sentence:
Training a model using labelled data where the labels are continuous quantities to predict
labels for new data is known as _____.
Feature Extraction
Regression
Classification
Clustering
Question 5
```

#### 5. Question 5

point

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?

```
k=1: Class 1
k=3: Class 0
k=1: Class 2
k=3: Class 1
k=1: Class 0
k=3: Class 2
k=1: Class 1
k=3: Class 2
k=3: Class 2
k=1: Class 1
```

## 6. Question 6

point

Which of the following is true for the nearest neighbor classifier (Select all that apply):

#### Memorizes the entire training set

Given a data instance to classify, computes the probability of each possible class using a statistical model of the input features

Partitions observations into k clusters where each observation belongs to the cluster with the nearest mean

A higher value of k leads to a more complex decision boundary

```
Question 7
1
point
```

### 7. Question 7

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

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

```
Question 8
1
point
```

#### 8. Question 8

The key purpose of splitting the dataset into training and test sets is:

To estimate how well the learned model will generalize to new data

To reduce the number of features we need to consider as input to the learning algorithm

To reduce the amount of labelled data needed for evaluating classifier accuracy

To speed up the training process

```
Question 9
1
point
```

#### 9. Question 9

The purpose of setting the random\_state parameter in train\_test\_split is: (Select all that apply)

To avoid predictable splitting of the data

To make experiments easily reproducible by always using the same partitioning of the data

To avoid bias in data splitting

To split the data into similar subsets so that bias is not introduced into the final results

```
Question 10
1
point
```

### 10. Question 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%.

X\_train: (2500,)
y\_train: (2500, 50)
X\_test: (7500,)
y\_test: (7500, 50)
X\_train: (7500, 50)

- y\_train: (7500,)
- X\_test: (2500, 50)
- y\_test: (2500,)
- X\_train: (2500, 50)
- y\_train: (2500, )
- X\_test: (7500, 50)
- y\_test: (7500,)
- X\_train: (10000, 50)
- y\_train: (10000,)
- X\_test: (10000, 50)
- y\_test: (10000,)
- X\_train: (10000, 28)
- y\_train: (10000, )
- X\_test: (10000, 12)
- y\_test: (10000,)
- I, **HIN-WENG WAN**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account. Learn more about Coursera's Honor Code

Submit Quiz Thumbs Up Thumbs Down Square Flag

#### **Confirm Navigation**

Are you sure you want to leave this page?
Stay on this Page Leave this Page
Confirm Navigation

Are you sure you want to leave this page? Stay on this Page Leave this Page