Congratulations! You passed!
TO PASS 80% or higher

Keep Learning

GRADE
100%

Module 4 Quiz

LATEST SUBMISSION GRADE
100%

1. Machine learning is suited to solve which of the following tasks? (Select all that apply.) — 1 / 1 point

- [x] Fraud Detection

Correct
Web traffic data can be used to predict fraudulent transaction.

- [x] Churn Analysis

Correct
Churn analysis predicts customer engagement.

- [x] A/B Testing

Correct
Machine learning optimizes between different versions of websites or emails.

- [x] Natural Language Processing

Correct
Natural Language Processing applies statistical methods to language.

- [x] Financial Forecasting

Correct
Time series analyses are commonly applied to financial data.

- [ ] Reporting
- [x] Image Recognition

Correct
Neural networks are effective for operating on image data.

2. Is a model that is 95% accurate at predicting breast cancer a good model? — 1 / 1 point

- ( ) Likely no because there are too many false positives
- ( ) Likely yes because it accounts for false negatives and we'd want to make sure we catch every case of cancer
- ( ) Likely yes because this is generally a high score
- (•) Likely no because there are not many cases of cancer in a general population

Correct
There are many ways of quantifying the success of a classification task. Accuracy (which is technically the true positive and true negative rates over the total observations) might not capture false negatives.

3. What is an appropriate baseline model to compare a machine learning solution to? — 1 / 1 point

- (•) The average of the dataset
- ( ) Zero
- ( ) The minimum value of the dataset

Correct
Predicting the average is a good benchmark to try to improve upon using maching learning.

4. What is Machine Learning? (Select all that apply.) — 1 / 1 point

- [x] Learning patterns in your data without being explicitly programmed

Correct
Machine learning uses linear algebra and calculus to learn patterns in data without being explicitly programmed.

- [ ] Statistical moments calculated against a dataset
- [ ] Hand-coded logic
- [x] A function that maps features to an output

Correct
Machine learning maps input features to an output.

5. (Fill in the blanks with the appropriate answer below.) **1 / 1 point**

Predicting whether a website user is fraudulent or not is an example of _____ machine learning. It is a _____ task.

- (•) supervised, classification
- ( ) supervised, regression
- ( ) unsupervised, regression
- ( ) unsupervised, classification

Correct
In this case, whether the user is fraudulent is the dependent variable and we are classifying fraudulent from non-fraudulent users.

6. (Fill in the blanks with the appropriate answer below.) **1 / 1 point**

Grouping similar users together based on past activity is an example of _____ machine learning. It is a _____ task.

- (•) unsupervised, clustering
- ( ) supervised, clustering
- ( ) supervised, classification
- ( ) unsupervised, classification

Correct
In this case, there is no clear dependent variable so it is an unsupervised problem involving clustering users.

7. Predicting the next quarter of a company's earnings is an example of... **1 / 1 point**

- ( ) Reinforcement
- ( ) Classification
- (•) Regression
- ( ) Clustering
- ( ) Semi-supervised

Correct
The value we're predicting is a continuous, theoretically unbounded value.

8. Why do we want to perform a train/test split before we train a machine learning model? (Select all that apply.) **1 / 1 point**

- [ ] To give us subsets of our data so we can compare a model trained on one versus the model trained on the other
- [x] To evaluate how our model performs on unseen data

Correct
We care about how the model performs on data it hasn't seen before.

- [ ] To calculate a baseline model
- [x] To keep the model from "overfitting" where it memorizes the data it has seen

Correct
Overfitting is where a model does not generalize well to unseen data.

9. What is a linear regression model learning about your data? **1 / 1 point**

- ○ The average of the data
- ○ The best split points in a decision tree
- ◉ The formula for the line of best fit
- ○ The value of the closest points to the one you're trying to predict

✓ **Correct**
Linear regression learns the coefficients (or formula) of best line through the data.

10. How do you define a custom function not already part of core Spark? **1 / 1 point**

- ○ By extending the open source code base
- ○ You can't write your own functions in Spark
- ◉ With a User-Defined Function

✓ **Correct**
A User-Defined Function (or UDF) allows you to extend core Spark.