

Final Exam | Coursera

coursera.org/learn/machine-learning-with-python/assignment-submission/0tMDt/final-exam/view-feedback



Final Exam

Assignment details

Due

February 10, 10:59 PM WIBFeb 10, 10:59 PM WIB

Submitted

February 3, 12:15 PM WIBFeb 3, 12:15 PM WIB

Attempts

2 left (3 attempts every 24 hours)

Your grade

To pass you need at least 70%. We keep your highest score.

93.33%

Final Exam

Graded Assignment • 45 min

Your grade: 93.33%

Your latest: 93.33%•

Your highest: 93.33%•

To pass you need at least 70%. We keep your highest score.

[Next item](#) →

Almost there! Let's review your answers and then try again.

Your work showed skill in understanding various machine learning concepts and techniques, but there is room for growth in evaluating clustering algorithms. Before retrying this assignment, we recommend focusing on this key area:

Evaluating Unsupervised Learning Models: Heuristics and Techniques: In Question 15, the metric chosen to determine the optimum number of clusters was not the most suitable. Try reviewing [this item](#) to refresh your memory on the appropriate metrics for evaluating clustering algorithms.

Good job on the rest of the assignment!

1.

Question 1

How does the one-versus-one classification strategy work?



Correct

One-versus-one classification involves training a classifier for each pair of classes, allowing the model to decide between two classes at a time.

1 / 1 point

2.

Question 2

You are tasked with building a regression tree to predict the sales prices of cars in a used car marketplace. The data includes a few high-end luxury vehicles that significantly skew the distribution of car prices. Upon evaluating the model, you observe that the predictions for mid-range cars are more accurate, but the predictions for high-end cars are off.

In this case, why would using the median instead of the mean to calculate predictions at the leaf nodes be beneficial?



Correct

The median is less sensitive to the influence of outliers and skewed data, making it a better choice for this scenario.

1 / 1 point

3.

Question 3

You're building a predictive model for stock prices, starting with a simple linear regression. After evaluating the model's performance, you decide to experiment with more complex models, such as polynomial regression and random forests. As the complexity increases, you notice changes in your model's performance metrics.

What happens to bias and variance as you make the model more complex?



Correct

As you make the model more complex, it can capture more details (decreasing bias), but it also becomes more sensitive to small changes in the data (increasing variance).

1 / 1 point

4.

Question 4

You are analyzing healthcare data to ensure accuracy and identify any unusual trends that might indicate potential errors or fraudulent activity.

Which machine learning task would help achieve this?



Correct

This task uses anomaly detection to identify patterns that deviate from the expected norms.

1 / 1 point

5.

Question 5

You are monitoring a machine learning model deployed to predict customer churn for a telecom company. After some time in production, you notice that the model's performance has degraded, and the predictions have become less accurate.

At which stage of the Machine Learning Model Lifecycle would you most likely identify these issues?



Correct

Production-related issues typically surface once the model is deployed and actively making predictions, highlighting the need to revisit earlier stages.

1 / 1 point

6.

Question 6

Which of the following models is best suited for binary classification when the decision boundary is based on the proximity of a data point to its neighbors?



Correct

KNN (K-Nearest Neighbors) is particularly useful when classification depends on proximity. The model classifies a data point based on its closest neighbors, making it an effective choice for tasks where decision boundaries are influenced by local similarity.

1 / 1 point

7.

Question 7

You are creating a model to predict a student's final grade in a course based on study hours per week and class attendance. Since the final grade is a continuous variable influenced by multiple factors, you need a regression method that considers both predictors.

Which type of regression is most appropriate for this case?



Correct

Multiple linear regression allows you to use more than one predictor, such as study hours and class attendance, to predict a continuous outcome like a final grade.

1 / 1 point

8.

Question 8

When performing logistic regression, which factor determines the classification of an observation?



Correct

Logistic regression assigns classes based on a threshold probability to distinguish between classes.

1 / 1 point

9.

Question 9

A researcher is investigating how seasonal rainfall impacts the distance a butterfly species migrates. They have data on average rainfall and migration distance for each season. The data suggests a curved trend between the two variables. Which regression technique is most appropriate for analyzing this relationship?



Correct

Polynomial regression models non-linear relationships, such as the curved pattern in this data.

1 / 1 point

10.

Question 10

A marketing analyst wants to segment customers into distinct groups based on demographic information, but they don't have prior knowledge of the number of clusters. They prefer a method that starts with all customers in one large group and splits them into smaller, distinct clusters. Which clustering technique should they consider?



Correct

Divisive clustering uses a top-down approach, beginning with one large cluster and dividing it into smaller clusters.

1 / 1 point

11.

Question 11

What is one of the main advantages of using HDBSCAN over DBSCAN?



Correct

HDBSCAN improves on DBSCAN by creating a hierarchy of clusters of varying densities and extracts the most stable ones.

1 / 1 point

12.

Question 12

If a clustering algorithm requires preserving both local and global structures of complex, high-dimensional data, which dimensionality reduction method would be the best choice?



Correct

UMAP is designed to preserve both local and global structures, making it ideal for clustering complex data.

1 / 1 point

13.

Question 13

A healthcare researcher is developing a model to predict if a patient has a certain disease. Since the treatment has significant side effects, they want to minimize the number of false positives (incorrectly predicting that a healthy patient has the disease). Which evaluation metric would be the most appropriate for this scenario?



Correct

Precision is effective when the cost of false positives is high, as it focuses on the proportion of true positives among all positive predictions, reducing false positives.

1 / 1 point

14.

Question 14

A data scientist is evaluating the performance of a regression model predicting the price of houses. They want a metric that measures the average magnitude of error in the model's predictions. Which metric would be most suitable?



Correct

MAE measures the average absolute difference between predicted and actual values, providing a straightforward assessment of prediction accuracy.

1 / 1 point

15.

Question 15

A researcher is comparing the results of two clustering algorithms. They want to determine the optimum number of clusters for their dataset. Which metric is most suitable for this task?



Incorrect

Review the Evaluating Unsupervised Learning Models: Heuristics and Techniques video.

0 / 1 point