IEMS5726B

Data Science in Practice (Summer 2024) Assignment 3

JIANG Xijie, 1155202866

1.

a. Answer:

It's a supervised learning and regression problem.

Mean Absolute Error (MAE). The MAE measures the average absolute difference between the predicted values and the actual target values. So if the MAE is lower, says the performance is better.

b. Answer:

i) Categorical Data. It is the data that belongs to several categories, which don't have a natural order or numerical value associated with it.

ii) Numerical Data. It is the data presents as a number, representing measurable quantities.

iii) Descriptive Data. It presents as text, provides descriptive information about the goods.

c. Answer:

No. He miss the procedure of repeating the training step for 10 times and calculating the average accuracy.

d. Answer:

Correlation analysis is used to explore the relationship between two or more variables.

Recommender system is an algorithmic approach that suggests relevant items (products, movies, music, etc.) to users based on their preferences, historical behavior, or similarity to other users.

The recommender systems of Netflix and TikTok are very famous and successful.

e. Answer:

Difficulty: Lack of Truth Labels: Unlike supervised learning tasks, there are no predefined class labels or ground truth to compare the predict value.

Two major considerations: within-cluster distance and between-cluster distance.