CSE362/CSE5058 Artificial Intelligence Project 2

Due: May _th, 2022 until 12pm

Prerequisites

Before you can start working on this assignment make sure that you have installed scikit-learn and pandas.

Project Definition

In this project, you are required to train a k-NN classifier to predict whether a customer makes a purchase given their gender, age, and salary features. Figure 1 illustrates some example outputs of given Python file

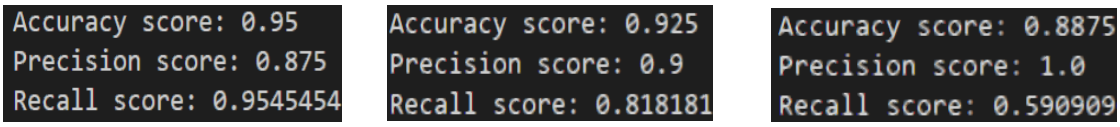

Figure 1 Performance metrics with n=10, 30 and 60 neighbors

To-Do

You are given a CSV file named *purchase_log.csv* that contains the necessary data for this assignment and a Python file with preprocessing steps such as loading the data from a file, encoding the *gender* feature to suitable format, splitting the data into training and test sets and feature scaling steps already completed.

Your task is to:

- Train a k-NN classifier with *Euclidean* distance and 10 nearest neighbors
- Make predictions with the test set
- Observe the performance metrics
- Repeat the experiment with 30 and 60 neighbors
- Explain the performance metrics we are using, how your performance metrics has changed and why.

Modify the Project2.py file to meet the requirements described above. You are **not** expected to implement your own k-NN algorithm, utilize the scikit-learn(sklearn) library to meet the requirements. Create a PDF report that describes your observations and explanations to requirements described above.

To-Submit

Only submit the Project2.py file and your report in PDF format. Do not upload the csv file given. Submitted file should be an archive (tar, zip, rar, etc.) named as your id-name-surname(e.g. 202051056016-berk-ercin.zip).