MACHINE LEARNING WORKSHOP

Questions:

1. Happy Singh bought a lot of apples and oranges for resale. He measured the weight and horizontal diameter of each fruit and noted them in a csv file. He bought a robot Sukimo who will go to market every week and buy the two fruits. help Happy Singh to train Sukimo to distinguish between apples and oranges. [Sukimo can measure weight and dimensions of any object
2. Vivek Rai is a very health conscious guy. He tracks the amount calories burnt, fat level, distance he walked, etc everyday using Samsung Fitbit device and app. Vivek decided to build an app that will predict the number of steps walked without using fitbit device, instead he will use some of the data that fitbit uses for generating health stats like calorie, distance walked and speed of walking, etc. He has his own fitbit data in the form of a csv file. Help Vivek to make an AI for his app.
3. There are many cab drivers in USA. The US Police wants to categorize the drivers into 4 groups based on their driving distance and drive speed. The data of 4000 drivers are recorded in a csv. Use machine learning technique to help the cops.
4. Your neighbor is a real estate agent and wants some help predicting housing prices for regions in the USA. It would be great if you could somehow create a model for her that allows her to put in a few features of a house and returns back an estimate of what the house would sell for.

   She has asked you if you could help her out with your new data science skills. You say yes, and decide that Linear Regression might be a good path to solve this problem!

   Your neighbor then gives you some information about a bunch of houses in regions of the United States,it is all in the data set: USA_Housing.csv.

   The data contains the following columns:

   - 'Avg. Area Income': Avg. Income of residents of the city house is located in.
   - 'Avg. Area House Age': Avg Age of Houses in same city
   - 'Avg. Area Number of Rooms': Avg Number of Rooms for Houses in same city
   - 'Avg. Area Number of Bedrooms': Avg Number of Bedrooms for Houses in same city
   - 'Area Population': Population of city house is located in
   - 'Price': Price that the house sold at
   - 'Address': Address for the house

5. The world remembers the Titanic disaster where many people lost their lives. But there are many others who survived the disaster. We have collected data about a lot of passengers in an excel file (titanic.xls). Suppose you were one of the passengers in the ship.Your Port of Embarkation was 'S', ticket fare 200, Passenger Class was 3 and your sex and age were your actual sex and current age. What could have been your chances of survival?

General instructions:

1. Try to solve each problem using the partially solved jupyter notebooks mailed to you all.
2. First ask yourself which ML model can be used to solve each problem.
3. Then find out the algorithm that is ideal for the given situation (Remember you learnt KNearestNeighbors, DecisionTree, KMeans and LinearRegression)
4. Follow this general approach:
    - Load dataset
    - extract features and labels [if present]
    - Split data into train and test
    - Select a model
    - Instantiate empty model [e.g.: model = KNeighborsClassifier()]
    - train the model [e.g.: model.fit(X_train, Y_train)]
    - Predict on testing data
    - check accuracy of prediction

Solution to all the problems can be found on our github repo

For any doubts, you can refer to AI-ML Seminar notes mailed to you all. If you still have doubts, feel free to contact us.

Happy learning!