CSCI 183 Homework 3

Programming Question : Heart Attack Prediction

What do you need to do?

1. Create and compile as many plots as you can using the matplotlib library for the given dataset.
2. Use the created plots to try to find some numerical features that can be used for good classification models (similar to the example in class find features that help in linearly separating data)
3. Split the dataset into 70% train and 30% test.
4. Implement the Logistic Regression classification algorithm
5. Use the evaluation metrics discussed in the class to see how well your models have performed for the given dataset.
6. Submit your code as an ipynb file and a document reporting your findings.

Observation Table:

For Logistic Regression

Features	Precision	Recall	Accuracy	F1-Score
:				
:				

Roadmap

1. Use as many visualization tools on the dataset. It is not necessary that you will get information from all of them, but if you are working on a dataset then any useful information you can have will be helpful in deciding which attributes are good/bad with respect to classification. Start with 1 attribute and try combinations with others to see if they are relevant attributes or not. Use this step to eliminate attributes instead of selecting [For example you might want to eliminate DoB from the attributes since it would not help for the classification task, but it would have age as an attribute then you might wanna use it for your classification task].
2. Once you have an idea about the attributes you want to use to build your model, use the sklearn library to build your model and then use the evaluation metrics to evaluate the performance of your model.

3. Our goal is to experiment with the dataset, not to get the most accurate model. (But try to get the best accuracy if possible :))

Dataset Description

This database contains 76 attributes, but all published experiments refer to using a subset of 13 (attributes) + 1 (target) of them.

The "target" field refers to the presence of heart disease in the patient. It is integer-valued 0 = no/less chance of heart attack and 1 = more chance of heart attack

Attribute Information

1) age

2) sex

3) chest pain type (4 values)

4) resting blood pressure

5) serum cholestoral in mg/dl

6)fasting blood sugar > 120 mg/dl

7) resting electrocardiographic results (values 0,1,2)

8) maximum heart rate achieved

9) exercise induced angina

10) oldpeak = ST depression induced by exercise relative to rest

11)the slope of the peak exercise ST segment

12) number of major vessels (0-3) colored by flourosopy

13) thal: 0 = normal; 1 = fixed defect; 2 = reversable defect

14) target: 0= less chance of heart attack 1= more chance of heart attack