Assignment 4

Applied Machine Learning

Fatemeh Saberi Khomami | 400422114 | fatima.saberi@gmail.com

Question 5.2

Assume you are asked to design a learning algorithm to predict whether patients are going to suffer a heart attack. Relevant patient features the algorithm may have access to include blood pressure (BP), body-mass index (BMI), age (A), level of physical activity (P), and income (I).
You have to choose between two algorithms; the first picks an axis aligned rectangle in the two dimensional space spanned by the features BP and BMI and the other picks an axis aligned rectangle in the five dimensional space spanned by all the preceding features.

5.2.1

1. Explain the pros and cons of each choice.

Answer:
The pros of a simpler model with only 2 feature are:

- it requires less data
- it is more robust to over-fitting

the cons of a simpler model with only 2 features are:

- it is more prone to under-fitting
- it is disregarding additional data that might be crucial and predictive

The pros of a more complicated model with all the 5 feature are:

- it is including additional data that might be crucial and predictive
- it is adding more flexibility to the hypothesis class

- it is more robust to under-fitting

The cons of a more complicated model with all the 5 feature are:

- it requires more data which may be hard to acquire
- it is more prone to over-fitting
- it is including additional data that might be irrelevant to the goal and only adds to our computational cost

5.2.2

2. Explain how the number of available labeled training samples will affect your choice.

Answer: If we have a small training set, we should go with the simpler model, because regarding our data, we will fail to build a complicated predictive model, and the learner will only memorise the data points instead of learning a predictive pattern.
But if we have a rich training set, it will be possible to extract a predictive model from vast number of features. Having a large data set for training will enable us to utilize the potential of all the provided information and features towards building a more accurate model. So, our choice of model complexity depends on the number of training examples.