Assign 2

Robert Joseph

February 17, 2021

Thought Question 1

Chapter 4
Subsection 4.1

Question

I assume that only Linear/Multiple Linear Regression has a Closed Form Analytic Solution compared to other Machine Learning Techniques. What are the disadvantages of using this rather than a numerical Approximation?

Solution

Linear Regression is represented by this (Multidimensional case)

$$\hat{\beta} = \arg\min \|X\beta - y\|_2$$

The closed form solution of Linear Regression is given by

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

For instance suggested by this closed form solution I think it would be computationally expensive to calculate this for very large Matrices. The time complexity for multiplication of matrices is in general $O(N^3)$ unless you use the faster algorithm and it goes down to $O(N^{2+\log_4(3)})$

Another reason why is that analytical solutions are strongly connected to the model, so implementing them can be inefficient if we plan to generalize/change the models in the future.

Finally for even Sparse Matrices this would be seem to be useless as well as the inverse of the matrix most of the entries would also have most of the entries as non-zero entries and it would be computationally expensive again to even store a large number of numbers for large N(size of the matrix).

Thought Question 2

Chapter 6
Subsection 6.2.2

Question
In the textbook its mentioned models obtained by directly estimating $p(y|x)$ are called discriminative models and models obtained by directly estimating $p(x|y)$ and $p(y)$ are called generative models. So is it possible to generate data from with the help of these models?

Solution
I would assume that a discriminative model would not be able to generate new data because does not know about dependencies between features, and it is irrelevant for prediction. It although is a superior model compared to a generative one.

Generative models on the other hand can generate data if there is sufficient data available. This is because I feel like they predict by modelling joint distributions plus generative models can actually learn the underlying structure of the data if we specify the model correctly. So I feel like this could work like just picking a random sample from this joint distribution similarly like in the way of a Generative Adversial Network or something.