ENPM808A: Introduction to Machine Learning

Mid-Term Exam

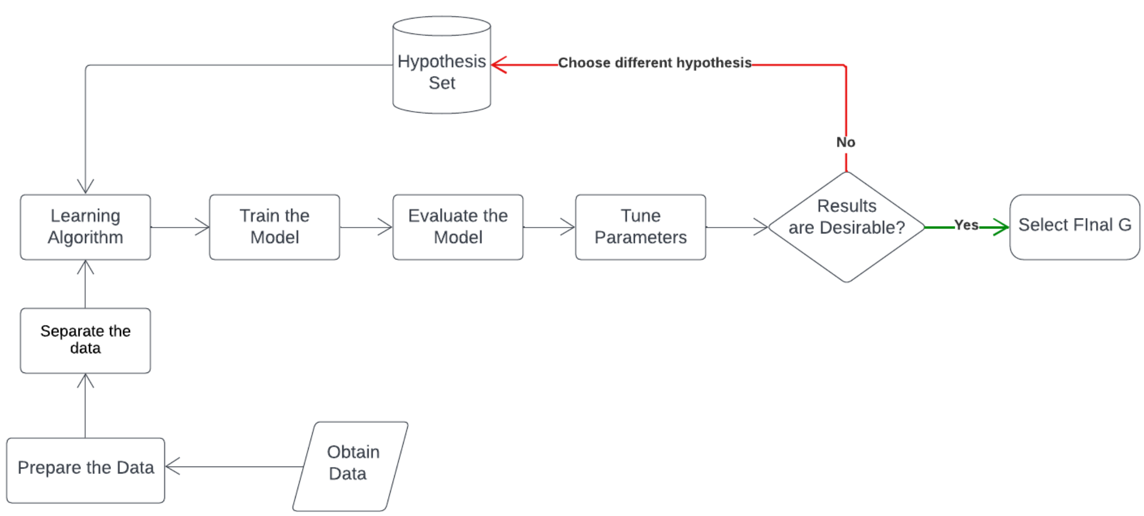
Problem 1

Answer:

Steps to implement a machine learning model:

- **Obtain the data** for implementation
- **Prepare the data** based on the desired result. (Data Cleaning and Filling in gaps if any Eg: Preprocessing of an image dataset)
- **Separate the data** into Training and Test sets used in further steps.
- **Choose Desired Features** in the data
- **Choose an ML Model/Learning Algorithm** based on the desired output.
- **Train the Model** on the Training Dataset, until we get a low in-sample error. (Fitting the Data with a hypothesis)
- **Evaluate the Model** using the Test Dataset (unseen data) and check its accuracy and performance.
- **Parameter Tuning** to improve the accuracy of the model or learning algorithm.
- Based on the results and less in sample error, we **select the final hypothesis** close to the target function.

Note: Keep Overfitting in mind and Data Snooping will lead to skewed results.



Problem 2

Answer: Option c

I think the best I would promise her is (c).

- The unknown target f can be highly complex, and learning might not be possible at all.
- But, if we can learn and produce a hypothesis g, since there are many data points (4000), the probability that g matches f is high according to Hoeffding inequality, and the Error(g) might be small as we have a large data set.

Problem 3

Please view the code for all values of all parts

Data separation → Linear regression → Predictions → E_in and E_out evaluations

Adding Weight Decay → Calculating Augmented Errors

d) According to my calculations k = -2 will give the least out-of-sample error

I have calculated the Errors using two methods

Method 1: I calculated the penalty value and added it to the E_in to calculate Augmented E_in

$$E_{\text{aug}}(h, \lambda, \Omega) = E_{\text{in}}(h) + \frac{\lambda}{N}\Omega(h).$$

Method 2: I have calculated the regularized weights and then calculated the new E_in

$$\mathbf{w}_{\text{reg}} = (\mathrm{Z}^{\mathrm{T}}\mathrm{Z} + \lambda \mathrm{I})^{-1}\mathrm{Z}^{\mathrm{T}}\mathbf{y}.$$

Problem 4

(a) There is a set of N points that can be shattered by H.

If there exists an H that can shatter a set of N points, then we can conclude that N might not be a breakpoint. Since VC Dimension can be no more than a breakpoint. In this case d_vc >=N

(b) Any set of N points can be shattered by H.

If there exists an H such that any set of N points can be shattered, we can certainly conclude that N is not a breakpoint of H. In this case d_vc > N.

(c) There is a set of N points that cannot be shattered by H.

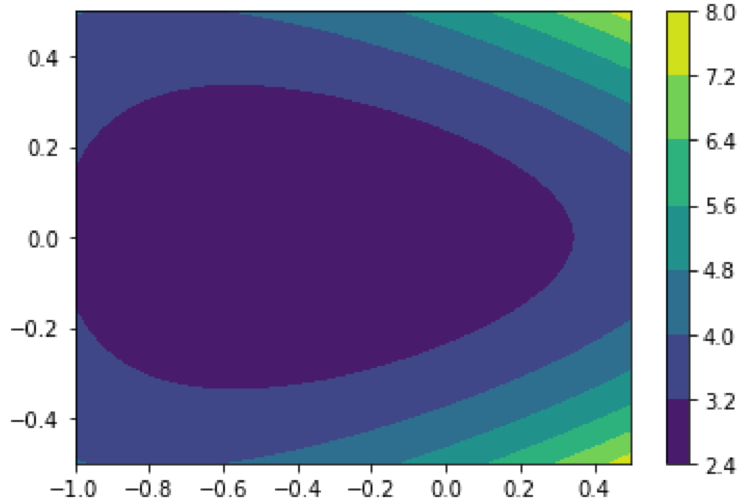
If there exists an H such that a set of N points cannot be shattered by N, it does not mean that N is the breakpoint. We cannot conclude that N is a breakpoint with insufficient information about all dichotomies possible. d_vc>=N

(d) No set of N points can be shattered by H.

If there exists an H such that any set of N points cannot be shattered, we can certainly conclude that N is a breakpoint of H. In this case d_vc <= N → d_vc = N-1.

Problem 5

a) Plotting the level sets of, F(x1, x2)



b) Calculating Gradient descent (check code)

c) Plotting the progressive convergence