

ENPM808A – Introduction to Machine Learning

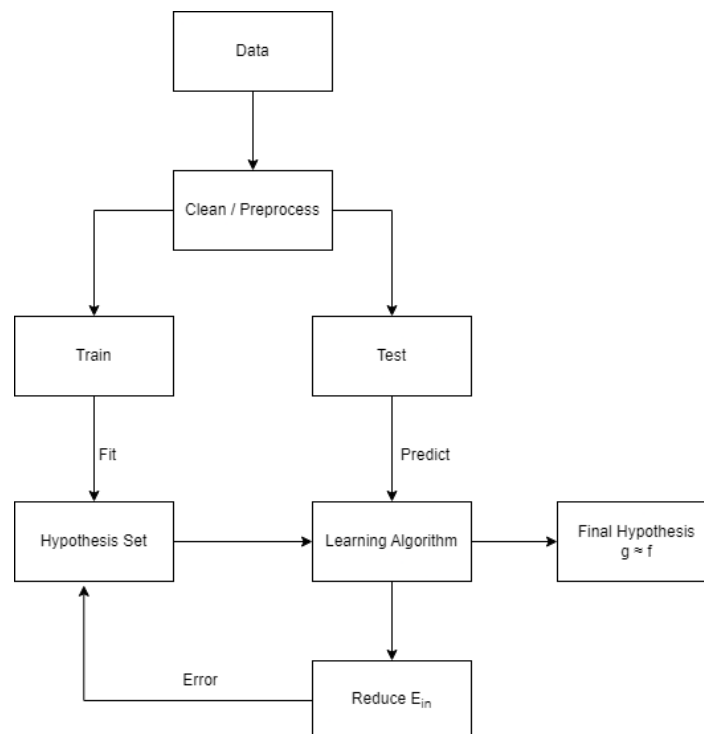
Mid-Term

Author: Syed Mohammad Ibrahim

Email: jamibi@umd.edu

1. Steps involved from getting the data to providing a final learned hypothesis are:

- Fetch the data
- Prepare the data
- Segregate the data into training set and test set
- Choosing or fitting the appropriate algorithm to take a formula “ $g: x \rightarrow y$ ”, the approximate target function “ f ”
- “ g ” is chosen to form a set of candidate formula, that is, hypothesis set H from the data D
- Next is training the data
- For efficient model, E_{in} must be lowest
- Finally, a prediction is made where hypothesis $g \approx f$



2. I would go with (c) option as the target function “ f ” is unknown and it can be very complex that could be impossible to learn at all. If at all, we can learn and produce a hypothesis “ g ” (since there are 4000

data points which is quantitatively high), the probability that “g” matches function “f” is high according to Hoeffding inequality, and the error on hypothesis “g” might be small since we have a large data set.

3.

4. a) The break point can be concluded as not N as the N points are shattered by the hypothesis H. Nothing conclusive can be said about the VC dimension of H. The only inference about d_{VC} is $d_{VC} > N$

b) Considering H can shatter any set of N points; N is surely not the break point of H, and it is an overkill. Again, nothing conclusive can be said about the VC dimension and the inference remains the same, that is d_{VC} is $d_{VC} > N$

c) Since only single set of N points cannot be shattered, we cannot say anything about the break point and VC dimension. It can be the case that there exists a different set of N points that H can shatter. This condition is insufficient to draw any conclusion on the break point. Thus, N may or may not be a break point

d) As H cannot shatter any set of N points, it can be concluded that N is the break point for H. It can be inferred for VC dimension that $d_{VC} \leq N$

5. Please see the question_5.py file for the code