Chapter - 2
Theory of Learning

Aviral Janveja

1 Can We Generalize?

Revisiting the machine learning process described in chapter 1 : We have an unknown target function f, which represents the underlying pattern that we would like to uncover. Next, we have a set of observations that will be used to approximate the unknown target function. Finally, our approximation of the target function, called the hypothesis function g, based on the sample of data that we have.

Our hypothesis function g may perform well on the available data points. However, remember that the goal in machine learning is not for g to perform well in-sample but for g to approximate f well, that is $g \approx f$. So, does our hypothesis function generalize well out of sample?

2 Answer : Probably, Approximately.

In order to answer the question raised above, first let us formalize the performance of our hypothesis function g in terms of its agreement or disagreement with f on the available data points.

For each point x in our data sample. If $g(\mathbf{x}) \neq f(\mathbf{x})$, that is, if our hypothesis disagrees with the given correct output, then it constitutes an **error**. Now, the sum-total of errors that our hypothesis function g makes in-sample, would then be called the in-sample error or E_{in}.

We try to minimize E_{in} as much as possible. But does a small E_{in} imply a small E_{out}, which is what actually matters ?

2. Does E_{in} track E_{out} well? Lower E_{out} means g approximates f well. Hence, E_{out} is what we care about. We use E_{in} to get a probabilistic bound on E_{out} via the Hoeffding inequality(from the law of large numbers in statistics. Adapted for our use-case in ML)

3. Well, it turns out we can say something about out-of-sample error E_{out} based on the in-sample error E_{in}, given that our sample is sufficiently large and we are ready to accept an approximation based on a tolerance value ϵ.

4. Intuitively, if the sample size is big, then it should help. If approximation is enough, that should help too. Finally M, which is the number of hypothesis, which is infinite for most relevant models. But this is not our final result in the theory of learning, we will deal with M going forward.

Intuitively. The probability of in sample and out of sample diverging will be low if you have reasonable error tolerance e and a lot of data points N. Model complexity denoted by M = the number of hypothesis.

3 References

1. CalTech Machine Learning Course - CS156, Lecture 2.