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 is based on the sample of data that we have.

What we have got until now is a function g that performs 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, how do we ensure that our hypothesis function generalizes well out of sample?

2 A Probabilistic Answer

circumventing the bin example and going directly with respect to the learning process. We can make a most probably statement. Rewriting it in terms of Ein and Eout.

The performance of our hypothesis g in sample can be formalized in terms of E_{in}. This is the error-rate in-sample or the number of data-points our hypothesis got wrong. E_{out} is what we care about, the error-rate out of sample. Low Ein means g approximates f well in sample. less Eout means g approximates f well out of sample as well. Hence, Eout is what we care about. We can use Ein to get a probabistic bound on Eout via the Hoeffding inequality.

Intuitively, if the large number of samples should help. Yes, then if error tolerance is not too strict, approximation is enough. Finally M, which is the number of possible hypothesis, which is infinite for most models. But we will deal with M going forward.

The question in the intro para can then be rephrased as, does E_{in} track E_{out} well?

The answer is yes, it is possible from our hypothesis function to approximate the target function in a way that is bound by probability. Called the Hoeffding inequality. from the law of large numbers in statistics that is adapted for our use-case in ML.

Hoeffding equation.

Let in sample performance of our hypothesis function be denoted by Ein which denotes the in sample error. Meaning the number if times the hypothesis function got the result wrong in the data set. Eout similarly. We ideally want Ein to track Eout or in sample performance to track out of sample performance. This is bound by Hoeffding inequality in terms of N = number of data points, epsilon = error tolerance and M = number of hypothesis.

Intuitively. The probability of in sample and out of sample diverging will be low if you have reasonable error tolerance and a lot of data points.

Model complexity and M = the number of hypothesis.

One chapter for all theory till NN.

3 References

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