CS 475 Machine Learning: Homework 1
Learning Foundations

Daniel Crankshaw

1 Analysis

1 Explain why you agree or disagree with the following statement.
It is always best to select a hypothesis class that contains the optimal hypothesis.

I disagree with the statement because sometimes the hypothesis class containing the optimal hypothesis may be too large to be practical. While the class may contain the optimal hypothesis, it could be the case that this makes the search space too large, and so we would never find the optimal hypothesis anyway. In this type of situation, it would be better to choose a hypothesis class that may not contain the optimal hypothesis but is small enough to effectively search through and contains a hypothesis better than what we are likely to find in the too large class containing the optimal hypothesis.

2 True or False: *An infinite hypothesis class always contains the optimal hypothesis.* If true, why? If false, give a counter example.

False. The counterexample given in class was the following. Say we are searching for strings and say the optimal hypothesis is a string that contains the letter 'z'. But the hypothesis class we are searching in has a limited alphabet that does not contain the letter z. Then we still have an infinite hypothesis class (there are an infinite number of strings of lowercase letters of the alphabet that do not contain the letter z) but our hypothesis class does not contain the optimal hypothesis.

3 Consider the following development scenario. A researcher collects 1000 labeled examples for learning, dividing them into a training set (500 examples) and a test set (500 examples). To develop a classifier, the researcher experiments by adding new features. To guide feature construction, the research tests each set of features by training on the train data and testing on the test data, measuring the resulting change in accuracy and only keeping features that help. When the researcher is finished, he collects 1000 new labeled examples and evaluates the classifier on these new examples.

Do you expect accuracy on these 1000 new examples to be the same, better or worse than the original 500 test examples? Why?

I expect accuracy on these new samples to be worse than the original test samples. Even though he is not using the original test set to actually train the algorithm, he is selecting for features based on the content of the test set. This is in a sense training the algorithm based on this test set, it is just a manual training (vs an algorithmic training). Therefore, when he tests the algorithm on the new examples, we expect worse accuracy than on the training data (which we can include the old test data in).

4 At the start of the semester, you arrive home to find three packages containing your three new textbooks. Each package was sent via Fedex or UPS with equal probability. Of the three packages, one of the packages is a brown box delivery by UPS. What is the probability that you have one UPS package and two Fedex packages? Why?

Given that one of the packages is from UPS, there are 4 possibilities for what the configuration of the remaining 2 packages could be:

$\{Fedex, Fedex\}, \{Fedex, UPS\}, \{UPS, Fedex\}, \{UPS, UPS\}$

Only one of those cases results in one UPS package and two Fedex packages. Therefore, the probability that we have one UPS package and two Fedex packages given that we know we have at least one UPS package is $\frac{1}{4} = 25\%$.

5 True/False (and why): Suppose you know your hypothesis class contains the optimal hypothesis, and you observe that changing any one part of your current hypothesis makes it worse than before. You can safely conclude that the current hypothesis must be optimal.

False. This is a confusion about the difference between local vs global optimality. The observation stated indicates that we have a locally optimal hypothesis. However, it could be that changing several parts of our hypothesis would - while first leading to a less optimal hypothesis - eventually find a more optimal hypothesis than our current one.