CS 6362 Machine Learning, Fall 2017: Homework 1

Venkataramana Nagarajan

Question 1: It is not always best to select a hypothesis class that contains the optimal hypothesis because the hypothesis class may be extremely large and it would just not be feasible to go through the entire class which may take extremely long or infinite time.

Question 2: False. For example, we may have an infinite hypothesis class containing infinite linear hypothesis, but the optimal hypothesis may be of a different class (say quadratic). In this case the infinite hypothesis class does not contain the optimal hypothesis.

Question 3: The accuracy on the 1000 new samples will be worse than the original 500 test samples. Here the researcher tries to adopt a feedback mechanism where he uses half of the training data to keep features that help. But by doing so, he overfits the test data and the predictor has learned some new information(noise) which will not be accurate on the new data.

Question 4: There are 4 events which are possible:

0 UPS 3 Fedex
1 UPS 2 Fedex
2 UPS 1 Fedex
3 UPS 0 Fedex

Now since it is given that one of the packets is from UPS our sample space has reduced from 4 to 3 as follows:

1 UPS 2 Fedex
2 UPS 1 Fedex
3 UPS 0 Fedex

Out of these 3 events there is only one event with 1 UPS packet and 2 Fedex packets which makes its probability 1/3.

Question 5: False. We cannot conclude that the current hypothesis must be optimal. It says that changing any one small part of the current hypothesis makes it worse than before, but it is

possible that we may be stuck in the local minimum. We could experience better behavior at some other point which could be the required optimal hypothesis.