CS 560 Statistical Machine Learning

Final Exam (Spring 2025)

Date – 5 May 2025

1. Discuss your understanding of a favorite subject (not necessarily in this course).

2. Describe a general problem that current machine learning techniques cannot be applied to solve a problem.

3. Active learning. Give an example to show that active learning can be used to exponentially reduce label complexity, and give another example to show that it cannot.

4. Online learning. Consider the online learning from experts model. Describe a scenario where it is possible to construct a learning algorithm whose performance is almost same as the best expert, but both of them suffer a large loss compared to the true labels.

5. Boosting confidence. Let D be the distribution on domain X, and h* € H be the target model. Therefore, any sample (x, y) that a learner receives satisfies the condition y = h* (x). Let d be the VCdimension of H. Suppose that Aw z is a weak PAC learner in the sense that for any given error rate € € (0, 1), it draws ne samples and outputs a hypothesis h such that err (h) ≤ e with probability 0.51. Since the success probability is a constant, Aw is not a PAC learner.

   Based on AwL, construct a new algorithm that is a PAC learner. That is, for any e, б € (0, 1), such algorithm outputs some h such that err (h) ≤ e with probability 1 -б. Hint: the success probability measures how lucky a learner is to get good training data. If the learner draws a fresh batch of ne samples (therefore, there are now two batches of samples), what will be the probability that at least one batch is good?