CSE446: Homework #4

Due: June 5, 2015

General Instructions

Please submit both your code and writeup online by 1:30pm PST on Friday, June 5, 2015. Your writeup should be a PDF document, and all code should be sufficiently commented.

1 VC Dimension

Mitchell 7.5a: Consider the space of instances X corresponding to all points in the xy plane. Give the VC dimension of the following hypothesis space: H_r = the set of all rectangles in the xy plane. That is, $H = \{((a < x < b) \wedge (c < y < d)) | a, b, c, d \in \mathbb{R}\}$. Justify your answer (you do not need to give a formal proof, but you should present the key ideas behind your reasoning).

2 PAC Learning

Mitchell 7.6: Write a consistent learner for H_r from Exercise 7.5. Generate a variety of target concept rectangles at random, corresponding to different rectangles in the plane. Generate random examples of each of these target concepts, based on a uniform distribution of instances within the rectangle from (0,0) to (100,100). Plot the generalization error as a function of the number of training examples, m. On the same graph, plot the theoretical relationship between ϵ and m for $\delta = 0.05$ *(Note the correction of the probable typo in the book, where $\delta = 0.95$).* Does theory fit experiment?

3 Bias and Variance

1. Read Pedro's ICML paper on bias-variance decomposition and check out the reference implementation for calculating average bias and net variance (defined in §2). In this problem we will be evaluating the bias and variance of your ID3 learner on a classification task using zero-one loss. Do not use pruning for this problem.

2. We will be using the SPECT Heart Data Set, which you can find more information about here. The first column of the data is a diagnosis (normal or abnormal), followed by 22 columns containing binary features (partial diagnoses computed from raw data).

3. Modify your ID3 learner to accept a parameter of maximum depth, beyond which no further splitting occurs. Do you think that the learner's bias will increase or decrease as this parameter increases/decreases? What about its variance? Why? Generate 25 bootstrap datasets using the method described in §4 of the ICML paper, and plot the average bias and net variance vs. maximum depth (varying the depth from 1-10).

4 SVMs

1. Download LIBSVM, currently the most widely used SVM implementation, and read the documentation to understand how to use it.

2. Download the new SPECT data set, which is the same as in Problem 3 except that it has been converted to LIBSVM format.

3. Run LIBSVM to classify heart conditions, using the kernels 0-3 and default parameters for everything else. How does the accuracy vary with different choices of kernel?

4. Use your bagged ID3 classifier from Homework 3 to classify heart conditions (using no pruning and no maximum depth). Use 25 bootstrap samples for the bags (note that you can use the same bootstrap samples from problem 3). How do your SVM accuracies compare with that of the bagged decision tree? Explain the difference (or lack of difference).

5. Sort the features by information-gain, and recompute the bagged ID3 and SVM accuracies using only the top k features, varying k from 1-10 (note that this will *not* necessarily produce the same trees as when varying the maximum depth from 1-10). How does the relative accuracy of bagged ID3 vs. SVMs vary with the number of features available? How do you explain this?