Problem Set 3

CS 6375

Due: 10/19/2017 by 11:59pm

Note: all answers should be accompanied by explanations for full credit. Late homeworks will not be accepted.

Problem 1: VC Dimension (35 pts)

1. Consider a binary classification problem for data points in $\mathbb{R}^3$ with a hypothesis space consisting of cylinders of a fixed radius $r > 0$ and fixed length $l > 0$ such that any point in the cylinder is labeled with a $+$ and any point outside the cylinder is labeled with a $-$. What is the VC dimension of this hypothesis space? Prove it.

   (a) How many samples would be sufficient to guarantee that an optimal learning algorithm will attain an accuracy of .8 with probability at least .95?

   (b) How does the VC dimension change if you also add cylinders of a fixed radius $r > 0$ and fixed length $l > 0$ such that any point in the cylinder is labeled with a $-$ and any point outside the cylinder is labeled with a $+$ to the above hypothesis space.

2. Consider a binary classification problem for data points in $\mathbb{R}^2$ with a hypothesis space consisting of a pair of axis aligned rectangles such that any point inside either rectangle is labeled with a $+$ and any point outside both rectangles is labeled with a $-$. What is the VC dimension of this hypothesis space? Prove it.

3. Consider a binary classification problem for data points in $\mathbb{R}^2$ with hypothesis space H consisting of exactly k distinct linear separators in $\mathbb{R}^2$. If $k = 3$ what is the largest and smallest that the VC dimension of H can be? Prove it by giving a specific H's whose VC dimensions match your upper and lower bounds.

Problem 2: Medical Diagnostics (65 pts)

For this problem, you will use the data set provided with this problem set. The data has been divided into two pieces heart_train.data and heart_test.data. These data sets were generated using the UCI SPECT heart data set (follow the link for information about the format of the data). Note that the class label is the first column in the data set.

1. Suppose that the hypothesis space consists of all decision trees with exactly three attribute splits (repetition along the same path is allowed) for this data set.

(a) Run the adaBoost algorithm with $M = 3$ to train a classifier for this data set. Draw the 3 selected trees in the order that they occur and report the ϵ, generated by adaBoost, for each.

(b) Run the adaBoost algorithm for 10 rounds of boosting. Plot the accuracy on the training and test sets versus iteration number.

2. Now, suppose that the hypothesis space consists of only height 1 decision trees for this data set.

(a) Use coordinate descent to minimize the exponential loss function over the training set. You can use any initialization and iteration order that you would like other than the one selected by adaBoost. What is the optimal value of α that you arrived at? What is the corresponding value of the exponential loss on the training set?

(b) What is the accuracy of the resulting classifier on the test data?

(c) What is the accuracy of adaBoost with $M = 20$ for this hypothesis space on the test data? How does the α learned by adaBoost compare to the one learned by gradient descent?