Problem Set 3

CS 7301

Due: 10/19/2016 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 axis aligned 3-d boxes such that any point in the box is labeled with a $+$ and any point outside the box is labeled with a $-$. What is the VC dimension of this hypothesis space? Prove it. How many samples would be sufficient to guarantee that an optimal learning algorithm will attain an accuracy of .8 with probability at least .95? Can you generalize your argument to axis aligned boxes in $\mathbb{R}^d$?

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 $\mathbb{R}^2$. Let H be the hypothesis space of all axis aligned rectangles in $\mathbb{R}^2$. Consider the boosted hypothesis space H' that takes a pair of hypotheses from H and takes the sign of their weighted combination (similar to what would be produced by two rounds of boosting). Specifically,

$$H' = \{f | f(x) = \text{sign}(\alpha_1 h_1(x) + \alpha_2 h_2(x)) \text{ for some } h_1, h_2 \in H \text{ and } \alpha_1, \alpha_2 \in \mathbb{R}\}.$$

   What is the $VC(H')$? Prove it.

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. Train a decision tree using the information gain heuristic to select attributes as described in class (break ties using a majority vote). What is the depth of the learned decision tree? What is the accuracy of your learned decision tree on the test set?

2. Suppose that the hypothesis space consists only of depth 2 decision trees for this data set.

   (a) Run the adaBoost algorithm with $M = 4$ to train a classifier for this data set. Draw the 4 selected trees in the order that they occur and report the ϵ, generated by adaBoost, for each.

   (b) How does the accuracy of the adaBoost classifier change with different values of M? Report the accuracy on the test set for $M \in \{8, 16, 32, 64\}$.

3. Now, suppose that the hypothesis space consists only of depth 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 $M = 16$ for this hypothesis space on the test data? How does the α learned by adaBoost compare to the one learned by gradient descent?