Introduction to Machine Learning CS436/580L

Weiying Dai, Fall 2016

Homework Assignment 2

Out: Sep. 26, 2016

Due: Questions 1-9 at beginning of class (1:15pm sharp!) on Oct 11, 2016; Question 10 at 11:59pm to the blackboard on Oct. 11, 2016

## **Hypothesis Evaluation** (35 points)

1. (5 points)

Suppose you test a hypothesis ***h*** and find that it commits ***r*** = 240 errors on a sample S of ***n*** = 800 randomly drawn test examples. What is the standard deviation in $Error_S(h)$? How does this compare to the standard deviation in the Lecture example (the example was to calculate the standard deviation in $Error_S(h)$ based on the finding of 12 errors on a sample of size 40)? State the reason for your observation.

2. (8 points)

Consider a learned hypothesis, h, for some boolean concept. When h is tested on a set of 100 examples, it classifies 79 correctly. What is the standard deviation and the 95% confidence interval for the true error rate for $Error_{\mathcal{D}}(h)$?

3. (10 points)

You are about to test a hypothesis h whose $Error_{\mathcal{D}}(h)$ is known to be in the range between 0.3 and 0.6. What is the minimum number of examples you must collect to assure that the width of the two-sided 95% confidence interval will be smaller than 0.1?

4. (12 points)

Suppose hypothesis h commits r = 10 errors over a sample of n = 75 independently drawn examples. What is the 90% confidence interval (two-sided) for the true error rate? What is the 95% one-sided interval (i.e. what is the upper bound U such that $Error_{\mathcal{D}}(h) \leq U$ with 95% confidence)? What is the 90% one-sided interval?

## **Computational Learning Theory** (65 points)

5. (7 points)

Consider training a linear separating hyperplane in a 2-dimensional space (i.e., to find a line that separates two classes). Give an upper bound on the number of training examples sufficient to assure with 90% confidence that the learned line

will have true error of at most 5%. Does this bound seem realistic?

6. (18 points)

Consider the space of instances $X$ corresponding to all points in the $x, y$ plane. Give the VC dimension of the following hypothesis spaces:

(a) $H_r$ = the set of all rectangles in the $x, y$ plane. That is, $H = \{((a < x < b) \wedge (c < y < d)) | a, b, c, d \in \Re\}$. Points inside the rectangle are classified as positive examples.

(b) $H_c$ = circles in the $x, y$ plane. Points inside the circle are classified as positive examples

7. (10 points)

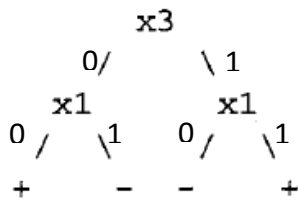
Consider the class C of concepts of the form $(a \leq x \leq b) \wedge (c \leq y \leq d)$,where a ,b , c, and d are integers in the interval (0,100). Note each concept in this class corresponds to a rectangle with integer-valued boundaries on a portion of the x, y plane. Hint: Given a region in the plane bounded by the points (0,0) and (n, n), the number of distinct rectangles with integer-valued boundaries within this region is $(n(n+1)/2)^2$.

(a) Give an upper bound on the number of randomly drawn training examples sufficient to assure that for any target concept $c$ in $C$, any consistent learner using $H = C$ will, with probability 95%, output a hypothesis with error at most .15.

(b) Now suppose the rectangle boundaries $a, b, c$, and $d$ take on *real* values instead of integer values. Update your answer to the first part of this question.

8. (16 points)

Consider the hypothesis class $H_{rd2}$ of "regular, depth-2 decision trees" over $n$ Boolean variables. A "regular, depth-2 decision tree" is a depth-2 decision tree (a tree with four leaves, all distance 2 from the root) in which the left and right child of the root are *required to contain the same variable*. For instance, the following tree is in $H_{rd2}$.




(a) As a function of $n$, how many syntactically distinct trees are there in $H_{rd2}$?

(b) Give an upper bound for the number of examples needed in the PAC model to learn $H_{rd2}$ with error $\epsilon$ and confidence $\delta$.

9. (14 points)

This question considers the relationship between the PAC analysis considered in this chapter and the evaluation of hypotheses discussed in Chapter 5. Consider a learning task in which instances are described by $n$ boolean variables (e.g., $x_1 \wedge \bar{x_2} \wedge x_3 \ldots \bar{x_n}$) and are drawn according to a fixed but unknown probability distribution $\mathcal{D}$. The target concept is known to be describable by a conjunction of boolean attributes and their negations (e.g., $x_2 \wedge \bar{x_5}$), and the learning algorithm uses this concept class as its hypothesis space $H$. A consistent learner is provided a set of 100 training examples drawn according to $\mathcal{D}$. It outputs a hypothesis $h$ from $H$ that is consistent with all 100 examples (i.e., the error of $h$ over these training examples is zero).

($a$) We are interested in the true error of $h$, that is, the probability that it will misclassify future instances drawn randomly according to $\mathcal{D}$. Based on the above information, can you give an interval into which this true error will fall with at least 95% probability? If so, state it and justify it briefly. If not, explain the difficulty.

($b$) You now draw a new set of 100 instances, drawn independently according to the same distribution $\mathcal{D}$. You find that $h$ misclassifies 30 of these 100 new examples. Can you give an interval into which this true error will fall with approximately 95% probability? (Ignore the performance over the earlier training data for this part.) If so, state it and justify it briefly. If not, explain the difficulty.

## Bias and Variance Dilemma

10. (30 points)

Generate 100 datasets from a known function $f(x) = 2\cos(2.8x) + 7.5$ with added noise following normal distribution $\mathcal{N}(0, 1)$. For each dataset, use the same 20 instances with $x \in [0\ 5]$ following uniform distribution. For each of 100 datasets, perform polynomial fits of order 1, 2, 3, 4, and 5. (a) Plot only the first five polynomial fits (using solid line) for each polynomial order and the average of the 100 fits (using dotted line), you do not have to show the 20 sample points. Plot the five polynomial fits of each order in a separate figure. (b) Plot bias, variance, and error for polynomials of order 1 to 5 using the equations in the lecture notes. (c) How do the bias and variance vary as the order of polynomial increases? Which order of polynomial has the minimal error?

Note: For Questions 1-9, the answers should be typed (handwritten answers will be penalized by 20% of the full credit). Questions 1-9 are due at the beginning of the class (at 1:15pm on Oct. 11). For Questions 10, source code, executables, and README (the answer of Question 10(c) needs to be included in the README file), should be compressed in one single tar file (with the file name as yourlastname_prog2.tar) and submitted through the Blackboard (due at 11:59pm on Oct. 11). Source code should be well commented. README file should clearly

describe how to compile the source and run the executables. 10% of the grade will be based on good coding style and meaningful comments.