CS 5350/6350: Machine Learining Spring 2015

Homework 3

Handed out: Feb 18, 2015
Due date: ~~Mar 4~~ Mar 8, 2015

General Instructions

- You are welcome to talk to other members of the class about the homework. I am more concerned that you understand the underlying concepts. However, you should write down your own solution. Please keep the class collaboration policy in mind.
- Feel free ask questions about the homework with the instructor or the TAs.
- Your written solutions should be brief and clear. Your assignment should be **no more than 10 pages**. X points will be deducted if your submission is X pages beyond the page limit.
- Handwritten solutions will not be accepted.
- The homework is due by midnight of the due date. Please submit the homework on Canvas.
- Some questions are marked **For 6350 students**. Students who are registered for CS 6350 should do these questions. Of course, if you are registered for CS 5350, you are welcome to do the question too, but you will not get any credit for it.

1 Warm up: Feature expansion

[10 points] Recall problem 2 in Homework 2, where we saw the concept class C consisting of functions function f_r defined by an integer radius r (with $1 \leq r \leq 128$) as follows:

$$f_r(x_1, x_2) = \begin{cases} +1 & x_1^2 + x_2^2 \leq r^2; \\ -1 & \text{otherwise} \end{cases} \tag{1}$$

Clearly the hypothesis class is *not* linearly separable in $\Re^2$.

Construct a function $\phi(x_1, x_2)$ that maps examples to a new space, such that the positive and negative examples are linearly separable in that space. (Note that ϕ should not depend on r).

2 PAC Learning

1. [20 points total] A factory assembles a product that consist of different parts. Suppose a robot was invented to recognize whether a product contains all the right parts. The rules of making products are very simple: 1) you are free to combine any of the parts

as they are 2) you may also cut any of the parts into two distinct pieces before using them. You wonder how much effort a robot would need to figure out the what parts are used in the product.

(a) [5 points] Suppose that a naive robot has to recognize products made using only rule 1. Given N available parts and each product made out of these constitutes a distinct hypothesis. How large would the hypothesis space be? Brief explain your answer.

(b) [5 points] Suppose that an experienced worker follows both rules when making a product. How large is the hypothesis space now? Explain.

(c) [10 points] An experienced worker decides to train the naive robot to discern the makeup of a product by showing you the product samples he has assembled. There are 6 available parts. If the robot would like to learn any product at 0.01 error with probability 99%, how many examples would the robot have to see?

2. [20 points, from Tom Mitchell's book] We have learned an expression for the number of training examples sufficient to ensure that every hypothesis will have true error no worse than ϵ plus its observed training error $error_D(h)$. In particular, we used Hoeffding bounds to derive

$$m \geq \frac{1}{2\epsilon^2}(\ln(|H|) + \ln(1/\delta)).$$

Derive an alternative expression for the number of training examples sufficient to ensure that every hypothesis will have true error no worse than $(1+\gamma)error_D(h)$. You can use general Chernoff bounds to derive such a result.

Chernoff bounds: Suppose $X_1, \cdots, X_m$ are the outcomes of m independent coin flips (Bernoulli trials), where the probability of heads on any single trail is $Pr[X_i = 1] = p$ and the probability of tails is $Pr[X_i = 0] = 1 - p$. Define $S = X_1 + X_2 + \cdots + X_m$ to be the sum of these m trials. The expected value of S/m is $E[S/m] = p$. The Chernoff bounds govern the probabilty that S/m will differ from p by some factor $0 \leq \gamma \leq 1$.

$$\begin{aligned} Pr[S/m > (1+\gamma)p] &\leq e^{-mp\gamma^2/3} \\ Pr[S/m < (1-\gamma)p] &\leq e^{-mp\gamma^2/2} \end{aligned} \tag{2}$$

3 VC Dimension

In this problem, we investigate a few properties of the Vapnik-Chervonenkis dimension.

1. [Shattering, 10 points] Recall the definition of shattering from class: a concept class shatters a set of points if for any labeling of those points, there is some function in the class that correctly labels it. Let C be the set of all conjunctions of n Boolean variables. Find a set $S \subseteq \{0,1\}^n$ consisting of exactly n examples that can be shattered by C. Prove the correctness of your answer.

2. [10 points] Show that a finite concept class C has VC dimension at most $\log |C|$. Hint: You can prove this by contradiction.

3. [15 points] We have a learning problem where each example is a point in $\Re^2$. The concept class H is defined as follows: A function $h \in H$ is specified by two parameters a and b. An example $\mathbf{x} = \{x_1, x_2\}$ in $\Re^2$ is labeled as $+$ if and only if $x_1 \geq a$ and $x_2 \leq b$ and is labeled $-$ otherwise.

   For example, if we set $a = 1, b = 4$, the grey region in figure 1 is the region of $\mathbf{x} = \{x_1, x_2\}$ that has label $+1$.

   What is the VC dimension of this class?

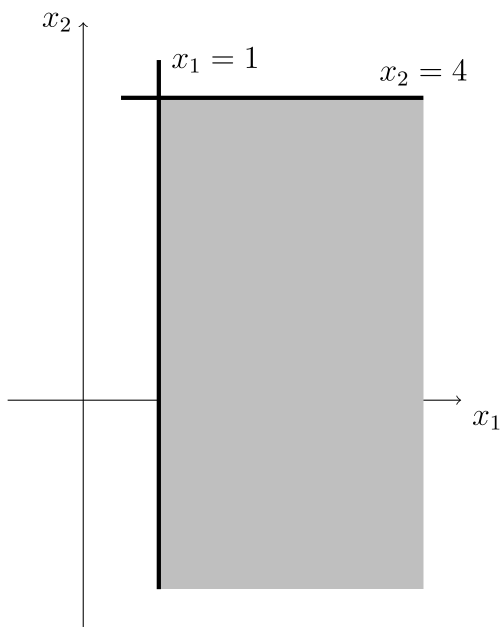

Figure 1: An example with $a = 1, b = 4$. All points in the gray region (extending infinitely) shows the region that will be labeled as positive.

4. [15 points] Determine the VC dimension of concept class consisting of the union of 2 intervals on the real line (that is, points within either of the intervals will be labeled as positive).

5. [**For 6350 Students**, 10 points] Generalize the result of the above question. Find the VC dimension of the union of k intervals on the real line.

6. [**For 6350 Students,** 15 points] Let two hypothesis classes H_1 and H_2 satisfy $H_1 \subseteq H_2$. Prove: $VC(H_1) \leq VC(H_2)$.

What To Submit

1. Your assignment should be **no more than 10 pages**. X points will be deducted if your submission is X pages beyond the page limit.

2. Your report should be in the form of a *pdf* file, LaTeX is recommended.

3. Please look up the late policy on the course website.