CS 5350/6350: Machine Learining Fall 2018

Homework 4

Handed out: October 26, 2018
Due date: November 13, 2018

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 discuss the homework with the instructor or the TAs.
- Your written solutions should be brief and clear. You need to show your work, not just the final answer, but you do *not* need to write it in gory detail. Your assignment should be **no more than 10 pages**. Every extra page will cost a point.
- 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.

Important Do not just put down an answer. We want an explanation. No points will be given for just a statement of the results of a proof. You will be graded on your reasoning, not just on your final result.

Please follow good proof technique; what this means is if you make assumptions, state them. If what you do between one step and the next is not trivial or obvious, then state how and why you are doing what you are doing. A good rule of thumb is if you have to ask yourself whether what you are doing is obvious, then it is probably not obvious. Try to make the proof clean and easy to follow.

1 PAC Learning

1. 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 for 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.

- [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.
- [5 points] Suppose that an experienced worker follows both rules when making a product. How large is the hypothesis space now? Explain.
- [10 points] An experienced worker decides to train the naive robot to discern the makeup of a product by showing it the product samples he has assembled. There are 6 available parts. If the robot has to learn any product at 0.01 error with probability 99%, how many examples would the robot have to see?

2. [20 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, d are integers in the interval $(0, 20)$. Each concept in this class consists of a rectangle with integer valued boundaries and labels all points inside the rectangle as positive and everything outside as negative.

   Give an upper bound on the number of randomly drawn examples needed to ensure that for any function c in the set C, a consistent learner that uses $H = C$ will, with probability 99% produce a classifier whose error is no more than 0.01.

   Hint: To answer this question, you could use the fact that in a plane bounded by the points $(0, 0)$ and (n, n), the number of distinct rectangles with integer-valued boundaries in the region is $\left(\frac{n(n+1)}{2}\right)^2$.

2 Shattering

[15 points] Suppose we have a set X_n consists of all binary sequences of a length n. For example, if $n = 3$, the set would consist of the eight elements {`000, 001, 010, 011, 100, 101, 110, 111`}.

Consider a set of functions H_n that we will call the set of *templates*. Each template is a sequence of length n that is constructed using `0`, `1` or `-` and returns $+1$ for input binary sequences that match it and -1 otherwise. While checking whether a template matches an input, a `-` can match both a `0` and a `1`.

For example, the template `-10` matches the binary strings `010` and `110`, while `-1-` matches all strings that have a `1` in the middle position, namely `010`, `011`, `110` and `111`.

Does the set of templates H_n shatter the set X_n? Prove your answer.

3 VC Dimensions

1. Consider learning problems where examples are points in the two dimensional plane. What is the VC dimension of the following concept classes? (In each case you need to prove your answer by showing both the upper and lower bounds.)

   (a) [15 points] The concept class H_c consisting of circles, with points strictly outside being negative.

(b) [15 points] 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?

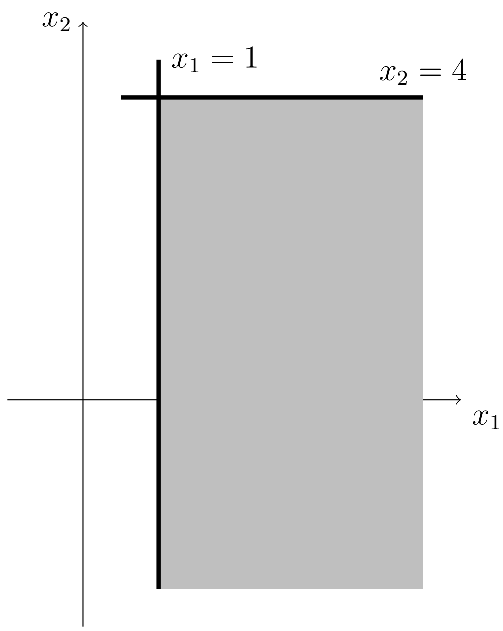

Figure 1: An example with $a = 1, b = 4$. All points in the gray region (extending infinitely) shows the region that will be labeled as positive.

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

4 AdaBoost

[15 points] You are given the following examples in the table below

$\mathbf{x} = [x_1, x_2]$	y
$[1, 1]$	-1
$[1, -1]$	1
$[-1, -1]$	-1
$[-1, 1]$	-1

Suppose you are also given the following 4 weak classifiers (i.e. rules of thumb).

$$
\begin{aligned}
f_a(\mathbf{x}) &= \operatorname{sgn}(x_1)\\
f_b(\mathbf{x}) &= \operatorname{sgn}(x_1 - 2)\\
f_c(\mathbf{x}) &= -\operatorname{sgn}(x_1)\\
f_d(\mathbf{x}) &= -\operatorname{sgn}(x_2)
\end{aligned}
$$

1. Step through the full AdaBoost algorithm by choosing f_t from the above weak classifiers in each round. Remember that you need to choose a hypothesis from f_a, f_b, f_c, f_d whose weighted classification error is **less than half**.
2. At the end, state the final hypothesis $H_{final}(x)$.

To get you started, we have chosen f_a as the first hypothesis and show the values of ϵ_1, α_1, Z_1, D_1 in the table below. Fill in the values of D_2. For ease of grading, please follow this template below: Report the hypothesis you choose, its predictions for all the examples, and the values of ϵ_t, α_t, Z_t, D_t, and D_{t+1} for *three subsequent rounds*

Round 1: Choose $h_1(\mathbf{x}) = f_a(\mathbf{x}) = \operatorname{sgn}(x_1)$.

$\mathbf{x} = [x_1, x_2]$	y_i	$f_a(x)$	D_1	$D_1(i) y_i h_t(\mathbf{x_i})$	D_2
$[1, 1]$	-1	1	1/4	-1/4	
$[1, -1]$	1	1	1/4	1/4	
$[-1, -1]$	-1	-1	1/4	1/4	
$[-1, 1]$	-1	-1	1/4	1/4	

$\epsilon_1 = 1/4, \alpha_1 = \frac{\ln 3}{2}, Z_1 = \frac{\sqrt{3}}{2}$