CS 5350/6350: Machine Leariining Fall 2017

Homework 4

Handed out: Thursday October 26th, 2017
Due date: Thursday November 9th, 2017

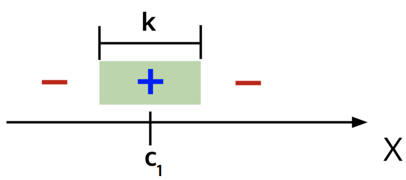
General Instructions

- 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. If you are registered for CS 5350, you are welcome to do the question too for extra credit.

1 Feature Transformation

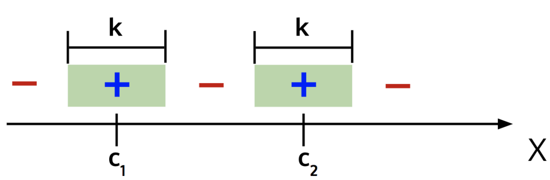
A feature transformation is a function that maps instances $\mathbf{x}$ to some new feature space $\phi(\mathbf{x})$. Recall that in class we saw an example of how to use a feature transformation to transform a set of instances which was not linearly seperable into one that was.

1. [5 points] Consider an instance space where all instances are real-valued and one dimensional. These instances are labeled according to a function which considers an interval of width k centered at c_1. All points within the intervals are labeled True, all points outside the intervals are labeled False.



These instances are not linearly seperable. Write a feature transformation to make these instances linearly seperable. You may **only** use the L_1 distance norm in your transformation, and the resulting instances should still be in 1 dimension. (A distance norm is a function which measures distance.)

2. [7 points] Now consider a variation: this new function now considers two intervals, both of width k, one of which is centered at c_1, and the other centered at c_2. All points within the intervals are labeled True, all points outside the intervals are labeled False.



This instance space is also not linearly seperable. Write a feature transformation to make these instances linearly seperable, by transforming it from 1 dimension into 2 dimensions, again using the L_1 distance norm in your transformation.

3. [3 points] "Fizzbuzz" is a classic interview question which asks you to write a function that assigns True to integers which are divisable by 3 or by 5, and False to all other integers. Show that the fizzbuzz instance space is not linearly seperable.

4. [15 points] Write a feature transformation to make the fizzbuzz instance space linearly seperable, by transforming it from 1 dimension into 2 dimensions. You may **only** use functions of the form cosine(ax) in your transformation (where a is a real valued number).

2 PAC Learning

Fall is here, which means it's time to go get a pumpkin spice latte! Unfortunately, there are so many options this year that you feel completely overwhelmed and decide to build a classifier to automate your feelings about a given combination of latte options. You have five possible reactions: *Exactly what I wanted*, *Delicious*, *Passable*, *Terrible*, *Too hot*.

1. [5 points] Suppose you have: M options for type of milk, W options for types of whipped cream, S options for latte size, C options for straw color. How large is the hypothesis space?

2. [10 points] You plan to train the classifier at a Starbucks location, and you plan to collect data by ordering lots of lattes which elicit each of the 5 emotions. This Starbucks offers 4 milk options, 2 whipped cream options, 3 size options, and 1 straw color option. If you want to the classifier to learn to recognize how you feel about which latte with 5 percent error with greater than 70% probability, how many lattes do you need to order?

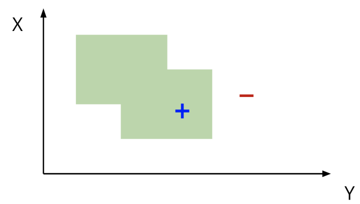
3. [20 points] Let X be the space of booleans in n dimensions, and let H be a set of hypotheses over X. H contains all "singleton" functions, and also the "all-negative" function. A singleton function f_z is true only when z is the input, and is false otherwise, where z is some element of X. Because H contains all singleton functions, it contains a singleton function for every element of X. The "all-negative" function is false for any input. This means that the true hypothesis f labels all examples in the domain negatively, perhaps except for one example.

   Show that H is PAC learnable. Provide an upper bound on the sample complexity.

   Hint: Recall that we saw an example in class of how to show that a different hypothesis set H was PAC learnable.

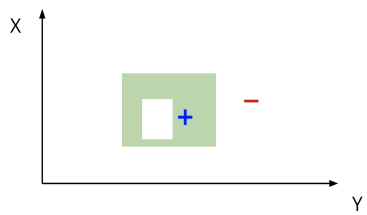
3 VC Dimension

1. [10 points] In class we saw how to find the VC dimension of axis parallel rectangles. Now let's consider a variation: consider the concept class that is made by taking the union of 2 axis aligned rectangles, as in the figure below. What is the VC dimension for this concept class?



Assume the following: the two rectangles do not need to be the same size or proportions, they may overlap in any orientation, and they must overlap (meaning two disjoint rectangles are not included in this concept class), and they are both 2-dimensional.

2. [10 points] Now consider a different variation: consider the concept class that is made by taking one axis parallel rectangle and subtracting another axis parallel rectangle, as in the figure below. What is the VC dimension for this concept class?



Assume the following: the rectangles do not need to have the same proportions, and they must overlap, and they are both 2-dimensional.

3. [15 points] Show that a finite concept class C has VC dimension at most log | C |. Hint: You can prove this by contradiction.

4 For 6350 Students only: Properties of VC Dimension

These problems are required for students registered for 6350. Students registered for 5350 can do these problems for extra credit.

1. [10 points] Prove that VC-dimension exhibits the following property: Given two hypothesis classes, H and G, if $H \subseteq G$ then VCdim(H) $\leq$ VCdim(G).

2. [5 points] Given a finite set of items X, and an integer k $<$ | X |, consider the hypothesis space which consists of all functions which assign a value of 1 to exactly k of items in X.

   What is the VC-dimension of this hypothesis space? Justify your answer.

3. [5 points] Now consider a variation: consider the hypothesis space which consists of all functions which assign a value of 1 to **at least** k of items in X.

   What is the VC-dimension of this hypothesis space? Justify your answer.