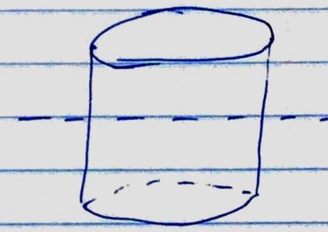
CS 6375
PSET 3

Student: TRI M. CAO
Date: 10/19/2017

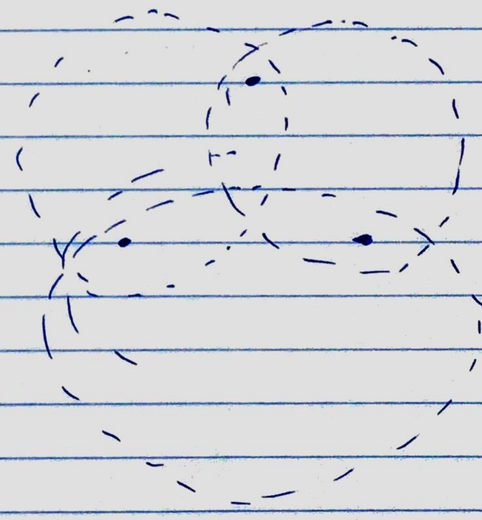
Problem 1:

1. cylinders of a fixed radius $r > 0$ and fixed length $l > 0$.

- If we ~~stide~~ slice the cylinder in the direction of the dashed line on the right, we will have a circle in 2-D with radius r.



We want to argue that the VC dimension of a circle is 3.
Given 3 points in $\mathbb{R}^2$, it's clear that we will always find a circle that cover all + points. For example, let ~~all~~ the 3 points be:



So the VC dimension ≥ 3.

- If we have 4 points, then a circle will not be able to cover all the + points.
We have two cases:

Case 1:

+

−

+ +

case 2:

− +

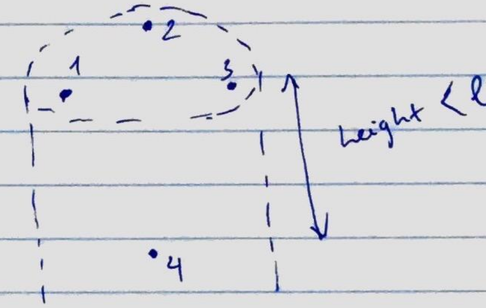
+ −

In both cases, we cannot find a circle that covers the + points.
Also, if We have more than 4 points, then we can always find a group of 4 points that looks like one ~~of~~ of the cases above

$\Rightarrow$ VC dimension of circle is 3.

- Now we will argue that the VC dimension of a cylinder in $\mathbb{R}^3$ is 4.

Given 4 points like this:
the 3 points at the top will make up a plane p. The point below will not be on plane p, but on a plane that is lower in height from plane p.
The height difference is less than l.



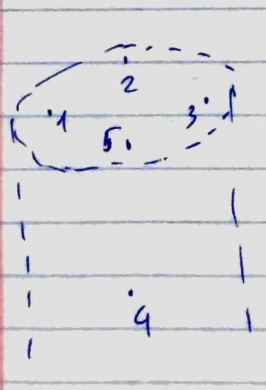
we choose those 4 points such that we ~~do~~ can classify the 3 points on plane p with a circle, and that circle can extend the ~~ta~~ length to also cover the lower point (point 4).

It's clear that we can classify the 3 points on plane p. For point 4, if it is + then we move the cylinder down to cover ~~the~~ it; if it is − then we move the cylinder up to avoid it. Either way, we can classify all 4 points.

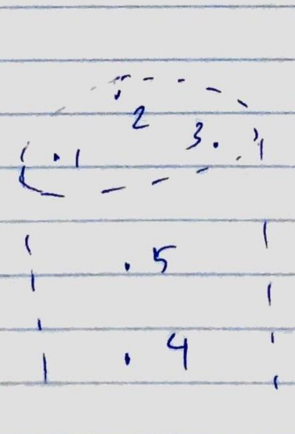
$$\Rightarrow \underline{VC(H) \geq 4}$$

- when we add one more point ~~the~~ to the 4 points above, we have 3 cases:

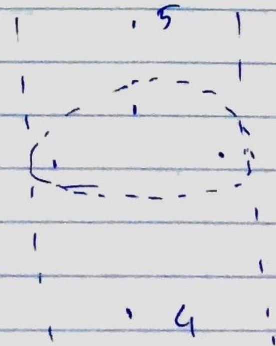
case 1



case 2



case 3



Case 1: point 5 is on the same plane as points 1, 2, 3.
In this case we cannot classify 4 points on the same plane with a circle so we fail to classify 5 points.

Case 2: point 5 ~~has a height~~ is not on plane p, but has a height that makes it between plane p and point 4.
If point 4 is + and point 5 is −, then it's impossible for a cylinder to cover all + points.

Case 3: point 5 is higher than plane p.
If height of 5 − height of 4 > l, then we cannot classify the case when point 5 and 4 are both +.
If height of 5 − height of 4 < l, then we cannot classify the case when both 5 and 4 are −.

1a) Sample complexity

$$M \geq \frac{1}{\epsilon}\left(4 \ln \frac{2}{\delta} + 8 \cdot VC(H) \ln \frac{13}{\epsilon}\right)$$

We have $\epsilon = 0.2$, $\delta = 0.05$, $VC(H) = 4$

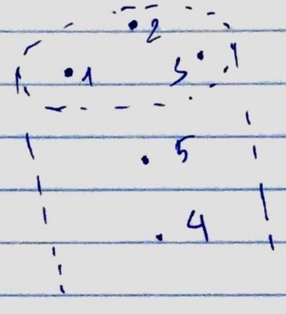
$$\Rightarrow M \geq \frac{1}{0.2}\left(4 \ln \frac{2}{0.05} + 8 \cdot 4 \ln \frac{13}{0.2}\right)$$

$$\boxed{M \geq 5\left(4 \ln (40) + 32 \cdot \ln (65)\right)}$$

1b) If we also have cylinders that cover − points then we will have VC dimension of 5.

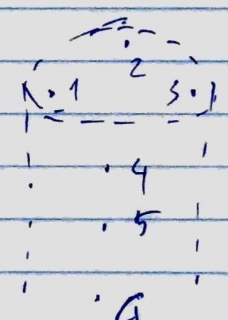
The new hypothesis space will solve case 2 in the discussion before.

If point 5 is − and point 4 is +, we will use a cylinder to cover all − points, that way we can ignore ~~the~~ point 4.
In other cases of point 5, we can still cover all + points.



~~I~~ Now let's try to add one more point, then we would see case 1 and case 3 again, and we cannot ~~solv~~ solve those cases. look at case 2, we would see something like this

If point 4 is −, point 5 is +, point 6 is −, then we cannot ~~find~~ classify those points.



Note that if we add more than one point we still have the 3 cases above.

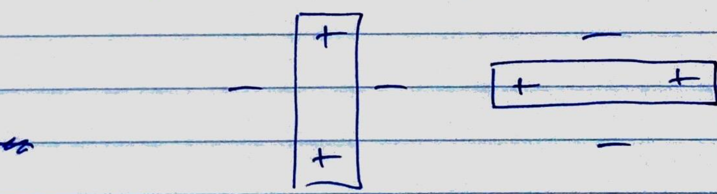
$\Rightarrow \boxed{VC(H) = 5}$

2 Pair of axis aligned rectangles.

- From the lecture, we have VC dimension of a single axis aligned rectangle is 4.
- we want to argue that two rectangles will have VC dimension of 8.

Given 8 points

For any labeling, a single axis aligned rectangle can classify the group of 4 points on the left, and the remaining rectangle will handle the group on the right. For example:

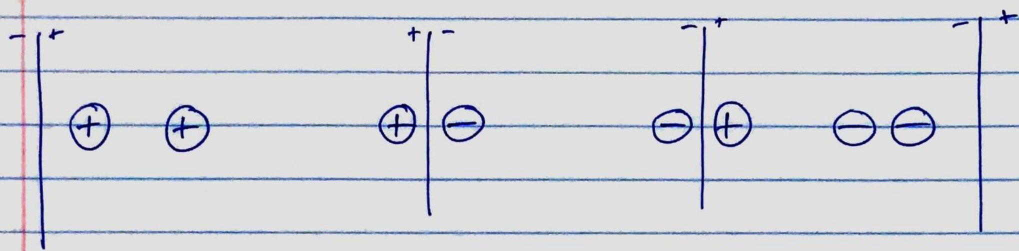


- If we add one or more points, then we no longer have two groups of 4 points. There will be at least one group of 5 or more points, which a single rectangle cannot classify.

$\Rightarrow \boxed{VC(H) = 8}$

Problem 1.3

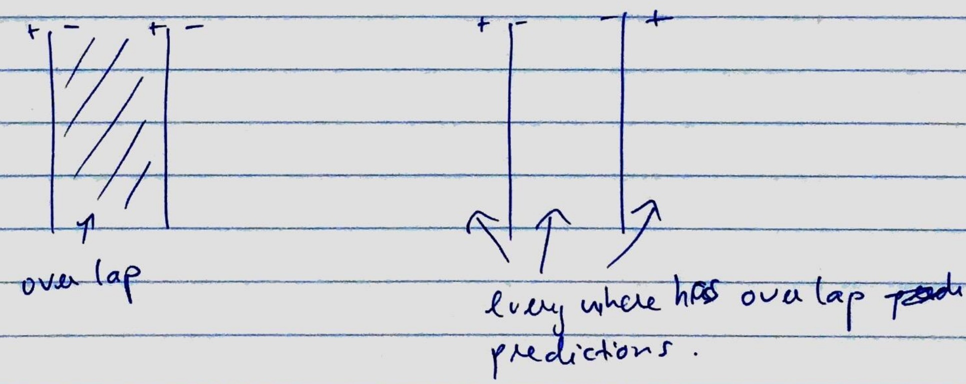
We have 3 linear classifiers.
let's try to classify two points:



So we need 4 different linear separators to classify two points.
$\Rightarrow VC(H) < 2$

— We can always classify one point with 3 linear separators
$\Rightarrow VC(H) \geq 1$
The reason is, if we have two distinct linear separators, then there will be a space where the predictions of the two linear separators overlap.

Examples:



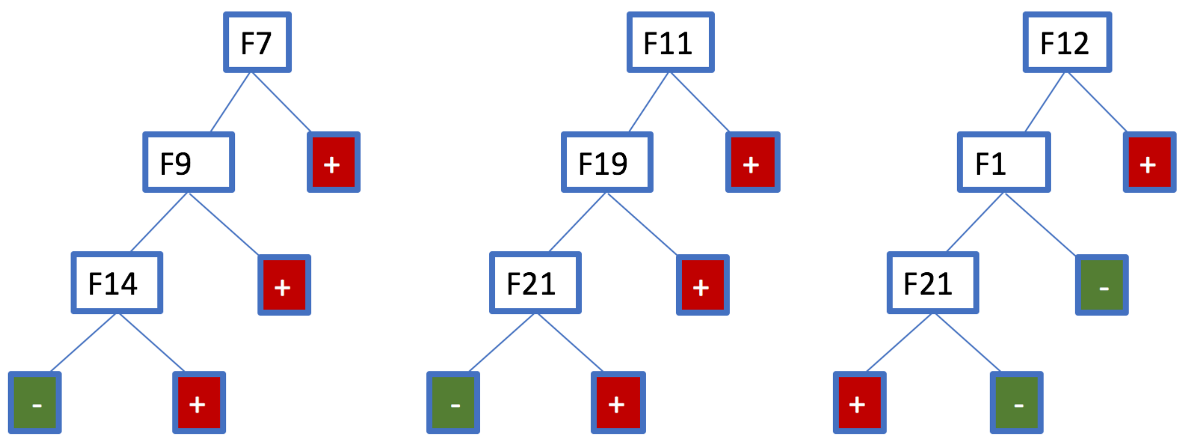
— Therefore I guess the min and max VC dimension of 3 distinct linear separators are both 1

CS 6375
Professor Nick Ruozzi
Student: Tri M. Cao
Problem Set 3
Date: October 19, 2017

Problem 2:

1.

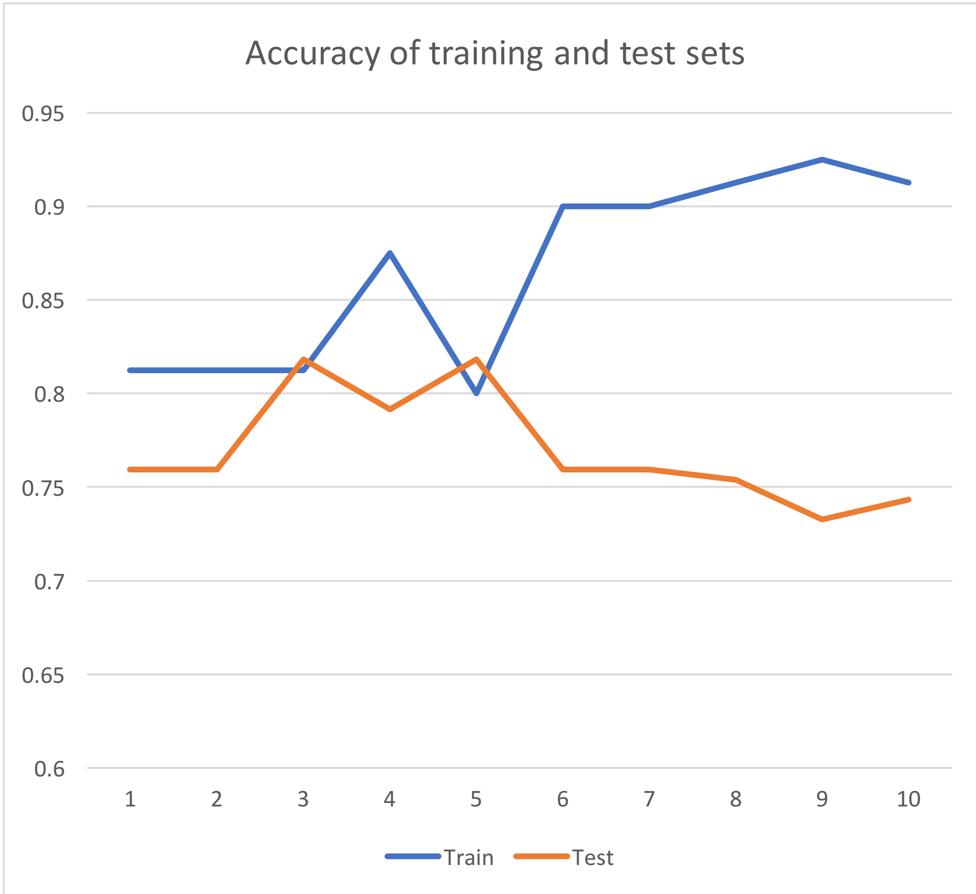
a. The 3 selected trees are (in order left to right):



The errors are respectively: 0.1875, 0.2692, 0.3404.

b. Run the adaBoost algorithm for 10 rounds.

Iteration	Train	Test
1	0.8125	0.7594
2	0.8125	0.7594
3	0.8125	0.8182
4	0.875	0.7914
5	0.8	0.8182
6	0.9	0.7594
7	0.9	0.7594
8	0.9125	0.754
9	0.925	0.7326
10	0.9125	0.7433



2. Coordinate descent
    a. When running coordinate descent, I just iterate over the trees in the order that I built them. After completing 500 loops over all trees, I get the **exponential loss of 39.691**.
    The values of alpha are listed at the end of this paper. There are 88 alpha values, each correspond to a tree in the hypothesis space.

    b. The accuracy of the classifier trained by **coordinate descent is 0.7005**.

    c. The accuracy of **adaBoost with M=20 is 0.668**.
    The alphas learned by adaBoost are very different from the ones learned by coordinate descent. No alpha value of adaBoost is negative, but there are many negative alphas in coordinate descent. Also, because we use M=20 in adaBoost, alpha would have at most 20 nonzero values. Meanwhile, alpha with coordinate descent has 44 nonzero values.

List of alphas generated by coordinate descent:

alpha 0 : 0
alpha 1 : -2.11136914605
alpha 2 : 0
alpha 3 : -0.764582190176
alpha 4 : 0
alpha 5 : -0.492332893361
alpha 6 : 0
alpha 7 : -0.414782041468
alpha 8 : 0
alpha 9 : -4.40585498748
alpha 10 : 0
alpha 11 : -2.81003779319
alpha 12 : 0
alpha 13 : 0.801991500919
alpha 14 : 0
alpha 15 : 0.519026835117
alpha 16 : 0
alpha 17 : 0.31751616799
alpha 18 : 0
alpha 19 : 0.141265303489
alpha 20 : 0
alpha 21 : 0.137296705114
alpha 22 : 0
alpha 23 : 0.108798260393
alpha 24 : 0
alpha 25 : 0.608145225077
alpha 26 : 0
alpha 27 : 0.322854292652
alpha 28 : 0
alpha 29 : 4.02545218409
alpha 30 : 0
alpha 31 : 2.67532600282
alpha 32 : 0
alpha 33 : -2.7467516857
alpha 34 : 0
alpha 35 : -2.09493771074
alpha 36 : 0
alpha 37 : 1.97905322099
alpha 38 : 0
alpha 39 : 0.754239138272
alpha 40 : 0
alpha 41 : 0.387701528922
alpha 42 : 0
alpha 43 : 0.256301751523
alpha 44 : 0
alpha 45 : -0.453426019943

alpha 46 : 0
alpha 47 : -0.277082109197
alpha 48 : 0
alpha 49 : 0.794651871233
alpha 50 : 0
alpha 51 : 0.386217574122
alpha 52 : 0
alpha 53 : 2.21784860527
alpha 54 : 0
alpha 55 : 1.68657589332
alpha 56 : 0
alpha 57 : 0.803818634551
alpha 58 : 0
alpha 59 : 0.783771319516
alpha 60 : 0
alpha 61 : 0.732113848047
alpha 62 : 0
alpha 63 : 0.588022904909
alpha 64 : 0
alpha 65 : 2.75415169632
alpha 66 : 0
alpha 67 : 2.69231034457
alpha 68 : 0
alpha 69 : 1.94955498116
alpha 70 : 0
alpha 71 : 1.93019470963
alpha 72 : 0
alpha 73 : 0.0518860455339
alpha 74 : 0
alpha 75 : 0.033996854833
alpha 76 : 0
alpha 77 : 0.317123147077
alpha 78 : 0
alpha 79 : 0.209027302585
alpha 80 : 0
alpha 81 : 0.167907866749
alpha 82 : 0
alpha 83 : 0.115825356003
alpha 84 : 0
alpha 85 : 0.369651353905
alpha 86 : 0
alpha 87 : 0.171223180525