Ritesh Sharma
Assignment 4(CS 5350)

Warm up

1.

a)

Yes, H is PAC-learnable.

The algorithm is polynomial in $1/\delta$, $1/\varepsilon$ if the number of examples required for an algorithm with probability of at least (1 - δ) and with error of at most ε.

The size of the instance space is infinite and hypothesis class is n.

This is only true if ln(|H|) is polynomial expression.

We know the equation from PAC-learning theory,

$$m > \frac{1}{\varepsilon}\left(\ln(|H|) + \ln\left(\frac{1}{\delta}\right)\right)$$

having at least a certain number of examples guarantees PAC-learnability for a consistent learner.

Since |H| is n in this case,

$\ln(|H|) = \ln(n)$

Because this is a polynomial expression the training set will always produce the same results, with 0 training error.

This means H is PAC-learnable.

b)

Because we only have one feature(which is a real number), the dimensionality of the instance space is 1.

Given a point x from instance space, Assume that it is an integer.

We can find h $\in$ H such that the threshold t > x.

This means that the VC dimension is at least 1, and this is because we can shatter one such number from R.

We cannot shatter a dataset of 2 points.

Following are configuration for 2 points:

CASE 1: x - y > 1

Since x > y. The difference between the number is greater than 1. This means there exists some integer between x and y.

We can choose a h $\in$ H with a t $\in$ T that will correctly classify the points

when,

1) h(x) = 1 and h(y) =0 (t is between x and y)
2) h(x) = 1 and h(y) = 1 (t is smaller than x and y)
3) h(x) = 0 and h(y) = 0 (t is larger than x and y)

For 1, (from above), we cannot choose a t that will correctly classify both.
If there was such a t, then y > t and x < t, this causes contradiction.
We cannot shatter this subset.

CASE 2: x - y < 1

This means that the difference between x and y is less than 1 i.e. no integer exists between them.

This means we can choose h ∈ H with a t ∈ T that will correctly classify the points when,

1) h(x) = 1 and h(y) =0 or h(x) = 1 and h(y) = 1(t is between x and y)
2) h(x) = 1 and h(y) = 1 (t is lower than x and y)
3) h(x) = 0 and h(y) = 0 (t is higher than x and y)

For 1, (from above), we cannot choose a t that will classify both correctly. If there was such t, then either y > t > x or y < t < x but that cannot be between the two. This causes a contradiction.
We cannot shatter this subset.

CASE 3: x - y = 1:

This means the difference between x and y is one, that means we can either choose h ∈ H with a t ∈ T that is either between them or is not.

From Case 1 and 2, both these splits yield a subset that cannot be shattered.

PAC Learning

1.

a)

Possibility of root node = 20
Possibility of child node = 19
Because each node has 4 possible ways to identify the labels of the data.
So, 20 * 19 * 4 * 4 = **6080** structurally different trees.

b)

Since decision tree is a consistent learner. The bound of m is defined by

$$m > \frac{1}{\varepsilon}\left(\ln(|H|) + \ln\left(\frac{1}{\delta}\right)\right)$$

Given, $\varepsilon = 1 - 0.99 = 0.01$

$\delta = 1 - 0.95 = 0.05$ and

$H = 6080$

$$m > \frac{1}{0.01} (\ln(|6080|) + \ln(\frac{1}{0.05}))$$

$$m > 1170.85$$

Hence, we need at least 1171 examples.

c)[**Extra Credit]**
Let's say we only care about the labels of the child nodes and not the feature valus.
This will give 4 possible values for each child node given root nodes.
So there are 20 possibilities for the root node and 16 for child node.
Hence |H| is 20 * 16 = **320**

2)
Assume nothing about the consistency of the learner.
Because H is PAC-learnable, the sample complexity is,

$$m >= \frac{1}{2\varepsilon^2} (\ln(|H|) + \ln(\frac{1}{\delta}))$$

$m(\varepsilon 1, \delta)$ and $m(\varepsilon 2, \delta)$ is then defined as:

$$m(\varepsilon 1, \delta) >= \frac{1}{2\varepsilon 1^2} (\ln(|H|) + \ln(\frac{1}{\delta}))$$

$$m(\varepsilon 2, \delta) >= \frac{1}{2\varepsilon 2^2} (\ln(|H|) + \ln(\frac{1}{\delta}))$$

We know, $\varepsilon 1 < \varepsilon 2$ so, $\frac{1}{2\varepsilon 1^2} > \frac{1}{2\varepsilon 2^2}$, therefore

$$\frac{1}{2\varepsilon 1^2} (\ln(|H| + \ln(\frac{1}{\delta})) > \frac{1}{2\varepsilon 2^2} (\ln(|H| + \ln(\frac{1}{\delta}))$$

Hence, $m(\varepsilon 1, \delta) > m(\varepsilon 2, \delta)$.

---

3 VC Dimension

1
a) Lower bound for VC(H) = 4.

Assume that x1, x2, x3, x4 $\in$ R such that x1 < x2 < x3x4.
Table of possibilities:

h(x1)	h(x2)	h(x3)	h(x4)	intervals
1	1	1	1	a < x1, b > x4
1	1	1	0	a <x1, b > x3
1	1	0	1	a < x1, b > x2 - c <x4, d > x4
1	1	0	0	a < x1, b > x2
1	0	1	1	a < x1, b > x1 - c < x3, d > x4
1	0	1	0	a < x1, b > x1 - c < x3, d > x3
1	0	0	1	a < x1, b > x1 - c < x4, d > x4
1	0	0	0	a < x1, b > x1
0	1	1	1	a < x2, b > x4
0	1	1	0	a < x2, b > x3
0	1	0	1	a < x2, b > x2 - c < x4, d > x4
0	1	0	0	a < x2, b > x2

This correctly classifies the subset of data for every partition with d = 4.

b)
Upper bound for VC(H) is 5.
Assume x1, x2, x3, x4, x5 $\in$ R,
The points can be split into subsets such that x1 < x2 < x3 <= x4 <= x5, x2 <= x3 <= x4 <= x5 <= x1, and so on.
We can always pick a partition of the data that cannot find two intervals that will correctly classify the data. This is independent of the subset we pick.
The table shows the partition that cannot be classified.
Here 1 is the lowest value and 5 is the greatest.

rank(xi)	1	2	3	4	5
h(xi)	1	0	1	0	1

If the data is partitioned this way, we cannot choose two intervals that correctly label the data. We can create such a partition for every subset from the instance space and we cannot shatter a dimension of 5 for this hypothesis space.
Hence, the upper bound for the VC-dimension is 5.

3)
The VC dimension of H is 2. We can show the proof of the statement as follows:
For any three points there must exist an arrangement where the hypothesis class is not able to shatte.
Lets consider two equivalence class:

1) 3 points placed in collinear fashion
2) 3 points placed in non collinear fashion

Case 1:
Since 3 points are placed in collinear fashion, the arrangement will be positive, negative, positive such that the hypothesis class cannot shatter these labeling.
Hence, in this particular case we can conclude that the equivalence is not shatterable.

Case 2:
Since 3points are placed in non collinear fashion, there will be an arrangement positive, negative, positive where negative will gall in the positive reason and thus this particular class is also not shatterable.
A Case where we consider two points. The given hypothesis class can shatter all the arrangements.
Hence the VC dimension is of H is 2.

4)
Consider any set of points N that are labelled G. The same set of points may not result in the same number of dichotomies by H as those produced by G.
The reason is that G may have hypotheses that are not present in H, whereas all hypothesis in H will be present in G. This means that G can be more expressive that H and can shatter more points. This means that $d_{VC}(H)$ <= $d_{VC}(G)$.

5)
Because H is a finite concept class and VC(H) = d, there exists d points that H can shatter.
We can arbitrarily give 0 or 1 label to each of the points, so there are 2^d ways to label them.
It does not matter how d points are labeled, there exists a hypothesis h ∈ H which can label them correctly.
H must consist of at least 2^d different hypothesis.
Hence, |H| >= 2^d.