<u>Machine Learning Homework 1</u>

Written by: Lior Shacohach - id - 206284960
Moriya Barel - id - 207304353

התרומה של כל אחד: עבדנו ביחד על כל השאלות.

<u>Question 1</u>

A) We have 6 areas, each of weight $\frac{\varepsilon}{6}$ -> the probability that a single point missed area 1 is $(1-\frac{\varepsilon}{6})$ -> the probability that all the sample points missed area 1 is $(1-\frac{\varepsilon}{6})^n$ which is approximated to $e^{-\frac{\varepsilon n}{6}}$.
Probability that the sample missed one of all areas ≤ $6e^{-\frac{\varepsilon n}{6}}$ (union bound).
We want $6e^{-\frac{\varepsilon n}{6}} \le \delta$,ergo $n \le 6\frac{ln(\frac{6}{\delta})}{\varepsilon}$.

B) The VC-dimension of 3 dimensional axis parallel rectangles is 6.
It can shatter some set of 6 points. Since we can't draw in 3d we will try to explain our thought process. Take the 2d version shown in class and add a fifth point at the middle of these 4 points, note that the 4 first points has the same z coordinate, the fifth point will have a different z value. For each of the labelings of the 4 points we can add the fifth point simply by stretching our 2d rectangle into a 3d version of the same rectangle. Now add a sixth point with z value equal to minus z value of the fifth point and extend rectangle to the other direction as well.

It cannot shatter any set of 7 points -
Proof: 2 cases
A - One point is in the convex hull of the others
B - The points are in general position
Part A - Assume one point is in the convex hull of the others:
A cube cannot label the outer points as + and the inner ones as - because the outer points must be inside the cube and since the cube is a convex object the inner point must be inside the cube as well.
Part B - general position:
Take the points with max-x, max-y, max-z, min-x, min-y and min-z.
Any rectangle that contains them must contain the seventh point as well

C) Define H to be our infinite set of 3d rectangles.

Let S be a sample drawn uniformly at random of size n.

Let $h \in (H)$ be a rule/rectangle from our set and $e(h)$ to be $h's$ error.

If h is consistent with S:

Take $e(h) = \varepsilon$

$\varepsilon \leq \frac{2}{n}(6\, log_2 \frac{2en}{6} + log_2 \frac{2}{\delta}) = \frac{2}{n}(6\, log_2 \frac{en}{3} + log_2 \frac{2}{\delta})$

If h is not consistent with S:

Take $e(h) = \varepsilon + \bar{e}(h)$

$\varepsilon + \bar{e}(h) \leq \bar{e}(h) + \sqrt{\frac{8*6ln\frac{2en}{6}+8ln\frac{4}{\delta}}{n}}$

$\varepsilon \leq \sqrt{\frac{48ln\frac{en}{3}+8ln\frac{4}{\delta}}{n}}$

D) We will compare $\varepsilon \leq \sqrt{\frac{48ln\frac{en}{3}+8ln\frac{4}{\delta}}{n}}$ with the bound of section A because both of them have empirical error on the sample.

First isolate ε from A bound so $n \leq 6\frac{ln(\frac{6}{\delta})}{\varepsilon}$ becomes $\varepsilon \leq 6\frac{ln(\frac{6}{\delta})}{n}$

Overall - $\varepsilon \leq 6\frac{ln(\frac{6}{\delta})}{n} \leq \sqrt{\frac{48ln\frac{en}{3}+8ln\frac{4}{\delta}}{n}}$

We can clearly see that the bound from section A is a lot tighter for several reasons -

- Same denominator
- Much bigger numerator in the VC bound
- Square root of a fraction enlarges it

Question 2

A) Total probability for A to win a round:
P(A Wins a given round) = P(A wins | B = 1) + P(A wins | B = 2) +…+ P(A wins | B = 6)

Probability for A to win a round assuming B rolled a 1:
P(A wins | B = 1) = P(A wins with both cubes ∪ A wins with one cube and lose with the other | B = 1)
P(A wins with both cubes | B = 1) = $5/6 * 5/6 * 1/6$ = 25/216
P(A wins with one cube and lose with the other | B = 1) =
$5/6 * 1/6 * 1/6 * 2 = 10/216$
P(A wins | B = 1) = $25/216 + 10/216 = 35/216$

Probability for A to win a round assuming B rolled a 2:
P(A wins | B = 2) = P(A wins with both cubes ∪ A wins with one cube and lose with the other | B = 2)
P(A wins with both cubes | B = 2) = $4/6 * 4/6 * 1/6 = 16/216$
P(A wins with one cube and lose with the other | B = 2) =
$4/6 * 2/6 * 1/6 * 2 = 16/216$
P(A wins | B = 2) = $16/216 + 16/216 = 32/216$

Probability for A to win a round assuming B rolled a 3:
P(A wins with both cubes | B = 3) = $3/6 * 3/6 * 1/6 = 9/216$
P(A wins with one cube and lose with the other | B = 3) =
$3/6 * 3/6 * 1/6 * 2 = 18/216$
P(A wins | B = 3) = $9/216 + 18/216 = 27/216$

Probability for A to win a round assuming B rolled a 4:
P(A wins with both cubes | B = 4) = $2/6 * 2/6 * 1/6 = 4/216$
P(A wins with one cube and lose with the other | B = 4) =
$2/6 * 4/6 * 1/6 * 2 = 16/216$
P(A wins | B = 4) = $4/216 + 16/216 = 20/216$

Probability for A to win a round assuming B rolled a 5:
P(A wins with both cubes | B = 5) = $1/6 * 1/6 * 1/6 = 1/216$
P(A wins with one cube and lose with the other | B = 4) =
$1/6 * 5/6 * 1/6 * 2 = 10/216$
P(A wins | B = 5) = $1/216 + 10/216 = 11/216$

Probability for A to win a round assuming B rolled a 6:
P(A wins | B = 6) = 0

Overall:

P(A wins a given round) = $(35 + 32 + 27 + 20 + 11 + 0)/216 = 125/216$

P(B wins a given round) = 1 - P(A wins a given round) = 1 - 125/216 = 91/216

Expected rounds A wins out of 100 -> E(A) = np = 100 * 125/216 = 57.8

Expected rounds B wins out of 100 -> E(B) = np = 100 * 91/216 = 42.1

B) Define:

$A_i := A\ wins\ the\ i'th\ round$ (A_i ~ Ber(p), p = 125/216)

$B_i := B\ wins\ the\ i'th\ round$ (B_i ~ Ber(p), p = 91/216)

$X := A_1 + A_2 + \ .. \ + A_{100}$

Markov:

$P(X \geq 65) \leq \ = \frac{E(X)}{65} \ = \ 57.8 / 65 \approx 0.89$

Hoeffding:

This time define - $X := (A_1 + A_2 + \ .. \ + A_{100}) / 100$

$P(X \geq 65 / 100) = P(X - E(X) \geq 0.65 - E(X)) \leq P(|X - E(X)| \geq 0.65 - E(X)) =$

$= P(|X - 0.57| \geq 0.08) \leq 2e^{-2*100*0.08^2} \approx 0.556$

C) Define - $X := \Sigma B_i$ and $\delta := 0.1$

$E(B_i) = \frac{91}{216} \ -> \ E(X) = \frac{91n}{216}$

$P(X \geq 35) = P(X - E(X) \geq 35 - E(X)) \leq P(|X - \frac{91n}{216}| \geq 35 - \frac{91n}{216}) \leq$

$\leq 2e^{-2*n*(35 - \frac{91n}{216})^2}$

Demand:

$2e^{-2n(35 - \frac{91n}{216})^2} \leq \delta = 0.1$

$ln(2e^{-2n(35 - \frac{91n}{216})^2}) \leq ln(0.1)$

$ln(2) + ln(e^{-2n(35 - \frac{91n}{216})^2}) \leq ln(0.1)$

$ln(2) - ln(0.1) - 2450n + \frac{3185n^2}{54} - \frac{8281n^3}{23328} \leq 0$

$82.75758 \leq n \leq 83.39503$

Consequently A,B needs to play at least 84 rounds.

Question 3

A) We want $e(h) \leq \varepsilon + 0.2$

And we know that $e(h) \leq \bar{e}(h) + \sqrt{\frac{ln(2(|H|)+ln\frac{1}{\ }}{2n}}$

So we bound $\bar{e}(h) + \sqrt{\frac{ln(2(|H|)+ln\frac{1}{\ }}{2n}} \leq \varepsilon + 0.2$

We know that $|H| = 2^{100}$ and that the empirical error is 20% so take $\bar{e}(h) = 0.2$

Overall we get:

$0.2 + \varepsilon \geq 0.2 + \sqrt{\frac{ln(2*2^{100})+ln\frac{1}{\ }}{2n}}$

$\varepsilon^2 \geq \frac{ln2^{101}+ln1-ln}{2n}$

$2\varepsilon^2 n \geq ln2^{101} - ln$

$ln \ \ \geq ln2^{101} - 2\varepsilon^2 n$

$\ \ \geq 2^{101}e^{-2\varepsilon^2 n} = \frac{2^{101}}{e^{2\varepsilon^2 n}}$

B) Take $\ \ = 0.1$ and $n = 1000$, we will do the same process like section a but isolate epsilon instead of $\ $.

$\varepsilon \geq \sqrt{\frac{ln2^{101}+ln\frac{1}{0.1}}{2*1000}}$

$\varepsilon \geq \sqrt{\frac{ln2^{101}+ln\frac{1}{0.1}}{2*1000}}$

$\varepsilon \geq 0.19$

C) The question is similar to the example given in class with the coffee house and the people drinking coffee/tea. We said that if the people of the ‘sample week’ are for whatever reason not the same people that will come afterwards (parallel to the different distribution), then we did not learn anything about these people and therefore cannot say anything about the true error of our program/rule on new people/points.