Quiz 4, Problem 1: Adaboost (50 points)

Part A: Venture Capitalism (38 points)

Congratulations—you've just won $10,000 in this month's lottery! You've heard that investing in startups is a good idea but aren't sure how to predict which startups will be successful. You decide to use Adaboost to classify upcoming startups as successful or not, so that you can make lots of money and stay on the cutting edge of new technology.

To begin, you look at six recent startups, noting some of their characteristics and whether they've become successful:

features

Startup ID	Name	Successful	Hires Business Majors	# of Team Members	Reached Kickstarter Goal	Founder from MIT
1	FaceStalk	**Yes**	Yes	5	Yes	No
2	CouchSurfer	**Yes**	No	6	No	Yes
3	Freeloader	**No**	Yes	4	No	Yes
4	MyFace	**No**	No	1	No	No
5	Googoo	**Yes**	No	3	No	Yes
6	NapApp	**No**	Yes	5	No	Yes

A1 (6 points) You've come up with several feature tests to help you predict whether a startup will succeed. For each of the tests in the table below, circle **all** the training points that the test **misclassifies**.

Test (if True, startup is Successful)	**Misclassified Training Points** (Circle all that apply)
Hires Business Majors = Yes	1 2 3 4 5 6
# of Team Members > 3.5	1 2 3 4 5 6
# of Team Members > 4.5	1 2 3 4 5 6

A2 (4 points) You've decided to use a different set of feature tests as weak classifiers to perform Adaboost. The classifiers and the errors they make are listed here.

Test ID	Test (if True, startup is Successful)	Misclassified Training Points
A	Reached Kickstarter Goal = Yes	2 5
B	# of Team Members < 3.5	4 5
C	Founder from MIT = Yes	1 3 6
D	Founder from MIT = No	2 4 5

note that this isn't true, but we are using this for the rest of the test

closest to 0, not furthest from 1/2

If you run Adaboost choosing **the weak classifier with the lowest error rate** in each round and breaking ties randomly, which tests from the table above would you **never** choose? (Circle **all answers** that apply)

A **B** **C** **D**

NONE OF THESE

(space to show work for Adaboost, part A3)

Errors from D: 2, 4, 5 } errors from D are a superset
Errors from A: 2, 5 } of errors from A

Could have looked @ B instead of A and reached the same conclusion

A3 (24 points) Perform two rounds of boosting using only the four weak classifiers (A, B, C, D) from part A2. In each round, *choose the weak classifier with the* **lowest error rate**. *In case of a* tie, choose the weak classifier that comes first **alphabetically**.

sum to 1/2

	Round 1	Round 2
$weight_1$	1/6	1/8
$weight_2$	1/6	1/4
$weight_3$	initialze as uniform 1/6	1/8
$weight_4$	1/6	1/8
$weight_5$	1/6	1/4
$weight_6$	1/6	1/8
Error rate of A	$\frac{1}{6}+\frac{1}{6} = \frac{2}{6}$ 2,5	$\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$ ✓ 2,5
Error rate of B	$\frac{1}{6}+\frac{1}{6} = \frac{2}{6}$ 4,5	$\frac{1}{8}+\frac{1}{4}=\frac{3}{8}$ 4,5
Error rate of C	$\frac{1}{6}+\frac{1}{6}+\frac{1}{6} = \frac{3}{6}$ 1,3,6	$\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$ 1,3,6
Error rate of D	$\frac{1}{6}+\frac{1}{6}+\frac{1}{6} = \frac{3}{6}$ 2,4,5	$\frac{1}{4}+\frac{1}{8}+\frac{1}{4}=\frac{5}{8}$ 2,4,5
weak classifier (h)	A (break AB tie)	B (break BC tie)
weak classifier error rate (ε)	2/6	3/8
voting power (α)	$\frac{1}{2}\ln(2)$	$\frac{1}{2}\ln\left(\frac{5}{3}\right)$

sum to 1/2

A4 (4 points) Here are the characteristics of a new startup called Glassr:

Name	Reached Kickstarter Goal	# Team Members	Founder from MIT
Glassr	No	3	Yes

According to the classifier you obtained from two rounds of boosting, should you invest in this startup? (That is, is it classified as Successful by your ensemble classifier?) (Circle one)

YES **NO**

$$H = \frac{1}{2}\ln(2)\left[\text{reached kickstarter}\right] + \frac{1}{2}\ln\left(\frac{5}{3}\right)\left[\#\text{ team members} < 3.5\right]$$

$\downarrow$ -1 $\downarrow$ -1

$H < 0$

Part B: Perfect Classifier (12 points)

Suppose you have six training points (P_1, P_2, P_3, P_4, P_5, P_6) and four weak classifiers (h_1, h_2, h_3, h_4), which make the following errors:

Classifier	Misclassified training points					
h_1	P_1		P_3	P_4		P_6
h_2		P_2			P_5	
h_3			P_3			
h_4				P_4		P_6

disjoint errors

B1 (5 points) Ben claims that by combining **<u>three</u>** of the weak classifiers above, he can construct an ensemble classifier H(x) that will correctly classify all the training data. If Ben is correct, list the three weak classifiers and assign them **integer voting powers** (α) to make a perfect ensemble classifier. If Ben is wrong, circle "CAN'T BE DONE" instead.

Weak classifier	Voting power
h_2	$\alpha =$ 1
h_3	$\alpha =$ 1
h_4	$\alpha =$ 1

CAN'T BE DONE

B2 (4 points) Alyssa claims that by combining **<u>two</u>** of the weak classifiers above, she can construct an ensemble classifier H(x) that will correctly classify all the training data. If Alyssa is correct, list the two weak classifiers and assign them **integer voting powers** (α) to make a perfect ensemble classifier. If Alyssa is wrong, circle "CAN'T BE DONE" instead.

If two tests disagree, cannot break tie

Weak classifier	Voting power
	$\alpha =$
	$\alpha =$

CAN'T BE DONE

B3 (3 points) What is the minimum number of rounds of boosting to produce a perfect ensemble classifier? If boosting will loop forever or terminate without producing a perfect classifier, circle CAN'T BE DONE instead.

Number of rounds: 3

CAN'T BE DONE

need at least 3 classifiers to break ties

Quiz 4, Problem 2: Bayesian inference (50 points)

Part A: *Senioritis* relapse (30 points)

Senioritis is a rare and treatable condition in general—however, here at MIT, it is positively epidemic, affecting 50% of the population. Experts have developed a cheap test for senioritis—the HACK scan—which is 80% sensitive and 60% specific. (This means that 80 out of 100 people with senioritis correctly test positive, and 60 out of every 100 people without senioritis correctly test negative. The HACK scan always reports either "positive" or "negative".)

Assume you are a typical member of the MIT population. For notation, we can let D be the variable "You have senioritis" and let T be the variable "You test positive for senioritis". Then the information above is:

$P(D)$	1 out of 2
$P(T \mid D)$	80 out of 100
$P(\overline{T} \mid \overline{D})$	60 out of 100

"nots" show up faintly online

Part A1 (10 points) What is the probability of obtaining a **negative test result**, regardless of whether you have senioritis?

The marginal probability of a negative test result is *approximately* (circle one):

0% 5% 10% 15% 20% 25% 30% 35% 40% 45% 50%

For credit, you must show your work. Write down the equations you intend to solve, if any, and indicate what values you're plugging in. You probably won't need a calculator, because you only need an approximate final answer.

Want the marginal $P(\overline{T})$

$$P(\overline{T}) = P(\overline{T}|D)\cdot P(D) + P(\overline{T}|\overline{D})P(\overline{D})$$

$$= \left(1 - \frac{80}{100}\right)\cdot \frac{1}{2} + \frac{60}{100}\cdot\frac{1}{2}$$

$$= \frac{10}{100} + \frac{30}{100}$$

$$= 40\%$$

Part A2 (10 points) Suppose your HACK scan returns a **negative result**. In this case, the probability that you indeed **don't have senioritis** is *most nearly* (circle one):

50% 55% 60% 65% 70% 75% 80% 85% 90% 95% 100%

<u>For credit, you must show your work.</u> Write down the equations you intend to solve, if any, and indicate what values you're plugging in. You probably won't need a calculator, because you only need an approximate final answer.

Want $$P(\bar{D} \mid \bar{T}) = \frac{P(\bar{T} \mid \bar{D}) P(\bar{D})}{P(\bar{T})}$$

$$= \frac{\frac{60}{100} \cdot \frac{1}{2}}{\frac{40}{100}}$$ from A1

$$= \frac{3}{4}$$

$$= 75\%$$

Part A3 (10 points) Out of a random sample of 100 MIT students, about how many of them are expected to be false negatives—that is, how many of them will both <u>have senioritis</u> and also <u>test negative</u>?

0 10 20 30 40 50 60 70 80 90 100

<u>For credit, you must show your work.</u> Write down the equations you intend to solve, if any, and indicate what values you're plugging in. You probably won't need a calculator, since you only need an approximate final answer.

Want $$P(D, \bar{T}) = P(\bar{T} \mid D) P(D)$$

$$= \left(1 - \frac{80}{100}\right) \cdot \frac{1}{2}$$

$$= \frac{10}{100}$$

$$= 10\%$$

Equivalently: $$P(D, \bar{T}) = P(D \mid \bar{T}) P(\bar{T})$$

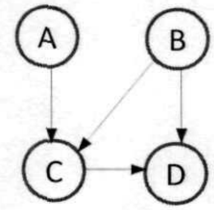
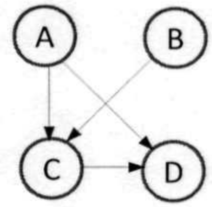
$$= (1 - .75) \cdot .4$$

$$= .1$$

Part B: This again, but different (16 points)

In the figure below, there are two Bayes nets and some independence statements. For each of the statements below and each Bayes net, circle TRUE if the statement is true for the net, and FALSE if the statement is false for the net.

Note: Assume that the only independence statements that are true are the ones enforced by the shape of the network.

		
A is independent of C.	TRUE FALSE	TRUE FALSE
Given C, A is independent of D.	TRUE FALSE	TRUE FALSE
P(B|DAC) = P(B|AC) aka $B \perp\!\!\!\perp D \mid AC$	TRUE FALSE	TRUE FALSE
Assuming all of the variables are boolean, how many parameters does each Bayes net have? (*The number of parameters is the total number of entries in all probability tables.*)		
# of parameters	$1+1+4+4=\boxed{10}$ A B C D	$1+1+4+4=\boxed{10}$ A B C D

Part C: What are the parameters in a binary net? (4 points)

Suppose you have training data, each with one feature X, and a classification Y. Both X and Y are boolean variables, meaning they can be either true or false. Consider the Naive Bayes classifier for this problem—which of the following probabilities are the parameters of the Naive Bayes model? (Circle **ALL** answers that apply, or circle "NONE OF THESE" instead.)

Hint: It may help to draw the Bayes net that corresponds to the Naive Bayes classifier for this problem.

$P(X)$ $P(X|Y)$ $P(X|\bar{Y})$ $P(X|Y)$ $P(X|Y)$ $P(X,Y)$ $P(X,Y)$

$P(Y)$ $P(Y|X)$ $P(Y|X)$ $P(Y|X)$ $P(Y|X)$ $P(X,\bar{Y})$ $P(\bar{X},Y)$

Bayes net

NONE OF THESE

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \propto P(X|Y)P(Y)$$

$$P(\bar{Y}|X) = \frac{P(X|\bar{Y})P(\bar{Y})}{P(X)} \propto P(X|\bar{Y})(1-P(Y))$$