6.034 Quiz 4
6 December 2013

Name	
email	

Circle your TA **(for 1 extra credit point)**, so that we can more easily enter your score in our records and return your quiz to you promptly.

Michael Fleder **Giuliano Giacaglia** **Dylan Holmes**

Casey McNamara **Robert McIntyre** **Duncan Townsend**

Mark Seifter **Sam Sinai** **Prashan Wanigasekara**

Problem number	Maximum	Score	Grader
1	50		
2	50		
Total	100		

SRN	6		

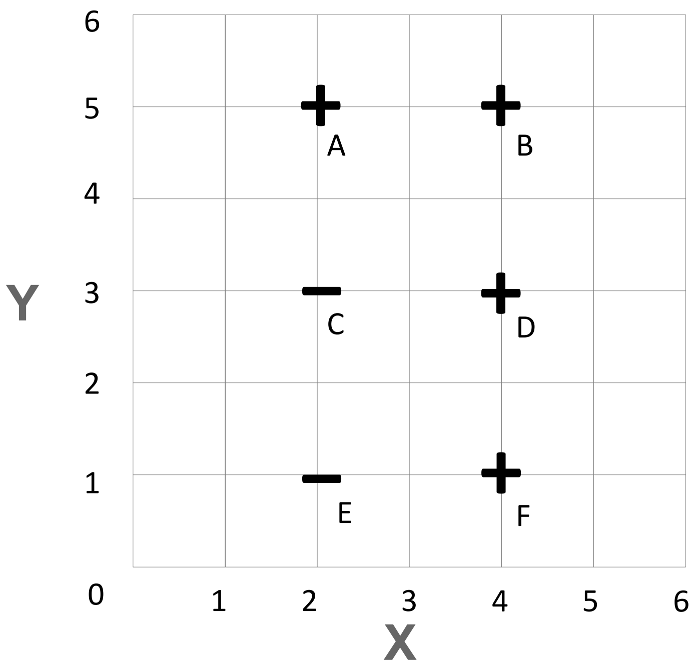
There are 10 pages in this quiz, including this one, but not including blank pages and tear-off sheets. As always, open book, open notes, open just about everything, including a calculator, but no computers.

Problem 1: Adaboost (50 points)

After attending a spectacular time traveler's convention, fearsome mathematician Ada Lovelace has hijacked a time machine, determined to investigate important developments in artificial intelligence.

Part A (50 points)

After attending an Adaboost lecture in 1995, Lovelace has rushed home to the 19th century to solve the following homework problem:



Classifier	Misclassified training points
X ≥ 3	
X ≥ 5	
Y ≥ 4	
Y ≥ 6	

<u>Reminder</u>: The classifier **X ≥ T** classifies a point as positive if it's to the *right* of a certain vertical line; otherwise, it classifies the point as negative.

Similarly, the classifier **Y ≥ T** classifies a point as positive if it's *above* a certain horizontal line; otherwise, it classifies the point as negative.

Feel free to draw these classifiers on the diagram.

A1 (4 points). Fill in the table of "Misclassified training points" above. For each weak classifier, list **<u>all</u>** the training points (A,B,C,D,E,F) which it misclassifies. If the classifier doesn't misclassify any points, write NONE instead.

A2 (24 points). Perform three rounds of boosting using the above training data and weak classifiers. In each round, choose the weak classifier with the lowest error rate. In case of a tie, choose the weak classifier that comes first in this list: **X≥3, X≥5, Y≥4, Y≥6**. You must use only these classifiers.

In any round, if boosting would terminate instead of choosing a classifier, write "NONE" for the weak classifier (h) and for the voting power (α). Then, leave all remaining spaces blank.

	Round #1	**Round #2**	**Round #3**
$weight_A$			
$weight_B$			
$weight_C$			
$weight_D$			
$weight_E$			
$weight_F$			
Error rate (ε) of X ≥ 3			
Error rate (ε) of X ≥ 5			
Error rate(ε) of Y ≥ 4			
Error rate (ε) of Y ≥ 6			
Weak classifier chosen this round (h)			
Voting power (α)			

You may show your work below for partial credit:

A3 (6 points). How would the ensemble classifier H produced after three rounds of boosting classify the following points? (Circle the best answer in each case.)

(X=0, Y=0)	POSITIVE	NEGATIVE	CAN'T BE DETERMINED
(X=0, Y=7)	POSITIVE	NEGATIVE	CAN'T BE DETERMINED
(X=7, Y=7)	POSITIVE	NEGATIVE	CAN'T BE DETERMINED

A4 (6 points). Suppose Lovelace decides to continue boosting until she completes a total of 2013 rounds. What classifier will she pick in the 2013th round?

(**Check the box for the correct answer, then fill in its blank.**)

☐ **Outcome A**: Adaboost must terminate at or before the 2013th round. In fact, it will terminate instead of picking a classifier in Round ______.

☐ **Outcome B:** Adaboost will reach the 2013th round without terminating. In the 2013th round, it will pick the weak classifier ______.

A5 (4 points). While keeping in touch with a brilliant yet harried cryptographer in the 20th century, Lovelace learns that there was a weak classifier missing from the original set. Apparently, the missing classifier misclassifies exactly three of the training points, and if you add it to the original list of classifiers, you can use boosting to classify the training data perfectly. (Note: The missing classifier could be any kind of classifier, not just a line classifier like **X ≥ T** or **Y ≥ T**)

Immediately, she deduces which three points it must have misclassified. Circle the **three** training points it must have misclassified.

A **B** **C** **D** **E** **F**

A6 (6 points). The cryptographer is quite surprised by these results. He remarks that since the missing classifier misclassifies three out of the six training points, it has an error rate of 50% and hence should be no better than a coin toss—it's useless as a weak classifier.

Explain what he means. (for example, briefly summarize what happens to the weights and to the voting power α if you pick a classifier with error rate of ½ during a round of boosting):

Based on the argument you just made, the cryptographer believes that Adaboost would never pick the missing classifier. Concisely explain why the Adaboost algorithm would, in fact, pick it.

Part B (bonus)

This is a bonus problem. Your score will be unaffected if you skip it or answer incorrectly. You can get +2 extra credit points if you answer correctly.

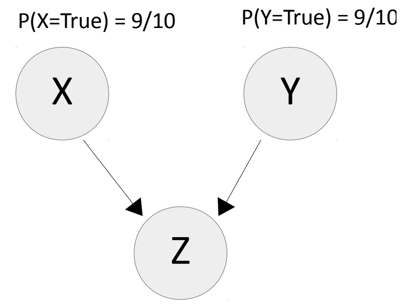
Having successfully solved a variety of boosting problems, Lovelace decides to supercharge her knowledge by kidnapping and interrogating a 21st century teaching assistant to check her work. (Needless to say, locked office doors are not a problem for time machines.)

Immediately, the TA points out that the following weights cannot possibly be correct, no matter what problem they're from. Why not?

	$weight_P$	**$weight_Q$**	**$weight_R$**	**$weight_S$**	**$weight_T$**	**$weight_U$**
Round ###	1/21	2/21	3/21	4/21	5/21	6/21

Problem 2: Bayesian Inference

Part A (5 points)



X	Y	P(Z=True|X,Y)
True	True	8/10
True	False	2/10
False	True	2/10
False	False	1/10

This Bayes Net depicts independence relations among three boolean variables X, Y, and Z. Consider the following probabilities:

(A) P(X = True | Y = True, Z = True)
(B) P(X = True | Y = False, Z = True)
(C) P(X = True | Z = True)

Your task is to arrange these probabilities from smallest to largest (Hint: this problem does not require significant calculation.) **Circle the best answer.**

A < B < C   B < A < C   C < A < B

A < C < B   B < C < A   C < B < A

Part B (45 points)

The *New York Times* has just acquired a batch of leaked NSA documents, and they have enlisted you to determine the security classification of each document: Secret (**S**), Restricted (**R**), or Lunch receipt (**L**).

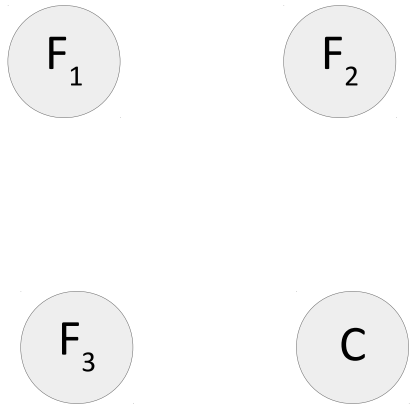
Based on your 6.034 expertise, you decide to build a Naïve Bayes classifier to do the job. First, you collect 600 previously leaked documents that are already marked as Secret (S), Restricted (R), or Lunch receipt (L). You notice several reoccurring keywords that could be relevant when determining classification:

	Keyword feature	# of **Secret** documents containing keyword	# of **Restricted** documents containing keyword	# of **Lunch receipt** documents containing keyword
F_1	"urgent"	70	50	10
F_2	"spaghetti"	1	2	50
F_3	"surveillance"	20	25	1

And the total number of documents with each security classification is:

Total # of **Secret** documents	Total # of **Restricted** documents	Total # of **Lunch receipt** documents
100	200	300

B1 (4 points): The unfinished Bayes net below has a node for each keyword feature F_1, F_2, F_3 and a node for the security classification C in {S,R,L}. Draw arrows between the nodes such that the arrows express the independence relations assumed when using a Naive Bayes classifier.



B2 (12 points): Based on the data in the tables above, compute estimates for the following probabilities (You may leave your answers in terms of fractions):

P(C = S)		P(C = R)		P(C = L)	
$P(F_1 \mid C = S)$		$P(F_1 \mid C = R)$		$P(F_1 \mid C = L)$	
$P(F_2 \mid C = S)$		$P(F_2 \mid C = R)$		$P(F_2 \mid C = L)$	
$P(F_3 \mid C = S)$		$P(F_3 \mid C = R)$		$P(F_3 \mid C = L)$	

Show your work for partial credit:

B3 (24 points): Now that you've collected your training data, you decide to look at the first unlabelled document from the NSA, which says:

> "**Urgent** notice: operation **spaghetti** is under **surveillance** by the 6.034 staff."
>
> (In other words, this document has F_1=true, F_2=true, F_3=true.)

In order to decide how to classify this document, let the variable Q denote the prior probability that a document contains all three keywords:

$$Q = P(F_1=\text{true}, F_2=\text{true}, F_3=\text{true}).$$

Compute the following conditional probabilities. Show your work for partial credit, and express your answers solely in terms of numbers/ratios and the variable Q. (You don't have to compute a numerical value for Q).

1. $P(C = S \mid F_1 = \text{true}, F_2 = \text{true}, F_3 = \text{true}) =$ ______

2. $P(C = R \mid F_1 = \text{true}, F_2 = \text{true}, F_3 = \text{true}) =$ ______

3. $P(C = L \mid F_1 = \text{true}, F_2 = \text{true}, F_3 = \text{true}) =$ ______

B4 (5 points). Based on your answers above, what security classification does your Naïve Bayes classifier assign to your document "Urgent notice: operation spaghetti is under surveillance by the 6.034 staff."? (**Circle one**.)

Secret	**R**estricted	**L**unch receipt	The classification cannot be determined with the given data.

↩ Note: The back of this page contains Problem 3, Spiritual and Right Now

Problem 3: Spiritual and Right-Now (6 points)

Circle the one **best** answer for each of the following questions. There is **no penalty for wrong answers**, so it pays to guess in the absence of knowledge.

1. Mansinghka demonstrated that probabilistic programs can:

    1. Open safes.
    2. Crack captchas.
    3. Predict the weather.
    4. Model the flight of bees.
    5. Explain irrational behavior.

2. Sinha's theory of autism suggests a correlation with:

    1. Malfunction of certain neurotransmitters.
    2. Inability to metabolize B vitamins.
    3. Reduced ability to habituate.
    4. Failure of brain centers involved in story understanding.
    5. Poverty.

3. Autistic children exhibit

    1. Attraction to insects.
    2. Fear of heights.
    3. Insensitivity to gaze cues.
    4. Curiosity about persistent loud noises.
    5. Voracious appetite.

4. The Genesis story understanding system:

    1. Performs listener aware story telling by restricting vocabulary to grade level.
    2. Summarizes stories by filling in templates, such as a template for earthquake reports.
    3. Uses if-then rules to build a graph of causal connections.
    4. Was motivated by the growing size of the game market.
    5. Uses crowd-sourcing to find concepts, such as *revenge*, in stories.

5. The Genesis story understanding system models:

    1. Different cultures with different sets of rules and concepts.
    2. Violent behavior as best explained via intrinsic personality traits.
    3. Story composition as a random walk through an elaboration graph.
    4. Personality differences using inference nets.
    5. Story understanding as a form of near-miss learning.

6. Cross-modal coupling:
    1. Uses Euclidean distance to merge regions.
    2. Uses vector dot product to determine distance.
    3. Rests on the premise that fairy tales link diverse cultures to common values.
    4. Uses labeled data to learn the sounds associated with vowels.
    5. Indicates the joint use of boosting and support vector machines.