6.034 Quiz 4

2 December 2015

Name	MAD HATTER
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.

Ryan Alexander	**Ben Greenberg**	**Neil Gurram**
Eeway Hsu	**Brittney Johnson**	**Veronica Lane**
Alice Liddell	**Robert Luo**	**Jessica Noss**
Mycal Tucker	**Sarah Vente**	**Jess Wass**

Problem number	Maximum	Score	Grader
1	50		
2	50		
Total	100		

SRN	6		

There are 8 pages in this quiz, including this one. As always, open book, open notes, open just about everything, including a calculator, but no computers.

Problem 1: Adaboost (50 points)

Part A: Furthest from ½ (35 points)

This is a boosting question with four training points (A, B, C, D). There are four weak classifiers (h_1, h_2, h_3, h_4) that make the following misclassifications:

Classifier	Misclassified training points (A, B, C, D)			
h_1	A			
h_2		B		
h_3			C	D
h_4	A	B	C	

A1 (23 points) Perform three rounds of boosting with these classifiers and training data. In each round, pick the classifier with the **error rate furthest from 1/2**. Break ties by picking the classifier that comes first in this list: h_1, h_2, h_3, h_4. Space for scratch work is provided on the following page.

In any round, if Adaboost 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	1/4	1/2 = 3/6	3/10
weight B	1/4	1/6	1/2 = 5/10
weight C	1/4	1/6	1/10
weight D	1/4	1/6	1/10
Error rate of h_1	1/4	1/2 = 3/6	3/10
Error rate of h_2	1/4	1/6	1/2 = 5/10
Error rate of h_3	2/4	2/6	2/10
Error rate of h_4	3/4	5/6	9/10
weak classifier chosen (h)	h_1	h_2	h_4 (or $-h_4$)
weak classifier error (ε)	1/4	1/6	9/10 (or 1/10)
voting power (α)	$\frac{1}{2}\ln 3$	$\frac{1}{2}\ln 5$	$\frac{1}{2}\ln\frac{1}{9} = -\frac{1}{2}\ln 9$ (or $\frac{1}{2}\ln 9$)

$\alpha = \frac{1}{2}\ln\left(\frac{1-\varepsilon}{\varepsilon}\right)$

$w_{new} = \frac{1}{2}\frac{1}{\varepsilon} w_{old}$ if incorrectly classified

$w_{new} = \frac{1}{2}\frac{1}{1-\varepsilon} w_{old}$ if correctly classified

Space provided for scratch work:

$H(x) = \frac{1}{2}\ln 3\, h_1 + \frac{1}{2}\ln 5\, h_2 - \frac{1}{2}\ln 9\, h_4$

A: $\text{sign}(\frac{1}{2}\ln 3 \cdot (-1) + \frac{1}{2}\ln 5 - \frac{1}{2}\ln 9 \cdot (-1))$

$= \text{sign}(\frac{1}{2}\ln(15)) > 0 \Rightarrow$ Correctly Classified

B: $\text{sign}(\frac{1}{2}\ln 3 + \frac{1}{2}\ln 5 \cdot (-1) - \frac{1}{2}\ln 9 \cdot (-1))$

$= \text{sign}(\frac{1}{2}\ln(\frac{27}{5})) > 0 \Rightarrow$ Correctly Classified

C: h_1, h_2 and $-h_4$ ~~both~~ all classify C correctly, so $\Rightarrow$ Correctly Classified

~~or for each of C, D: sign~~ or $\text{sign}(\frac{1}{2}\ln 3 + \frac{1}{2}\ln 5 - \frac{1}{2}\ln 9 \cdot (-1))$

$= \text{sign}(\frac{1}{2}\ln 135) > 0 \Rightarrow$ correctly classified

D: $\text{sign}(\frac{1}{2}\ln 3 + \frac{1}{2}\ln 5 - \frac{1}{2}\ln 9)$

$= \text{sign}(\frac{1}{2}\ln \frac{5}{3}) > 0 \Rightarrow$ Correctly classified

A2 (8 points) After three rounds of boosting, how does the ensemble classifier *H(x)* classify each training point? Circle the one best answer in each case. If the answer can't be determined from the available information, circle "Can't tell".

A:	Correctly classified	Misclassified	Can't tell
B:	Correctly classified	Misclassified	Can't tell
C:	Correctly classified	Misclassified	Can't tell
D:	Correctly classified	Misclassified	Can't tell

The following rules for manipulating logarithms may be helpful:

$\log(x) + \log(y) = \log(x * y)$ $\qquad \log(\frac{1}{x}) = -\log(x)$ $\qquad \log(1) = 0$

A3 (4 points) For each statement below, circle the one best answer (true or false).
(Hint: consider this problem.)

TRUE   FALSE   1. If there are exactly 3 weak classifiers that make non-overlapping errors, it is always possible to use them to construct a perfect ensemble classifier (either by using Adaboost or by picking values of α by hand).

Assign α's as 1.

TRUE   FALSE   2. Whenever possible, Adaboost will pick a classifier that makes non-overlapping errors with the previously picked classifiers.

Our example here had h_4 chosen despite overlapping errors with h_1 and h_2, even though h_3 had no overlapping errors

Part B: More True/False (15 points)

This section consists of questions about Adaboost *in general*—they do not rely on the preceding section. Decide whether each of the statements below is true or false. Circle the one best answer in each case.

TRUE   FALSE   1. In each round, the error rate ε of each weak classifier h is in the range 0 ≤ ε ≤ 0.5.

We had $\varepsilon = \frac{9}{10}$ for h_4 in round 3 in ~~example~~ our given problem

TRUE   FALSE   2. If a weak classifier has an error rate ε ≤ 1/3, it can only misclassify up to 1/3 of the training points.

We had $\varepsilon = \frac{2}{6}$ for h_3 in round 2 in example, but $2/4 = 1/2$ misc points

TRUE   FALSE   3. Given 3 classifiers, a weak classifier can never be picked two rounds in a row.

would mean $\varepsilon = \frac{1}{2}$ would have been chos second round $\Rightarrow \alpha = 0$.

TRUE   FALSE   4. The error rate of the ensemble classifier *H(x)* never increases from one round to the next.

Discussed in Lecture

error rate oscillates b is bounded by exponen

TRUE   FALSE   5. Adaboost terminates only if a specific number of rounds is specified.

We can stop if all points are correctly classified even if no limit on number of rounds is specified.

Problem 2: Bayes Nets

Part A: Bayes Nets (16 points)

For each Bayes Net described below, answer this question:

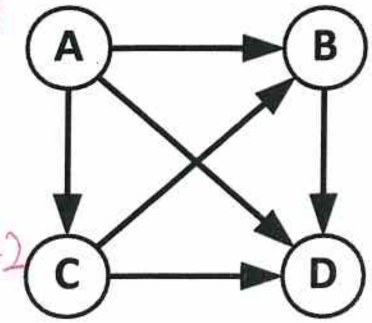
Assuming all of the variables are boolean, how many parameters does the Bayes net have?

(The number of parameters is the total number of entries in all probability tables.)

A1 (4 points)

P(A) 1

$4 = 2^2$ A | C | P(B|A,C)



A | P(C|A) $2^1 = 2$
T
F

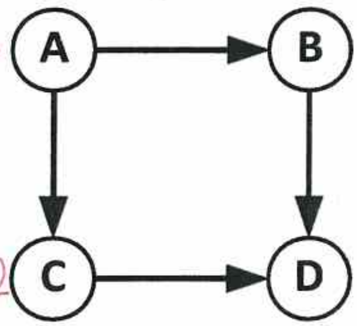
$8 = 2^3$ A | B | C | P(D|A,B,C)

1 + 2 + 4 + 8 = 15

Number of parameters = 15

A2 (4 points)

1 (A), $2^1 = 2$ (B), $2^1 = 2$ (C), $2^2 = 4$ (D)



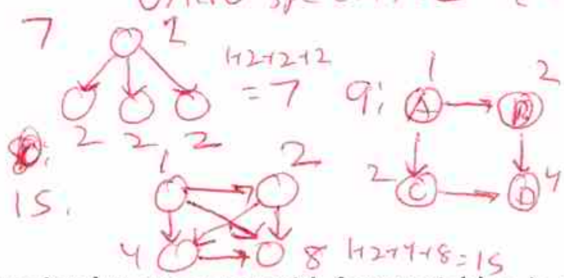
Number of parameters = 9

1 + 2 + 2 + 4 = 9

(similar way to construct truth tables as above)

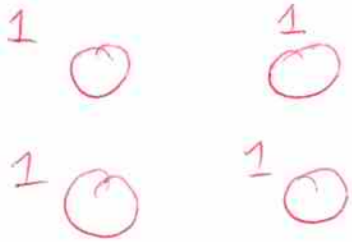
A3 (4 points) A Bayes net with four variables, in which NO variables are assumed to be independent. Underspecified can be anything from 7 to 15

e.g.: 7



Number of parameters = 7 to 15

A4 (4 points) A Bayes net with four variables, in which ALL variables are assumed to be independent.



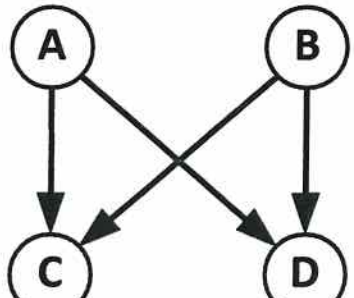
Number of parameters = 4

1+1+1+1 = 4

Part B: Some Probability (18 points)

Here is a Bayes net with 4 boolean variables and their associated probability tables. Each probability is represented by a lowercase variable. Below, show your work for partial credit.

P(A)
a



P(B)
b

A	B	P(C|AB)
T	T	p
T	F	q
F	T	r
F	F	s

A	B	P(D|AB)
T	T	w
T	F	x
F	T	y
F	F	z

B1 (8 points) Write an expression for $P(\bar{A}B\bar{C}\bar{D})$ in terms of the probabilities specified in the Bayes net.

Cross out variables by Bayes Assumption: ~~Children~~ $\perp\!\!\!\perp$ Non-Descendants Given Parents

$$P(\bar{A}B\bar{C}\bar{D}) = P(\bar{D}|\cancel{\bar{C}}B\bar{A})P(\bar{C}|B\bar{A})P(B|\cancel{\bar{A}})P(\bar{A}) = (1-y)(1-r)b(1-a)$$

Chain Rule used above

B2 (10 points) Write an expression for $P(B\bar{C}\bar{D})$ in terms of the probabilities specified in the Bayes net.

$$(1-w)(1-p)ba + (1-y)(1-r)b(1-a)$$

So, we see that ~~$B\bar{C}\bar{D}$~~ $\bar{C}$ and $\bar{D}$ depend on both A and B, but we don't have an A. So, we should marginalize by A through exhaustion.

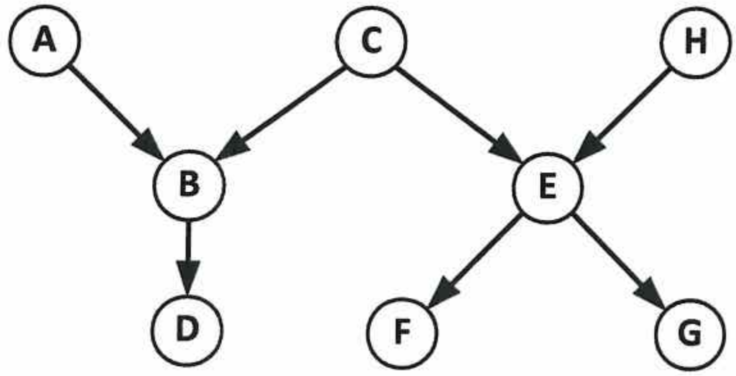
$$P(B\bar{C}\bar{D}) = \sum_{\text{all } A} P(AB\bar{C}\bar{D}) = \cancel{\sum_{\text{all } A} P(\bar{D}|BA)P(\bar{C}|}$$

$$= \sum_{\text{all } A} P(\bar{D}|BA)P(\bar{C}|BA)P(B)P(A) \quad \text{(same as express above)}$$

$$= P(\bar{D}|BA)P(\bar{C}|BA)P(B)P(A) + P(\bar{D}|B\bar{A})P(\bar{C}|B\bar{A})P(B)P(\bar{A}$$

Part C: Independence (16 points)

Here is a Bayes net with 8 variables, which are not necessarily boolean. Assume that the only independence statements that are true are the ones enforced by the shape of the network. Show your work for partial credit.



C1 (6 points) List every variable that is marginally independent of A.

C, H, E, F, G

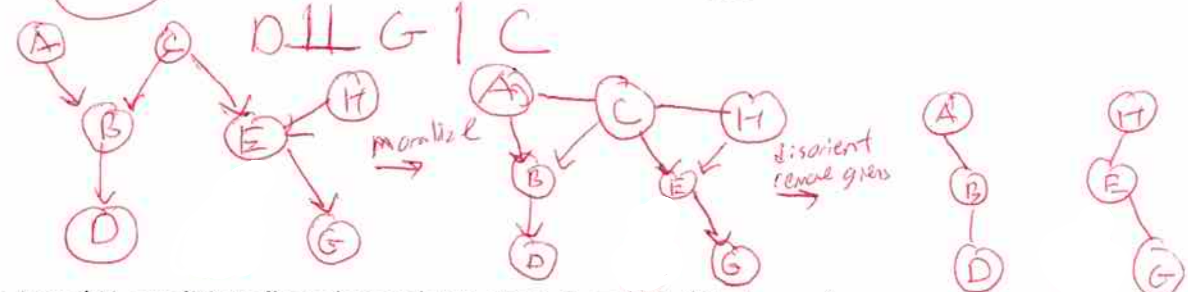
a variable v is marginally independent of A if after drawing ancestral graph of v and A, there is no path between v and A. (by d-separation)

C2 (5 points) Is P(D|C) = P(D|CG)? (Circle one)

YES NO

$D \perp\!\!\!\perp G \mid C$

ancestral graph

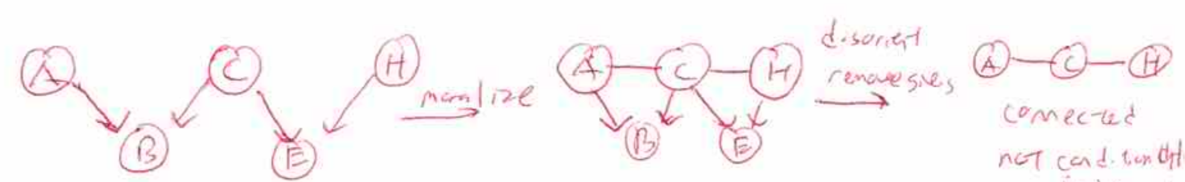


C3 (5 points) Are A and H conditionally independent, given B and E? (Circle one)

YES NO

$A \perp\!\!\!\perp H \mid B, E$

ancestral graph



(SRN QUESTIONS ON NEXT PAGE →)

Problem 3: Spiritual and Right-Now

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 According to Winston, Minsky believes that we will not really understand intelligence until:

1. We develop provably a universal representation.
2. We develop a provably universal theorem prover.
3. We build robots that operate in the real world.
4. We understand how humans deploy many kinds of reasoning.
5. We develop intelligent machines that can pass the Turing test.

2 Jegelka noted that submodularity has been used with success to:

1. Analyze effects of vaccination programs.
2. Place water-quality sensors.
3. Improve race-horse breeding programs.
4. Program vacuum-cleaner robots.
5. Develop proofs of program correctness.

3 Winston described how the Genesis story-understanding system:

1. Tells stories persuasively by filtering content.
2. Uses Bayesian inference to infer conceptual content.
3. Deploys machine-learning techniques to cluster stories into tragedies and comedies.
4. Authors new stories by combining precedents using a support vector machine.
5. Extends its story database by crawling the web.

4 Sinha explained that a newly sighted adult:

1. Initially sees in red-green, and only later gains the ability to see blue.
2. Can be modeled by a convolutional neural net with few recurrent connections.
3. Tends to see more objects in drawings than a normally sighted person would.
4. Has trouble identifying moving elements in drawings.
5. Exhibits the same kind of errors seen in classifiers based on deep neural net technology.

5 Mansinghka explained how probabilistic programming has been used to:

1. Score SAT essays.
2. Place bets on fantasy sports sites such as Draft Kings.
3. Select answers on 6.034 SRN questions.
4. Guide self-driving cars.
5. Recognize faces in various postures.

6 Katz noted that his START system translates English questions into:

1. Statistical inference models.
2. Sets of ternary expressions.
3. Logical statements in predicate calculus.
4. An inner language based on Esperanto.
5. Transition space diagrams.