6.034 Quiz 4

4 December 2019

Name	SOLUTIONS
Email	

For 1 extra credit point: Circle the TA whose recitations you attend so that we can more easily enter your score in our records and return your quiz to you promptly.

Sydney Gibson	**Rui Li**	**Allison Tam**
Udgam Goyal	**Jennifer Madiedo**	**Héctor Vazquez**
Jenna Hong	**Jack Murphy**	**Eric Wong**
Damon Jones	**Mira Partha**	

Problem number	Maximum	Score	Grader
1 - Bayes	50		
2 - Boosting	50		
Total	**100**		

SRN	7		

There are **14** pages in this quiz, including this one, but not including tear-off sheets. Tear-off sheets with duplicate drawings and data are located after the final page of the quiz. **We do not collect tear-off sheets, so please show your work on the quiz pages, not the tear-off sheets.**

As always, the quiz is open book, open notes, open just about everything, including a calculator, but no computers or cell phones.

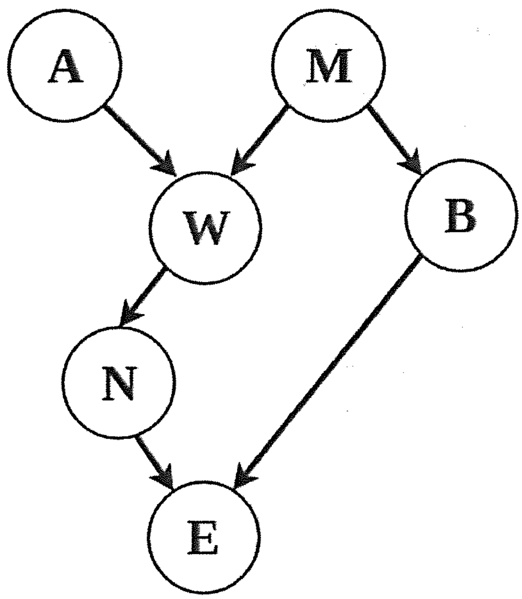
Problem 1: Bayesian Inference (50 points)

Detective Sherlock Holmes has discovered that his partner, Dr. John Watson, has gone missing. Holmes suspects foul play and decides to investigate Watson's disappearance using a BayesNet.

The following (Boolean) variables are in the BayesNet:

- **A**: Irene **A**dler is acting suspiciously.
- **M**: Professor **M**oriarty is acting suspiciously.
- **W**: **W**atson has been kidnapped.
- **B**: Professor Moriarty is robbing a **b**ank.
- **N**: Watson left a **n**ote for Holmes on their dining table.
- **E**: The note on the dining table is **e**ncrypted.

Holmes devises the following BayesNet:



For your convenience, a copy of the network is provided on a tear-off sheet.

Part A: Where's Watson? (30 points)

A1 (2 points) Using the BayesNet assumption, simplify the following expression. Write your answer in the box provided. If the expression cannot be simplified, write "Not possible" in the box.

$$P\,(\mathrm{N} \mid \mathrm{AMWB})$$

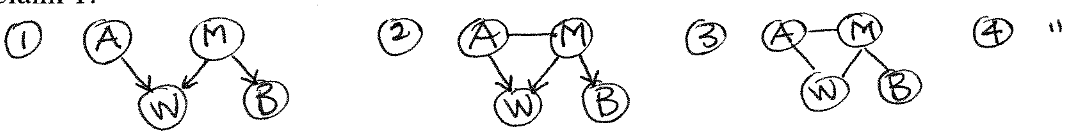
$\boxed{P(N|W)}$

A2 (20 points) Holmes wants to use d-separation to investigate the following claims. For each claim, circle the one best answer.

Claim 1:	$W \perp\!\!\!\perp B$	True	False	Can't Tell
Claim 2:	$W \perp\!\!\!\perp B \mid AM$	True	False	Can't Tell
Claim 3:	$W \perp\!\!\!\perp B \mid N$	True	False	Can't Tell
Claim 4:	$W \perp\!\!\!\perp B \mid E$	True	False	Can't Tell

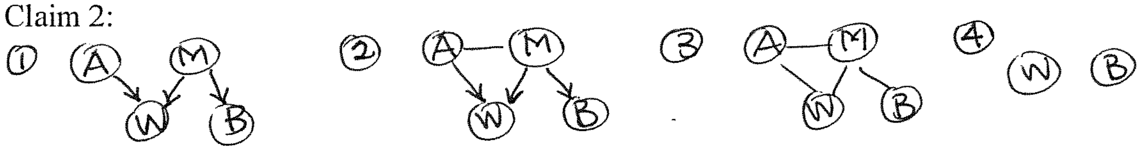
For partial credit, show your work in the boxes below.

Claim 1:



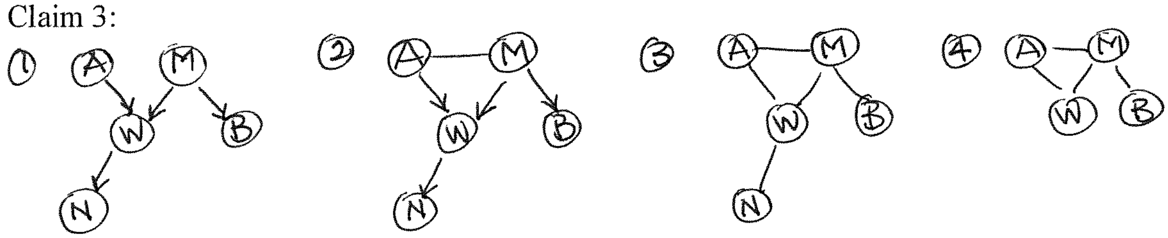
⑤ Not independent

Claim 2:



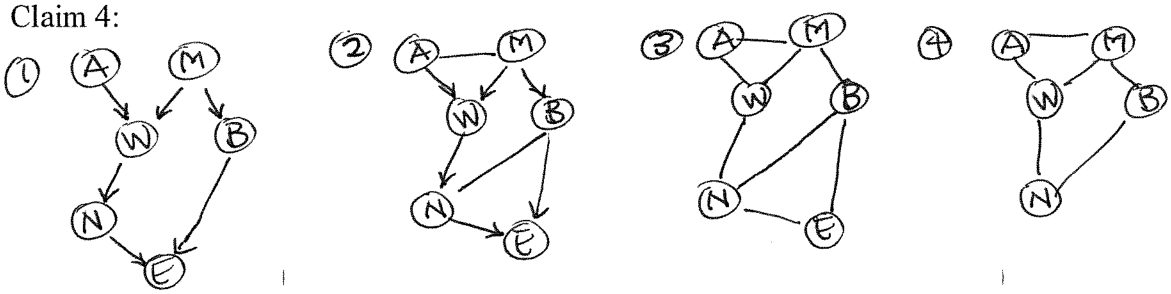
⑤ independent

Claim 3:



⑤ Not independent

Claim 4:



⑤ Not independent

A3 (2 points) Trying to deduce the culprit, Holmes considers the probability that Moriarty is acting suspiciously, given that Watson has been kidnapped: $P(\mathbf{M}\text{=true} \mid \mathbf{W}\text{=true})$. He then discovers that Irene Adler has **not** been acting suspiciously: **A**=false. How does the **conditional probability change** with the new given value? Circle the one best answer.

a. $P(\mathbf{M}\text{=true} \mid \mathbf{A}\text{=false}, \mathbf{W}\text{=true}) > P(\mathbf{M}\text{=true} \mid \mathbf{W}\text{=true})$

b. $P(\mathbf{M}\text{=true} \mid \mathbf{A}\text{=false}, \mathbf{W}\text{=true}) < P(\mathbf{M}\text{=true} \mid \mathbf{W}\text{=true})$

c. $P(\mathbf{M}\text{=true} \mid \mathbf{A}\text{=false}, \mathbf{W}\text{=true}) = P(\mathbf{M}\text{=true} \mid \mathbf{W}\text{=true})$

d. Can't Tell

A4 (6 points) Given the independence relations shown in the BayesNet, how many parameters does Holmes need to fully describe this system?

$\boxed{14}$

How many parameters would Holmes need if no independence relations were known? (An unsimplified answer is fine.)

$\boxed{2^6-1 = 63}$

Part B: Bayes to the Rescue (20 points)

Holmes discovers that his arch-nemesis, Professor Moriarty, is holding Watson hostage. Moriarty challenges Holmes to the following coin game for Watson's freedom.

Moriarty has two coins, a fair coin and a biased coin. He will select a coin and flip it 3 times. Holmes's objective is to figure out, based on the observed sequence of flips, which coin Moriarty chose. Holmes is given the following information about the probabilities of each coin landing on Heads (**H**) or Tails (**T**).

Coin 1 (fair)	**Coin 2 (biased)**
$P(\text{H}) = 0.5$ $\quad P(\text{T}) = 0.5$	$P(\text{H}) = 0.8$ $\quad P(\text{T}) = 0.2$

P (Moriarty selects Coin 1) = 0.6
P (Moriarty selects Coin 2) = 0.4

Moriarty secretly chooses one of the coins and flips it three times. Holmes and Moriarty observe the following sequence of flips: **HHT**

B1 (8 points) Calculate the following probabilities. Leave the marginal probability of observing the given sequence, P(HHT), uncalculated in your answer.

P (Coin 1 | HHT) = $\boxed{\frac{0.075}{P(HHT)}}$

P (Coin 2 | HHT) = $\boxed{\frac{0.0512}{P(HHT)}}$

For partial credit, show your work in the box below.

$$P(\text{coin } 1 | HHT) = \frac{P(HHT | \text{coin}1) * P(\text{coin}1)}{P(HHT)} = \frac{(0.5 * 0.5 * 0.5)(0.6)}{P(HHT)}$$

$$= \frac{0.075}{P(HHT)}$$

$$P(\text{coin}2 | HHT) = \frac{P(HHT | \text{coin}2) * P(\text{coin}2)}{P(HHT)} = \frac{(0.8 * 0.8 * 0.2)(0.4)}{P(HHT)}$$

$$= \frac{0.0512}{P(HHT)}$$

B2 (2 points) Given the observed data, which coin did Moriarty most likely select? Circle your answer.

Coin 1   Coin 2

B3 (8 points) Wanting to further test Holmes's powers of deduction, Moriarty flips the coin a fourth time. The full flip sequence is now: **HHTH**

Calculate the following updated probabilities. Leave the marginal probability of observing the given sequence, *P*(HHTH), uncalculated in your answer.

$$P(\text{Coin 1} \mid \text{HHTH}) = \boxed{\frac{0.0375}{P(HHTH)}}$$

$$P(\text{Coin 2} \mid \text{HHTH}) = \boxed{\frac{0.04096}{P(HHTH)}}$$

For partial credit, show your work in the box below.

$$P(coin 1 \mid HHTH) = \frac{P(HHTH \mid coin1) * P(coin1)}{P(HHTH)} = \frac{(0.5^4)(0.6)}{P(HHTH)}$$

$$= \frac{0.0375}{P(HHTH)}$$

$$P(coin 2 \mid HHTH) = \frac{P(HHTH \mid coin2) * P(coin2)}{P(HHTH)} = \frac{(0.8^3 \cdot 0.2)(0.4)}{P(HHTH)}$$

$$= \frac{0.04096}{P(HHTH)}$$

B4 (2 points) Given the observed data, which coin did Moriarty most likely select? Circle your answer.

Coin 1   Coin 2

This page is intentionally blank (except for this sentence).

Problem 2: Friends Who Boost Together (50 points)

Part A: The One with the Thanksgiving (35 points)

For Thanksgiving dinner this year, Monica decides to go all out. Her friends place bets on which parts of the dinner will go wrong: Turkey (**T**), Potatoes (**P**), Gravy (**G**), Cranberry sauce (**C**), Stuffing (**S**), or Dessert (**D**). Treating each friend as a **weak binary classifier**, the table below shows where their predictions went wrong. Correct predictions are +; incorrect predictions are -.

	T	P	G	C	S	D
h1 (Ross)	-	+	-	-	-	+
h2 (Rachel)	+	+	-	+	+	-
h3 (Chandler)	-	+	+	-	+	-
h4 (Phoebe)	-	+	+	-	+	+
h5 (Joey)	+	+	-	-	-	+
Actual	+	+	+	-	+	-

A1 (2 points) Fill out the table below to show each friend's misclassifications.

	Misclassified Training Points
h1	T, G, S, D
h2	G, C
h3	T
h4	T, D
h5	G, S, D

A2 (30 points) Fill out the table below to complete the first 3 rounds of Adaboost, choosing the classifier with the error rate **furthest from ½**. Break ties according to which classifier comes first in this list: **h1, h2, h3, h4, h5.**

For partial credit, show your work on the next page.

	Round 1	**Round 2**	**Round 3**
Weight of T (w_T)	$1/6$	$1/2$	$5/16$
Weight of P (w_P)	$1/6$	$1/10$	$1/4$
Weight of G (w_G)	$1/6$	$1/10$	$1/16$
Weight of C (w_C)	$1/6$	$1/10$	$1/4$
Weight of S (w_S)	$1/6$	$1/10$	$1/16$
Weight of D (w_D)	$1/6$	$1/10$	$1/16$
Error rate of h1 (ε_{h1})	$4/6$	$8/10$	$1/2$
Error rate of h2 (ε_{h2})	$2/6$	$2/10$	$5/16$
Error rate of h3 (ε_{h3})	$1/6$	$5/10$	$5/16$
Error rate of h4 (ε_{h4})	$2/6$	$6/10$	$6/16$
Error rate of h5 (ε_{h5})	$3/6$	$3/10$	$3/16$
Chosen classifier (h)	h_3	h_1	h_5
Error rate of classifier (ε)	$1/6$	$4/5$	$3/16$
Voting power (α)	$\frac{1}{2}\ln(5)$	$\frac{1}{2}\ln(1/4)$	$-\frac{1}{2}\ln(3/13)$

or $-\frac{1}{2}\ln(4)$ or $\frac{1}{2}\ln(13/3)$

For partial credit on A2, you can show your work in the box below.

R1 Initial weights : $1/6$ T G S D

error rates = Σ misclassified w_i $\quad \varepsilon_{h_1} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6}$ etc.

choose classifier w/ ε farthest from $1/2$ $\rightarrow h_3$

$$\alpha = \frac{1}{2}\ln\left(\frac{1-\varepsilon}{\varepsilon}\right) = \frac{1}{2}\ln\left(\frac{1-\frac{1}{6}}{\frac{1}{6}}\right) = \frac{1}{2}\ln 5.$$

R2 update weights : $w' = \begin{cases} 1/2 \left(\frac{1}{1-\varepsilon}\right) w_{old} & \text{if correctly classified} \\ 1/2 \left(\frac{1}{\varepsilon}\right) w_{old} & \text{if incorrectly classified} \end{cases}$

OR : $W_T = \frac{1}{2}$

rest of the weights are equally proportioned : $\frac{1/2}{5} = \frac{1}{10}$.

$\varepsilon = \sum_{\text{missed}} w_i$

classifier : h_1 , $\alpha = \frac{1}{2}\ln\left(\frac{1-\frac{4}{5}}{\frac{4}{5}}\right) = \frac{1}{2}\ln\left(\frac{1}{4}\right) = -\frac{1}{2}\ln(4)$

R3 $W_T + W_G + W_S + W_D = \frac{1}{2}$; $W_P + W_C = \frac{1}{2}$

ratio 5 : 1 : 1 : 1 $\qquad$ 1 : 1

A3 (3 points) What ensemble classifier **H(x)** would you generate after 3 full rounds of Adaboost?

$$H(x) = \frac{1}{2}\ln(5)h_3 + \frac{1}{2}\ln(1/4)h_1 - \frac{1}{2}\ln(3/13)h_5$$

or SIGN(" " ")

Part B: Central Perk (5 points)

Joey is attempting to settle the hotly debated argument: *Did Ross cheat on Rachel, or were they "on a break"?* Joey identifies seven key events **(A-G)** as training points and chooses 3 random people **(P_1, P_2, P_3)** to classify each event as having happened or not. He treats each person as a **weak classifier** and records their misclassifications in the table below.

Classifier	Misclassifications
Person 1 (P_1)	A, D
Person 2 (P_2)	B, E, G
Person 3 (P_3)	C, F

B1 (3 points) Joey constructs the following classifier:

$$H(x) = 3*P_2(x) + 4*P_3(x)$$

Is this a perfect classifier (i.e., correctly classifies all training points)? Circle YES or NO. If NO, also write the misclassified training points in the box provided.

YES **NO, H(x) misclassifies:** C, F

B2 (2 points) Assume Person 1 misclassified A, D, and **one** additional training point (B, C, E, F, or G). In this case, is it possible to create a perfect classifier by running Adaboost? Circle one.

YES **NO** **CAN'T TELL**

Part C: Smelly Cat 2.0 (10 points)

Phoebe is writing a song about Adaboost, but she got her facts all mixed up. For each claim below, either **write True,** or **write False and explain in <u>10 words or less</u> why the claim is false.**

1. Adaboost is fast and prone to overfitting.

False: Adaboost is robust to overfitting; generalizes well with each iteration/round.

2. When choosing classifiers with the smallest error rate less than ½, the weights of all training points in a particular round can be negative.

False: Weights of training points cannot be negative.

3. If more than two weak classifiers misclassify the same training point, the created classifier cannot be a perfect classifier.

False: could be a perfect classifier if other weak classifiers (which correctly classify the training point) have larger voting power.

4. If you let Adaboost run on any dataset, it will not terminate until instructed to do so.

False: Adaboost can terminate if all points are correctly classified or if no weak classifiers are "good enough".

5. The following is a possible set of weights for an unknown round of Adaboost: 1/10, 1/10, 2/10, 6/10.

False: These weights are not possible; no combination of weights adds up to ½.

Problem 3: Spiritual and Right Now (6 points + 1 bonus)

For each question, **write in the box** provided the letter corresponding to the **one** best answer **and circle** the answer. There is **no penalty for wrong answers,** so it pays to guess in the absence of knowledge.

B

1. Davis explained that the problem with the general problem solver (GPS) was:
   A. It couldn't generalize.
   B. It didn't have domain-specific knowledge for problems.
   C. There was no way to quantify the difference between two states.
   D. It had no sense of universal subgoaling.

B

2. According to Pratt, which of the following statements is **TRUE?**
   A. Only fully autonomous cars can save lives.
   B. The hardest part of autonomous vehicle research is prediction.
   C. Billions of miles of simulated driving data solves the need for training data.
   D. The main motivators behind autonomous driving research are safety and traffic mitigation.

D

3. What did Katz mean when he said the human vision system is biased?
   A. Humans recognize other humans faster than they recognize animals.
   B. Humans recognize objects faster than they recognize images of objects.
   C. Humans recognize ImageNet images more accurately than machines recognize ImageNet images.
   D. Humans recognize objects that make sense in a scene.

D

4. According to Berwick, what is the critical step that occurs uniquely in human children aged 3-4?
   A. They learn to emulate the cadence of their native language.
   B. They understand words in their native language.
   C. They learn to string together words in their native language to ask for things they want.
   D. They understand and can create hierarchical phrases in their native language.

A

5. Which of the following was **NOT** an example Holmes discussed as an application of concept patterns in the Genesis story understanding system?
   A. Concept patterns enable alternate endings to be generated from a story.
   B. Concept patterns enable comparison between stories.
   C. Concept patterns enable different viewpoints of the same story.
   D. Concept patterns enable summarization of stories.

C

6. Holmes illustrated the Genesis system's capabilities with a story about:
   A. Henry VIII
   B. Graduate advising
   C. International conflict
   D. A criminal and a policeman

A

7. Davis mentioned which companies that developed technology and let consumers figure out applications?
   A. Apple and Boston Dynamics
   B. iRobot and SharkNinja
   C. Android and Netflix
   D. Bank of America and GE