# 6.034 Quiz 4

# 5 December 2018

| | |
|---|---|
| Name | JUSTIN TIMBERLAKE |
| Email | nsync@mit.edu |

**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.

**Suri Bandler**

**Sanchit Bhattacharjee**

**Alex Charidis**

**Samir Dutta**

**Marie Feng**

**Ariel Jacobs**

**Victoria Longe**

**Smriti Pramanick**

**Kifle Woldu**

**Matt Wu**

**Richard Yip**

| Problem number | Maximum | Score | Grader |
|---|---|---|---|
| 1 - Adaboost | 56 | | |
| 2 – Bayes Nets | 44 | | |
| Total | 100 | | |

| | | | |
|---|---|---|---|
| SRN | 6 | | |

There are 12 pages in this quiz, including this one, but not including tear-off sheets. Tear-off sheets with duplicate drawings are located after the final page of the quiz.

As always, open book, open notes, open just about everything, including a calculator, but no computers.

This page is intentionally blank.

# Problem 1: Adaboost (56 points)

## Part A: Six of Crows (40 points)

Six (6) students from the bestselling book *Six of Crows*—**R**achel (R), H**O**lden (O), **M**ahgi (M), **T**asha (T), **K**yle (K), and **S**car (S)—are headed back to the Ketterdam University District after an evening out.

The students want to use Adaboost to determine if they will each arrive home safely, and they poll patrons at five (5) establishments about how safe each of their trips home would be. For each student, the patrons at each establishment act as a weak classifier and predict the student's trip home as safe (yes) or not safe (no). The resulting **five (5) weak classifiers**—Kaelish Prince, **KP(x)**; Emerald Palace, **EP(x)**; Crow Club, **CC(x)**; Blue Paradise, **BP(x)**; The Sweet Shop, **SS(x)**—make the following misclassifications:

| *Weak Classifiers* | *Misclassified Training Points* |
|---|---|
| KP(x) | R, O, M |
| EP(x) | R, O |
| CC(x) | M |
| BP(x) | T, K, S |
| SS(x) | O, M, T |

**For your convenience, the table is provided on a tear-off sheet after the last page of the quiz.**

**A1 (30 points)** On the next page, complete the first three (3) rounds of Adaboost,

**A1 (30 points)** Fill out the table below to complete the first three (3) rounds of Adaboost, choosing the classifier with the **error rate furthest from ½**. Some of the values have been filled in for you.

Break ties according to which classifier comes first in this list: KP(x), EP(x), CC(x), BP(x), SS(x).

You can show your work for partial credit on the next page.

| | Round 1 | Round 2 | Round 3 |
|---|---|---|---|
| Weight of R $(w_R)$ | $1/6$ | $1/10$ x | $\frac{4}{16} = \frac{1}{4}$ |
| Weight of O $(w_O)$ | $1/6$ | $1/10$ x | $\frac{4}{16} = \frac{1}{4}$ |
| Weight of M $(w_M)$ | $1/6$ x | $\frac{5}{10} = 1/2$ | $\frac{5}{16}$ |
| Weight of T $(w_T)$ | $1/6$ | $1/10$ | $1/16$ |
| Weight of K $(w_K)$ | $1/6$ | $1/10$ | $1/16$ |
| Weight of S $(w_S)$ | $1/6$ | $1/10$ | $1/16$ |
| Error rate of KP(x) $(\varepsilon_{KP})$ | $\frac{3}{6} = 1/2$ | $7/10$ | $13/16$ |
| Error rate of EP(x) $(\varepsilon_{EP})$ | $\frac{2}{6} = 1/3$ | $2/10 = 1/5$ | $1/2$ |
| Error rate of CC(x) $(\varepsilon_{CC})$ | $1/6$ | $\frac{5}{10} = 1/2$ | $5/16$ |
| Error rate of BP(x) $(\varepsilon_{BP})$ | $\frac{3}{6} = 1/2$ | $3/10$ | $3/16$ |
| Error rate of SS(x) $(\varepsilon_{SS})$ | $\frac{3}{6} = 1/2$ | $7/10$ | $5/8$ |
| Weak classifier chosen (h) | CC(x) | EP(x) | KP(x) |
| Weak classifier error (ε) | $1/6$ | $1/5$ | $13/16$ |
| Voting power (α) | $\frac{1}{2}\ln(5)$ | $\frac{1}{2}\ln(4)$ | $\frac{1}{2}\ln\left(\frac{3}{13}\right) = -\frac{1}{2}\ln\left(\frac{13}{3}\right)$ |

$\frac{1}{2}\ln\left(\frac{5/6}{1/6}\right)$ $\qquad$ $\frac{1}{2}\ln\left(\frac{4/5}{1/5}\right)$

**The following rules for manipulating logarithms may be helpful.**

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

*For partial credit on A2, you can show your work here.*

Round 1: initialize weights to $1/N \rightarrow 1/6$.

Error rate for each $h = \sum_{\substack{\text{misclassified} \\ \text{points, } i \\ \text{by } h.}} W_i$

cc(x) only misclassifies m, $\varepsilon = 1/6$.

$\rightarrow \alpha_1 = \frac{1}{2}\ln\left(\frac{1-\varepsilon}{\varepsilon}\right) = \frac{1}{2}\ln\left(\frac{1-1/6}{1/6}\right) = \frac{1}{2}\ln\left(\frac{5/6}{1/6}\right) = \frac{1}{2}\ln(5)$.

Round 2:

update weights: incorrect $\rightarrow$ just m $\Rightarrow W_m = 1/2$ (Alt: $\frac{1}{2}\cdot\frac{1}{6}\cdot\frac{1}{1/6} = \frac{1}{2}\cdot W_{old}\cdot\frac{1}{\varepsilon}$)

correct $\rightarrow$ R, O, T, K, S in equal ratio in round 1 $\Rightarrow \frac{\frac{1}{2}}{5} = \frac{1}{10}$ each.

1:1:1:1:1

$\hookrightarrow$ Alt: $\frac{1}{2}\cdot W_{old}\cdot\frac{1}{1-\varepsilon} = \frac{1}{2}$

Round 3:

update weights: incorrect $\rightarrow$ R, O

correct $\rightarrow$ m, T, K, S $\Rightarrow \frac{5}{16}, 1/16, 1/16, 1/16$.

(5:1:1:1, $1/2 / 8 = 1/16$)

**A2 (3 points)** What ensemble classifier **H(x)** did you generate after three (3) rounds of Adaboost?

**H(x) =** $\text{SIGN}\left(\frac{1}{2}\ln(5) + \frac{1}{2}\ln(4) - \frac{1}{2}\ln\left(\frac{13}{3}\right)\right)$

**A3 (1 points)** Which weak classifier **h** contributes the most to H(x)? <u>Circle one</u>.

**KP(x)** **EP(x)** **CC(x)** **BP(x)** **SS(x)**

**A4 (6 points)** How does the H(x) produced after 3 rounds of Adaboost classify each of the following training points? <u>Circle one</u>.

| | | | |
|---|---|---|---|
| **O:** | Correctly Classified | Misclassified | Can't Tell |
| **M:** | Correctly Classified | Misclassified | Can't Tell |
| **T:** | Correctly Classified | Misclassified | Can't Tell |

even though we don't have the original labels, we can see if the voting power of h's that correctly classify the point outweighs the summed voting power of the h's in H(x) that misclassify the point.

## Part B: Ensembles? (16 points)

The students disagree about ensemble classifiers. **Circle** whether each student's statement below is correct, incorrect, or can't be determined from available information. **In one sentence explain your choice** in the box provided. Assume all Adaboost parameters, including weak classifiers, are the same as in Part A. Also assume as in Part A that the best classifier is the one with error rate furthest from ½.

**B1 (4 points) Mahgi:** If you performed Adaboost with no limit on the number of rounds and only used two (2) of our five (5) classifiers, then the accuracy will never exceed 80%.

**Correct** **Incorrect** **Can't Tell**

**Explanation:**

pick CC(x) with higher voting power than the 2nd $h_1(x) \rightarrow$ Accuracy will be $1 - 1/6 \geq 80\%$ (b/c CC(x) only misclassifies M and there are 6 points).

**B2 (4 points) Rachel:** Some of our five (5) classifiers will never be picked by the Adaboost algorithm.

**Correct** **Incorrect** **Can't Tell**

**Explanation:**

accepted multiple answers: ① due to tiebreaking, b/c $\varepsilon(BP(x)) = 1 - \varepsilon(KP(x))$, KP(x) comes first and BP(x) will never be selected. ② Adaboost terminates after 3 rounds! At max, pick 3/5.

*note: we don't have here that the pts misclassified by some $h_i$ is both superset and subset of others.

**B3 (4 points) Tasha:** By choosing appropriate voting powers by hand, you can make a perfect ensemble classifier out of three (3) of our five (5) weak classifiers.

**Correct** **Incorrect** **Can't Tell**

**Explanation:**

Use the same $h_i$ found by Adaboost in H(x) and pick any weights in triangle inequality. (satisfying)

**B4 (4 points)** Kyle decides to make his own ensemble classifier from three (3) of the weak classifiers:
**H(x) = 4*CC(x) + 2*EP(x) + BP(x)**

**Kyle:** My classifier H(x) is made from three (3) weak classifiers that make disjoint errors. Therefore, it has perfect accuracy on the original training data.

**Correct** **Incorrect** **Can't Tell**

**Explanation:**

Disjoint is not enough to gaurentee perfect, weights must be valid, ex: in triangle inequality.

Here, H(x) misclassifies $M^6$ b/c 2EP + BP not enough to outweigh CC(x) and CC(x) misclassifies M.

# Problem 2: Bayes Nets (44 points)

## Part A: Genie in a Bayes Net (28 points)

Christina Aguilera asks 6.034 to help her write next summer's big hit song to earn her nineteenth GRAMMY nomination. She uses her understanding of the music scene to compile a probability table with six (6) boolean variables, but she needs your help to analyze the table.

i.e. need full table.

**A1 (4 points)** Without making any assumptions about independence, how many parameters do we need in the joint probability table? (You don't need to simplify your answer.)

For full table, need one row for every possible combo of the 6 vars. B/c boolean, $\rightarrow 2^6$, by exhaustion, last row can be determined $\Rightarrow 2^6 - 1$

| *Number of parameters =* |
|---|
| $2^6 - 1 = 63$ |

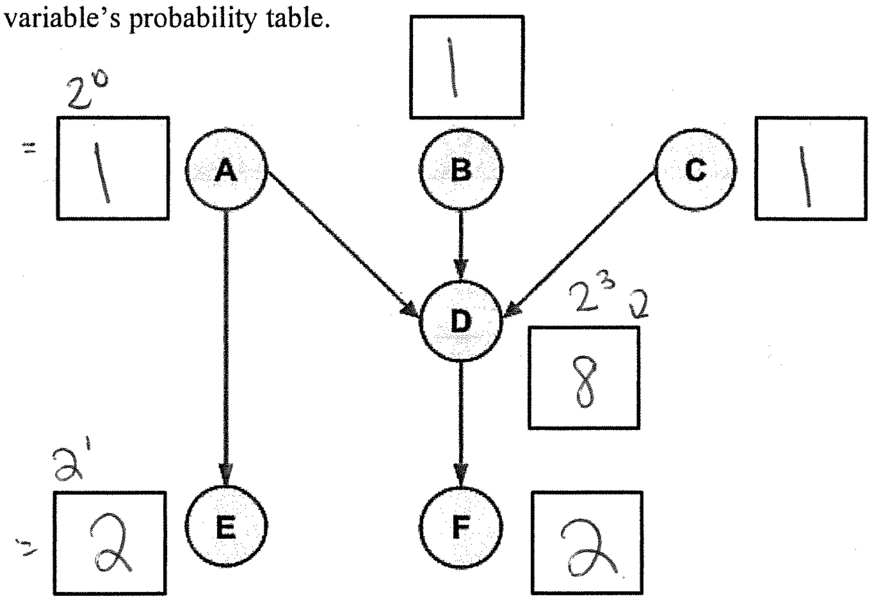
**A2 (6 points)** Christina thinks that with no independence assumptions, the joint probability table is much too big. To make her model computationally feasible with her six (6) boolean variables, she defines the Bayes net below so that she can rely on the Bayes net independence assumption.

Assume the only independence statements that are true are the ones enforced by the shape of the net. How many parameters are needed for this Bayes net?

| *Number of parameters =* |
|---|
| $1+1+1+8+2+2 = 15$ |

**For partial credit,** next to each variable in the Bayes net, clearly write the number of parameters in that particular variable's probability table.

Each var requires $2^{\#\text{parents}}$ parameters.



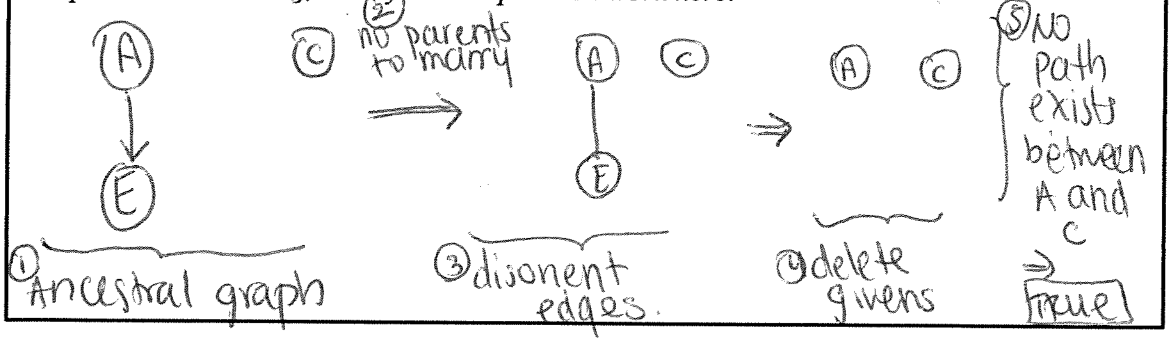
D-separation shown for each.

**For problems A3 through A5, refer to the Bayes net on the previous page. The net is also duplicated on a tear-off sheet.** Again assume the only independence statements that are true are the ones enforced by the shape of the net (all other independence statements are false).

**A3 (6 points)** P(A | C E) = P (A | E). Circle one. "A is conditionally independent of C given E"

**True** **False**

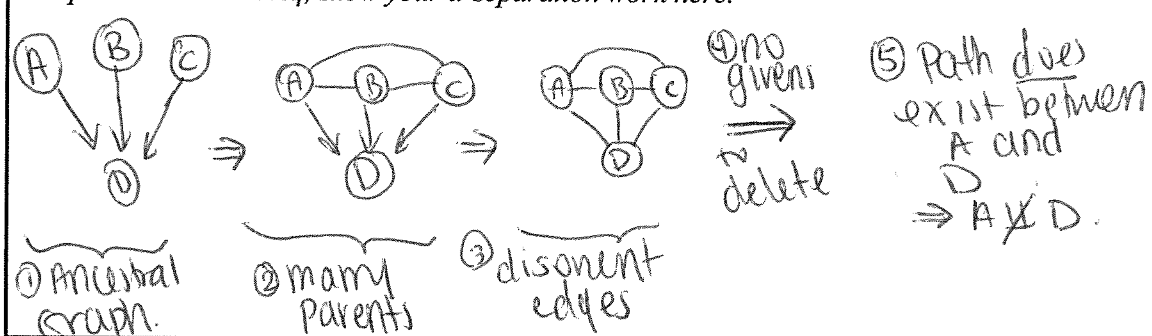
*For partial credit on A3, show your d-separation work here.*




**A4 (6 points)** A and D are marginally independent. Circle one.

**True** **False**

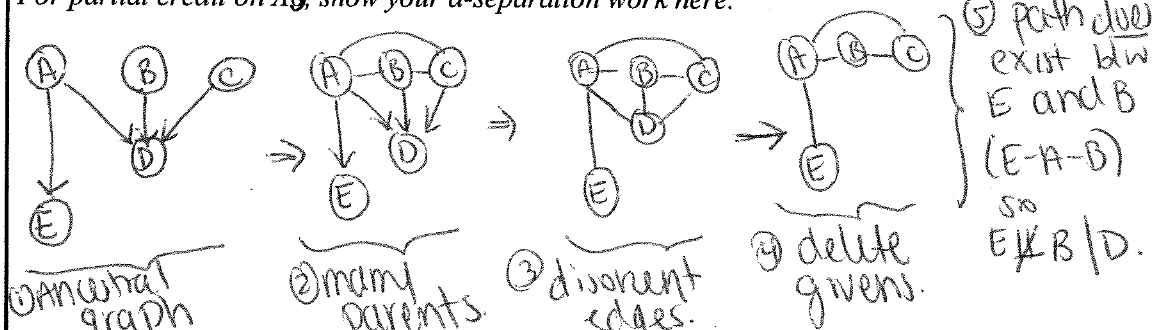
*For partial credit on A4, show your d-separation work here.*




**A5 (6 points)** E and B are conditionally independent, given D. Circle one.

**True** **False**

*For partial credit on A5, show your d-separation work here.*

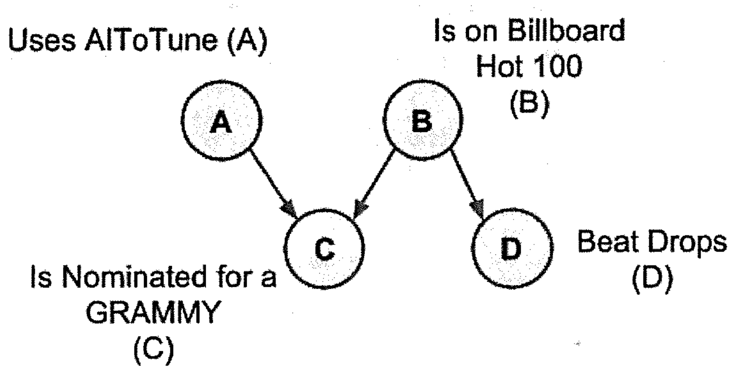


# Part B: Hit Me Bayesian One More Time (16 points)

Jealous of Christina, Britney Spears asks 6.034 to help her write next summer's even bigger hit song so she can earn her ninth GRAMMY nomination.

She believes that whether a song is nominated for GRAMMY consideration (C) depends on whether it uses AIToTune (A) and whether it is on the Billboard Hot 100 (B). She also believes that being on the Billboard Hot 100 (B) influences whether a song's rhythm changes with a beat drop (D).

**For your convenience, a copy of this Bayes net is provided on a tear-off sheet.**




Britney wants to know her chances of being considered for a GRAMMY (C = TRUE), given that A = FALSE, B = FALSE, and D = FALSE.

In other words, Britney is computing an expression for the joint probability $P(\bar{A}\,\bar{B}\,C\,\bar{D})$. You plan to use the Chain Rule to expand the joint probability as a product of conditional probabilities with increasing numbers of given variables.

**B1 (4 points)** Expand $P(\bar{A}\,\bar{B}\,C\,\bar{D})$ as a product of conditional probabilities. (*Hint: Consider expanding the expression in a particular order to best take advantage of the structure of the network, but don't apply any independence assumptions to simplify your conditional probabilities yet.*)

$$P(\bar{A}\bar{B}C\bar{D}) = P(\bar{D}\mid\bar{A}\bar{B}C)\,P(C\mid\bar{A}\bar{B})\,P(\bar{A}\mid\bar{B})\,P(\bar{B})$$

OR

$$P(C\mid\bar{A}\bar{B}\bar{D})\,P(\bar{D}\mid\bar{A}\bar{B})\,P(\bar{A}\mid\bar{B})\,P(\bar{B})$$

(we accepted any valid expansion via chain rule but the above are the ideal orderings to maximize simplification).

**B2 (6 points)** Simplify your expression in **B1** using the Bayes net independence assumption. *Your answer should be expressed in terms of conditional and marginal probabilities that can be looked up in the Bayes net.*

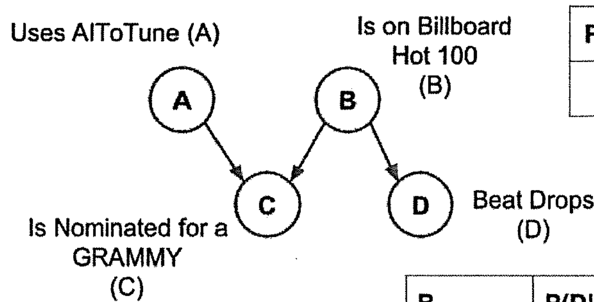
$P(\bar{A}\,\bar{B}\,C\,\bar{D}) = P(\bar{D}\mid\bar{B})\,P(C\mid\bar{A}\,\bar{B})\,P(\bar{A})\,P(\bar{B})$

via Bayes net Assumption: "a var is conditionally independent of its non descendents given its parents"

**B3 (6 points)** Brittney provides you with the parameters of the Bayes net. In the associated probability tables, each probability is represented by a lowercase letter.

**For your convenience, a copy of this Bayes net with the tables is provided on a tear-off sheet.**

| $P(\bar{A})$ | P(A) |
|---|---|
| $1-a$ | a |




| P(B) | $P(\bar{B})$ |
|---|---|
| b | $1-b$ |

| A | B | P(C\|AB) |
|---|---|---|
| T | T | c |
| T | F | d |
| F | T | e |
| F | F | f |

| B | P(D\|B) | P(D=False\|B) |
|---|---|---|
| T | g | $1-g$ |
| F | h | $1-h$ |

Evaluate your expression in **B2** and write it as an algebraic expression of the parameters a through h specified in the Bayes net.

$P(\bar{A}\,\bar{B}\,C\,\bar{D}) = P(\bar{D}\mid\bar{B})\,P(C\mid\bar{A}\,\bar{B})\,P(\bar{A})\,P(\bar{B})$

$(1-h)\cdot(f)\cdot(1-a)(1-b)$

(product)

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

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

C

1. Winston argued that inheritance is crucial in role frame representations because inheritance:
   a. makes it easier to put the representation into code and translate between coding languages.
   b. demonstrates how climbing the hierarchy contributes increasingly large jumps in knowledge.
   c. gives us names over the concepts and their relationships to one another.
   d. can be modeled as linking cause and effect, growing our semantic net.

b

2. Winston uses the term "parasitic semantics" to refer to:
   a. semantic theories riddled with conceptual holes, as if infected by parasites.
   b. supposing that a program has humanlike knowledge about labels in its vocabulary.
   c. a Marxist theory of knowledge representation developed to condemn social parasites.
   d. informal theories that gain power by mathematical mappings to Aristotelian logic.

b

3. Katz argued that a good way to collect information needed for constructing a large set of natural language questions is to:
   a. download relational databases from public resources, e.g., The Census or Google Scholar.
   b. gather semi-structured data from the information boxes on the top of Wikipedia pages.
   c. parse Wikipedia articles to understand what the articles are about.
   d. download encyclopedias such as World Book and Encyclopedia Brittanica.

a

4. Berwick argued that Aristotle's model of language, with components sound and meaning, is incomplete and requires a third component, called merge, which he referred to as a:
   a. language CPU, which processes speech into meaning and vice versa.
   b. language cache, which stores utterances to help determine context.
   c. language Turing Machine, which maintains and changes states.
   d. language encoder, which turns symbols into neurological impulses and vice versa.

d

5. Kanwisher claimed that we do not need to be able to speak or understand a language, e.g., English, in order to think because:
   a. a particular region in the brain, rather than the whole brain, is activated for language.
   b. there is no particular region in the brain for language, so language is not internalized uniquely as thought.
   c. babies are able to respond to instructions from their parents even before they can speak.
   d. even people without any functioning language region of the brain are able to perform mathematical tasks. (aphasia patients)

a

6. Winston explained that distinguishing external language from inner language matters because:
   a. external language is just a hack to get us understood; inner language is what enables our ability to be symbolic and therefore intelligent.
   b. external language, and the numerous variations of external language, is what distinguishes us from primates; inner language is just an abstraction.
   c. external language and how it changes mimics changes in human thought; inner language is a representation disconnected from how we think.
   d. external language changes too frequently to study; inner language is constant through time.

This page is intentionally blank.