6.034 Quiz 4
7 December 2016

Name: [Blank]
Email: [Blank]

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

- Jake Barnwell
- Michaela Ennis
- Rebecca Kekelishvili
- Vinny Chakradhar
- Phil Ferguson
- Nathan Landman
- Alex Charidis
- Stevie Fine
- Samarth Mohan
- Brian Copeland
- Brittney Johnson
- Jessica Noss

<table>
<thead>
<tr>
<th>Problem</th>
<th>Maximum</th>
<th>Score</th>
<th>Grader</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - Boosting</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 - Bayes</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRN</td>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There are 10 pages in this quiz, including this one.

As always, open book, open notes, open just about everything, including a calculator, but no computers or phones.
Problem 1: Football Boosting (50 points)
Part A: Modeling the NFL (33 points)
You want to predict the results of some of this week’s key football games. Many factors have an impact on each game, and no individual metric exceeds 65% accuracy in predicting games, so you decide to combine many of these weak classifiers using Adaboost.

In the table below, each training point is a game featuring “Away team @ Home team”, and each game is labeled numerically for ease of reference. Each of four weak classifiers predicts a winner for each game, either the Home team or the Away team. The actual winner of each game is listed in the last row of the table.

<table>
<thead>
<tr>
<th>Game 1</th>
<th>Game 2</th>
<th>Game 3</th>
<th>Game 4</th>
<th>Game 5</th>
<th>Game 6</th>
<th>Game 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHI @ NYG</td>
<td>DEN @ OAK</td>
<td>BUF @ SEA</td>
<td>NO @ SF</td>
<td>NYJ @ MIA</td>
<td>CAR @ LA</td>
<td>PIT @ BAL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Weak Classifier</th>
<th>Misclassified training points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fan Poll</td>
<td>1, 7</td>
</tr>
<tr>
<td>Home Team</td>
<td>4, 6</td>
</tr>
<tr>
<td>Better Defense</td>
<td>1, 2, 6</td>
</tr>
<tr>
<td>Better Quarterback</td>
<td>5, 7</td>
</tr>
</tbody>
</table>

A1 (3 points) Complete the table below by filling in which training points are misclassified by the fourth weak classifier, Better Quarterback:

A2 (30 points) On the next page, perform 2.5 rounds of boosting using these classifiers and training data. In each round, pick the classifier with the error rate furthest from ½. Break ties by choosing the classifier that comes earlier in this list: Fan Poll, Home Team, Better Defense, Better Quarterback.

In any round, if Adaboost would terminate instead of choosing a classifier, write NONE for that round’s weak classifier (h), then leave all remaining spaces blank.
<table>
<thead>
<tr>
<th>weight 1</th>
<th>Round 1</th>
<th>Round 2</th>
<th>Round 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{1}{7} \times )</td>
<td>( \frac{1}{4} \times \frac{5}{32} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weight 2</td>
<td>( \frac{1}{10} \times \frac{2}{32} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weight 3</td>
<td>( \frac{1}{10} \times \frac{2}{32} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weight 4</td>
<td>( \frac{1}{10} \times \frac{1}{4} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weight 5</td>
<td>( \frac{1}{10} \times \frac{2}{32} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weight 6</td>
<td>( \frac{1}{10} \times \frac{1}{4} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weight 7</td>
<td>( \frac{1}{4} \times \frac{5}{32} )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**error rate of Fan Poll:**
- Round 1: \( \frac{2}{7} \)
- Round 2: \( \frac{10}{20} \)

**error rate of Home Team:**
- Round 1: \( \frac{2}{7} \)
- Round 2: \( \frac{4}{20} \)

**error rate of Better Defense:**
- Round 1: \( \frac{3}{7} \)
- Round 2: \( \frac{9}{20} \)

**error rate of Better Quarterback:**
- Round 1: \( \frac{2}{7} \)
- Round 2: \( \frac{7}{20} \)

**weak classifier chosen (h):**
- Round 1: Fan Poll
- Round 2: Home Team

**weak classifier error (ε):**
- Round 1: \( \frac{2}{7} \)
- Round 2: \( \frac{1}{5} \)

**voting power (α):**
- Round 1: \( \frac{1}{2} \ln \left( \frac{1-\frac{2}{7}}{\frac{2}{7}} \right) \)
- Round 2: \( \frac{1}{2} \ln \left( \frac{\frac{5}{10}}{\frac{5}{20}} \right) \)

---

**Show your work for partial credit:**

\[
\omega_{new} = \begin{cases} \frac{1}{2} \omega \text{ if incorrectly classified} \\ \frac{1}{2} \frac{1}{1-\varepsilon} \omega \text{ if correctly classified} \end{cases}
\]

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

**Round 1**
- \( \alpha = \frac{1}{2} \ln \left( \frac{\frac{5}{10}}{\frac{2}{7}} \right) \)
- \( \omega_{new} = \frac{1}{2} \cdot \frac{2}{7} \cdot \frac{1}{4} = \frac{1}{10} \)
- \( \omega_{new} = \frac{1}{2} \cdot \frac{2}{7} \cdot \frac{1}{5} = \frac{1}{10} \)

**Round 2**
- \( \alpha = \frac{1}{2} \ln \left( \frac{\frac{5}{10}}{\frac{5}{20}} \right) \)
- \( \omega_{new} = \frac{1}{2} \cdot \frac{5}{10} \cdot \frac{5}{10} = \frac{5}{80} = \frac{2}{32} \) or \( \frac{5}{8} \cdot \frac{1}{10} = \frac{5}{80} = \frac{2}{32} \)
Part B: Predicting the NFL (17 points)

Bruce Headstrong constructs this ensemble classifier:

\[ H(x) = 5 \cdot (\text{Fan Poll 's vote}) + 4 \cdot (\text{Home Team 's vote}) - 3 \cdot (\text{Better Defense 's vote}) \]

B1 (5 points) Which training points does Bruce's ensemble classifier \( H(x) \) misclassify? Circle all that apply, or circle NONE if no points are misclassified:

\[
\begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & \text{NONE} \\
\end{array}
\]

B2 (3 points) Is there some assignment of voting powers \( \alpha \) that could form a perfect ensemble classifier out of these three weak classifiers? If so, assign voting powers to the three weak classifiers to form a perfect ensemble classifier. If not, instead circle CAN'T BE DONE.

<table>
<thead>
<tr>
<th>Weak classifier</th>
<th>Voting power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fan Poll</td>
<td>( \alpha = )</td>
</tr>
<tr>
<td>Home Team</td>
<td>( \alpha = )</td>
</tr>
<tr>
<td>Better Defense</td>
<td>( \alpha = )</td>
</tr>
</tbody>
</table>

B3 (6 points) Now, you want to use Bruce's ensemble classifier to predict three games. Based on the data below, how does Bruce’s classifier classify each game? (That is, which team is predicted to win?) Fill in the Predicted Winner row by writing Home or Away in each box. If the answer can’t be determined from the available information, instead write CAN'T TELL.

<table>
<thead>
<tr>
<th>Test Points: Three upcoming games from NFL Week 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Game 8</td>
</tr>
<tr>
<td>CIN @ NYG</td>
</tr>
<tr>
<td>Fan Poll</td>
</tr>
<tr>
<td>Home Team</td>
</tr>
<tr>
<td>Better Defense</td>
</tr>
<tr>
<td>Better Quarterback</td>
</tr>
<tr>
<td>Predicted Winner</td>
</tr>
</tbody>
</table>

B4 (3 points) The next week, you find out that the actual winners of the Week 10 games were:

<table>
<thead>
<tr>
<th>Game</th>
<th>Game 8</th>
<th>Game 8</th>
<th>Game 10</th>
</tr>
</thead>
</table>

Given this information, what was the accuracy of Bruce's classifier (i.e. fraction of games correctly predicted) on this test dataset? Fill in the box: \[ \frac{1}{3} \]
Problem 2: Fantastic Bayes (50 points)

A British wizard has accidentally released some magical creatures at MIT! He offers a reward for safely capturing and returning the creatures. Eager to help, you set off in search of clues!

Part A: Big-Picture Bayes (7 points)
On the chalkboards in Stata, you discover a drawing of an enormous, densely packed Bayes net with variables arranged in a square grid. Upon closer inspection, you find that each variable within the grid has exactly three parents and three children, as shown below, zoomed in:

On the next board, you find eight (8) sketches. In each sketch, the shaded region represents some subset of nodes in the enormous Bayes net:

Which one sketch most closely represents...

...the ancestors of the center node?  
\[ \text{A} \]

...the descendants of the center node?  
\[ \text{B} \]

...the nodes that are conditionally independent of the center node, assuming that the center node’s parents are given?  
\[ \text{F} \]
Part B: Classifying Fantastic Beasts (25 points)

While walking through Lobby 10, you notice an unfamiliar 4-legged creature running through Killian Court. You know that three 4-legged creatures are still on the loose: a unicorn, a hippogriff, and a chimaera. In your guidebook on Fantastic Beasts, you find the following information on a torn page:

The Bayes net below can be used to distinguish between three 4-legged creatures (classification \( Y \)) using naïve Bayes classification: a unicorn, a hippogriff, and a chimaera. Some of these creatures Look friendly, and each creature may Chase people, in accordance with the probabilities given below:

<table>
<thead>
<tr>
<th>( Y )</th>
<th>( P(Y=\text{unicorn}) )</th>
<th>( P(Y=\text{hippogriff}) )</th>
<th>( P(Y=\text{chimaera}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>unicorn</td>
<td>0.5</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>hippogriff</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chimaera</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Classification

\( Y \)

\( L \) : Looks friendly

\( C \) : Chases people

**B1 (3 points)** Based on the possible values of each variable and the structure of the Bayes net, what is the minimum number of parameters required in the Bayes net?

Show your work for partial credit:

\[
Y: 2 \rightarrow \text{need } 2 \text{ of } P(Y=u), P(Y=h), P(Y=c); \text{ use exhaustion to get the third}
\]

\[
P(L|Y): 3 \rightarrow \text{need } P(L|Y) \text{ for all } 3 \text{ values of } Y
\]

\[
P(C|Y): 3 \rightarrow \text{need } P(C|Y) \text{ for all } 3 \text{ values of } Y
\]

**B2 (2 points)** You want to classify the unknown creature as a unicorn, hippogriff, or chimaera. Based on the prior probabilities alone, circle the one most likely classification. If it’s impossible to tell, instead circle CAN’T TELL:

unicorn     hippogriff     chimaera     CAN’T TELL
**B3 (18 points)** Rafael Reif comes running into the lobby, looking terrified. You ask him for more information about the creature, such as whether it looked like a horse with a single horn, but he doesn’t remember. He replies: “It was chasing me too fast for me to notice details, but it didn’t look friendly!”

Given that the creature chases people (C=True) and does NOT look friendly (L=False), use the Bayes net to determine the likelihood of each classification. **Below, list the classifications (unicorn, hippogriff, chimaera) from most to least likely:**

<table>
<thead>
<tr>
<th></th>
<th>Most likely</th>
<th></th>
<th>Least likely</th>
</tr>
</thead>
<tbody>
<tr>
<td>hippogriff</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chimaera</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unicorn</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Show your work for partial credit:**

\[
P(Y|C \& \bar{L}) = \frac{P(C \& \bar{L}|Y)P(Y)}{P(C \& \bar{L})} = \alpha \frac{P(C|Y)P(\bar{L}|Y)P(Y)}{P(C)}
\]

\[
\alpha \cdot P(C|Y)P(\bar{L}|Y)P(Y)
\]

- **Y = unicorn:** \((.4)(1-.8)(.5) = 0.04\)
- **Y = hippogriff:** \((.4)(1-.6)(.4) = 0.064\)
- **Y = chimaera:** \((.6)(1-.1)(.1) = 0.054\)

**B4 (2 points)** How would your naïve Bayes classifier classify the creature, given the evidence? (Circle one)

- unicorn
- hippogriff
- chimaera
- CAN’T TELL
Part C: Chalkboard Malfunction (18 points)

You enter 10-250 and notice that the Boards appear to be sliding up and down by themselves! You realize that this could either be caused by technical Malfunction, or by an invisible Demiguise (a magical creature) pushing the buttons. Armed with your guidebook on Fantastic Beasts, you draw the following Bayes net:

You wonder whether you should be surprised that the boards are sliding without any people present. Without knowing anything about whether there is a Demiguise and/or a Malfunction, you calculate the probability of the Boards would sliding and get $P(B) = 5/100$.

C1 (10 points) Given that the Boards are sliding ($B=True$), what is the probability that there is an invisible Demiguise near the boards ($D=True$) and that there is NOT a Malfunction ($M=False$)? Write your answer as a number or numeric expression:

$$P(D\bar{M}|B) = \frac{24}{50}$$

Show your work for partial credit:

$$P(D\bar{M}|B) = \frac{P(B|D\bar{M})\cdot P(D\bar{M})}{P(B)} = \frac{P(B|D\bar{M})\cdot P(D)\cdot P(\bar{M})}{P(B)}$$

$$= \frac{3/10 \cdot 1/10 \cdot (1 - 2/10)}{5/100}$$

$$= \frac{24}{1000}$$

$$= \frac{24}{50}$$
C2 (3 points) It's difficult to find an invisible creature, so you wonder whether you should just call MIT FIXIT to report a malfunction. Accordingly, you start computing the probability that there is a Malfunction given that the Boards are sliding, that is, \( P(M|B) \).

But suddenly, a Demiguise becomes visible, and you see that it is pushing the buttons! With this new information, the probability of a malfunction is \( P(M|BD) \). Given the Demiguise's appearance, the probability of a malfunction has: (Circle one)

- INCREASED
- DECREASED
- STAYED THE SAME

C3 (2 points) Which effect or principle best explains your answer to part C2 above? Circle the one best answer.

1. Conditional Independence
2. Occam's Razor
3. Exhaustion
4. Explaining Away
5. Mutual Exclusion

C4 (3 points) You carefully capture the Demiguise and move away from the buttons. Surprisingly, the boards are still sliding! Now, it should be easy to determine the probability of a Malfunction, given that the Boards are sliding and there is no Demiguise pushing the buttons: \( P(M=True|B=True, D=False) \). This probability is closest to: (Circle one)

- 0
- 0.1
- 0.2
- 0.3
- 0.4
- 0.5
- 0.6
- 0.7
- 0.8
- 0.9
- 1.0

1. The boards are sliding
2. The only possible causes of the board sliding are malfunction or Demiguise because \( P(M|MB)=0 \)
3. There is no Demiguise
4. Therefore, there must be a malfunction

See reverse for SRN questions →
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. Mansinghka explained how probabilistic programming has been used to:
   1. Improve Wall Street statistical arbitrage schemes.
   2. Build cyber defenses against ransomware.
   3. Find characteristics associated with diabetes.
   4. Examine credit card histories to predict likely bankruptcies.
   5. Speed up deep neural net learning.

2. AlphaGo:
   1. Models the way humans play Go.
   2. Picks the move with the highest probability of success using Bayesian inference.
   3. Evaluates all move combinations 20 levels deep using a computing cloud.
   4. Uses Monte Carlo game play in its static evaluator.
   5. Determines each move using k-nearest neighbors in a vast library of human games.

3. Berwick argued that:
   1. All primates have the merge operation.
   2. Merge capability increases with brain size.
   3. Merge is associated with completion of an anatomical loop in the brain.
   4. Recently discovered Neanderthal artifacts demonstrate that they too had the merge capability.
   5. Nim Chimpsky was taught to use merge via training in American Sign Language.

4. Winston described how the Genesis story-understanding system:
   1. Tells stories persuasively by reference to parables found in sacred texts.
   2. Reflects our human tendency to seek explanations.
   3. Deploys neural-net techniques trained on millions of stories to make inferences.
   4. Summarizes stories by retaining only the first $n$ and final $m$ events.
   5. Uses word co-occurrence statistics to identify concepts such as revenge.

5. Pratt emphasized that:
   1. Statistical evidence indicates self-driving cars will cut fatalities per mile to near zero in 10-20 years.
   2. Establishing responsibility for self-driving car accidents will require legal innovation on an unprecedented scale.
   3. The “handoff” problem is about researcher reluctance to transfer development work to production engineers.
   4. Texting and other new distractions are a major reason for increased self-driving car research.
   5. Human drivers are extremely competent when viewed from a fatalities-per-mile perspective.

6. According to Wilson, place fields are:
   1. Neurons in rats’ brains that fire based on locations they’re thinking about.
   2. Nearby regions in a maze in which rats can localize other rats.
   3. Electrical fields that help rats determine their location.
   4. Heuristic estimates that help rats find their food based on visual stimuli.
   5. Mental maps that rats form to store the shortest path through each maze.