6.034 Quiz 4
December 2, 2009

Name	
email	

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

TAs
Erica Cooper
Matthew Peairs
Lisa Fisher
Mark Seifter
Yuan K. Shen
Jeremy Smith
Olga Wichrowska

Thu	
Time	Instructor
12-1	Gregory Marton
1-2	Berwick/Marton
2-3	Berwick/Marton
3-4	Berwick/Marton

Fri	
Time	Instructor
1-2	Randall Davis
2-3	Randall Davis
3-4	Randall Davis

Problem number	Maximum	Score	Grader
1	49		
2	50		
Total	100		

There are 10 pages in this quiz, including this one. In addition, tear-off sheets are provided at the end with duplicate drawings and data. As always, open book, open notes, open just about everything.

Problem 1: Bayesian Inference (49 points)

Part A: Bayes Nets (30 Points)

0.1

0.2

M	H
F	0.7
T	0.8

V	R
F	0.1
T	0.3

M	V	Z
F	F	0.01
F	T	0.4
T	F	0.5
T	T	0.6
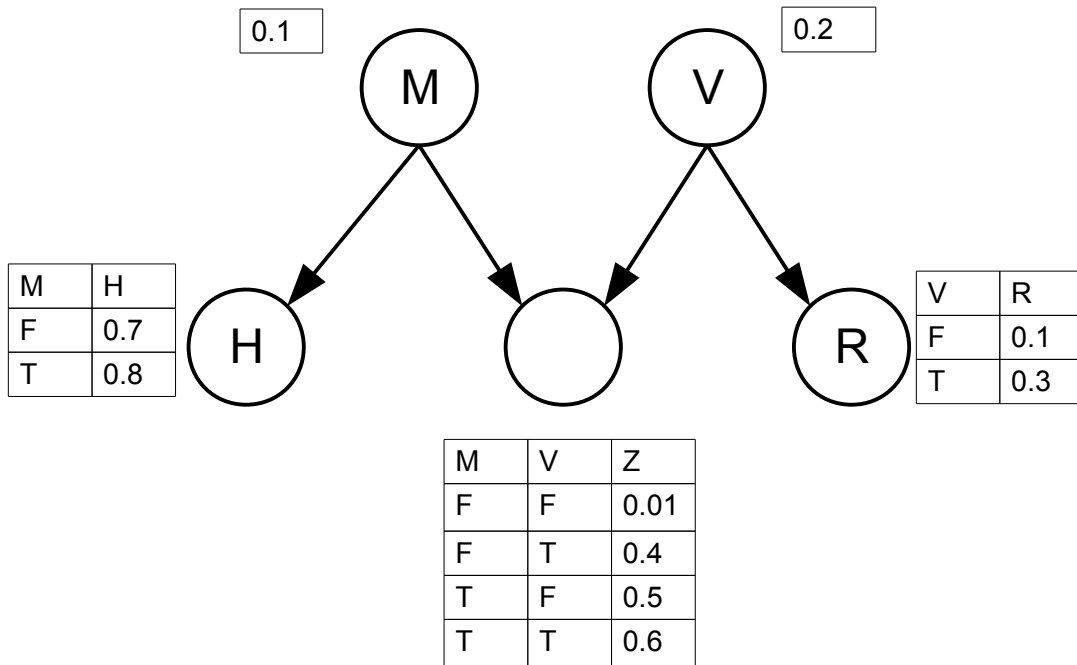
You decide to use Bayes Nets to help you explain recent sightings of Zombies on campus.   In your net you define the following variables:

M = A **M**agical spell was cast { T, F }
V = There is a **V**iral outbreak of H1Z1   {T, F}
H = You see a **H**ippogriff (a magical creature)   { T, F }
Z = You see a **Z**ombie { T, F}
R = You see Time t**R**avelers in hazmat suits trying to contain the viral outbreak. {T, F }
Using this Bayes Net, answer the following questions:

A1 (3 points): What is the probability of NOT seeing a Zombie given that NO magic spell has been cast and there is NO viral outbreak? **Write down the probability you are computing and give an numeric answer.**

A2 (3 points): Joint Probability: What is the probability of the following state of the world?
P(M=T, H=T, Z=T, V=T, R=T)   **Give an expression in terms of the probabilities that can be directly read from the network. You need not compute the final numeric answer.**

A3 (4 points): What is the probability of seeing a Zombie given that there is a viral outbreak?
P(Z=T | V=T)

Give an expression in terms of probabilities read from the network. You need not compute a final numeric answer.

A4 (6 points): What is the probability of seeing a Zombie? P(Z=T)   **Give an expression in terms of probabilities read from the network. You need not compute a final numeric answer.**

A5 (8 points): What is the probability of a Viral outbreak if you see a Zombie? $P(V=T \mid Z=T)$: **Give an expression in terms of probabilities read from the network and/or probability computed from a previous questions. You need not compute a final numeric answer.**

A6 (6 points): Write inequalities to give an ordering to these probabilities from smallest to largest $P(V=T \mid Z=T)$, $P(V=T \mid Z=T, M=F)$, $P(V=T \mid Z=T, M=T)$.

Part B: Naïve Bayes (19 Points)

You decide to create a Naïve Bayes classifier to distinguish Zombies from MIT students.
You survey Zombies and Healthy MIT students based on the following true and false questions:

      W – Do you wear tattered clothes?
      S – Do you get very little sleep?
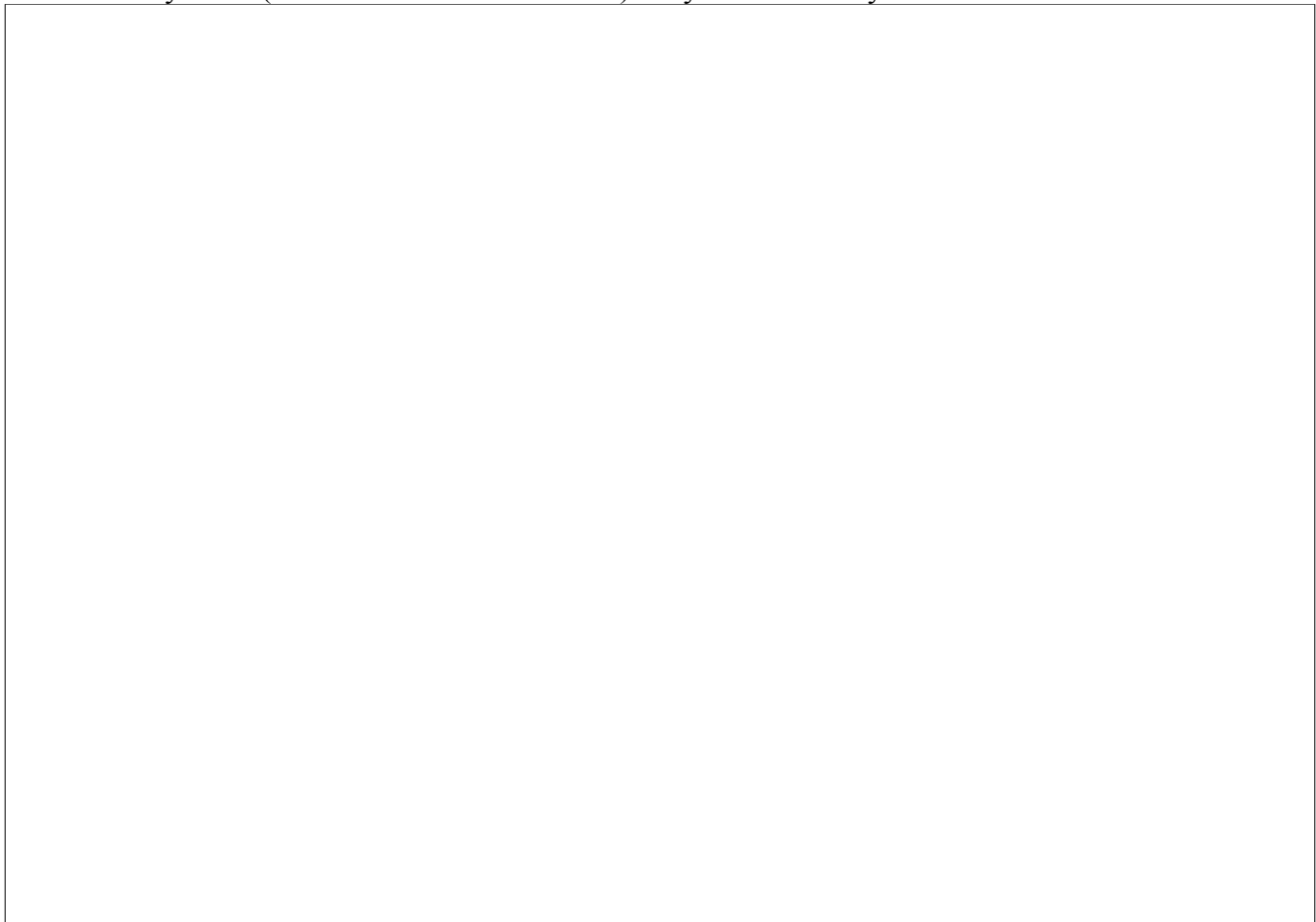   N – Do you often venture out at night?
      B – Do you enjoy a healthy diet of human brains?

Your survey returned the following data set:

	W = True	S = True	N = True	B = True	# surveyed
Zombie	6	8	9	9	10
Non-Zombie	10	15	12	1	20

B1 (9 points):
Draw the Bayes net (with all CPT tables filled in) for your Naïve Bayes Classifier.

B2 (10 points): You decide to test your classifier on Eric.   He has the following characteristic:

He is a night owl (likes to go out late into the night).   He wears old and tattered clothes, and because of his classes he gets very little sleep. He admits liking the occasional Kidney Pie, but he would never ever eat human brains.

What is Eric according to your classifier? (Circle one)

Healthy   Zombie   Can't be determined

To get full credit, you must show your calculations:

Question 2: Boosting (50 Points)

After graduating MIT, you get a job working for Van Helsing and Summers, a famous vampire hunter consulting agency. Gabriel Van Helsing, one of the two founders, once attended several 6.034 lectures as a guest, and he remembers Professor Winston's Vampire Identification Tree lecture. He assigns you the task of creating a superior classifier for vampires by using boosting on the following data.

Vampires:

ID	Name	Vampire	Evil	Emo	Transforms	Sparkly	# Romantic Interests	ID
1	Dracula	Y	Y	N	Y	N	5	1
2	Angel	Y	N	Y	Y	N	5	2
3	Edward Cullen	Y	N	Y	N	Y	1	3
4	Saya Otonashi	Y	N	Y	N	N	3	4
5	Lestat de Lioncourt	Y	N	Y	N	N	5	5
6	Bianca St. Claire	Y	Y	N	Y	N	5	6
7	Mircalla Karnstein	Y	Y	N	Y	N	5	7
8	Sailor Moon	N	N	N	Y	Y	1	8
9	Squall Leonhart	N	N	Y	N	N	1	9
10	Circe	N	N	N	N	N	5	10

Part A (10 points)

To make it easier for yourself, you first determine for each possible classifier (including the classifiers True and False which mean that everyone is a Vampire or not a Vampire, respectively), which of the data points would be misclassified. For example, for the first classifier, the one that says a person is a vampire if he is Evil (and not a vampire if not evil) the classifier is wrong on sample 2, 3, 4, and 5. We have given each classifier a name to make it easier for you to refer to them.

Classifier	Test	Value	Misclassified
A	Evil	Y	2, 3, 4, 5
B	Emo	Y	
C	Transforms	Y	
D	Sparkly	Y	
E	# of Romantic interests	> 2	
F	# of Romantic interests	> 4	
G	TRUE		
H	Evil	N	
I	Emo	N	
J	Transforms	N	
K	Sparkly	N	
L	# of Romantic interests	< 2	
M	# of Romantic interests	< 4	
N	FALSE		

Part B (5 points)

Again to help yourself later, only 6 of these classifiers would ever be used because the other 8 make all the same errors as one of the other classifiers and then make additional errors. **Circle the 6 above** that you would consider using. **If you are sure you have the right 6, you don't have to waste your time with any of the others in Part C.**

Part C (15 points)

Now you are ready to perform Boosting on the collected Vampire dataset. Fill in the following blanks for the weights, classifiers, errors, and alphas of the first three rounds of Boosting.   If there is ever a tie, break it by choosing the classifier that is higher on the list in Part A.   Remember to only use the classifiers you circled in Part B.

	Round 1	Round 2	Round 3
w1			
w2			
w3			
w4			
w5			
w6			
w7			
w8			
w9			
w10			
h			
ε			
α			

Part D (10 points)

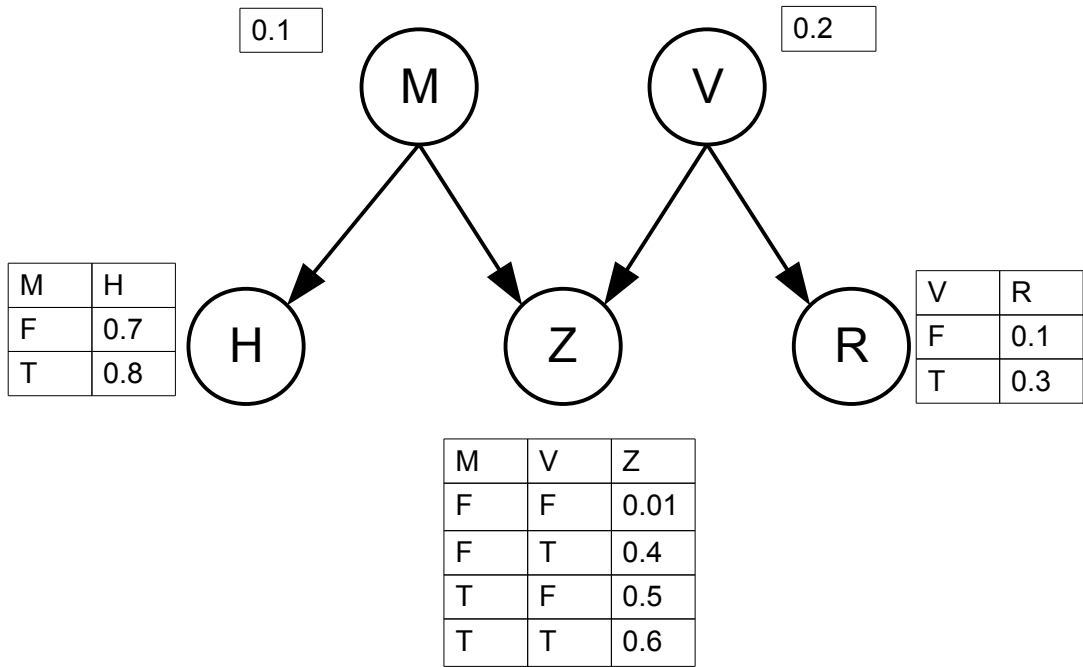What is the final classifier you find when performing three rounds of Boosting:

How many of the data points does your final classifier classify correctly?

Part E (10 points)

Wesley Windham-Pryce, a fellow consultant, has noticed a few correlations between some of the classifiers you used, and so he suggests using a new set of weak classifiers consisting of each pair of your original weak classifiers logically ANDed or ORed together (so for instance, two of the new classifiers would be "Emo=Y OR Evil=Y" and "Sparkly=N AND Transforms=Y").   He believes that you will be able to classify large vampire datasets more quickly and with fewer rounds of boosting using his system.   Do you agree or disagree with Wesley?   Explain your argument briefly and precisely.

Tear off sheet—you need not hand this in.

ID	Name	Vampire	Evil	Emo	Transforms	Sparkly	# Romantic Interests	ID
1	Dracula	**Y**	Y	N	Y	N	5	1
2	Angel	**Y**	N	Y	Y	N	5	2
3	Edward Cullen	**Y**	N	Y	N	Y	1	3
4	Saya Otonashi	**Y**	N	Y	N	N	3	4
5	Lestat de Lioncourt	**Y**	N	Y	N	N	5	5
6	Bianca St. Claire	**Y**	Y	N	Y	N	5	6
7	Mircalla Karnstein	**Y**	Y	N	Y	N	5	7
8	Sailor Moon	**N**	N	N	Y	Y	1	8
9	Squall Leonhart	**N**	N	Y	N	N	1	9
10	Circe	**N**	N	N	N	N	5	10

| 0.1 |

M   V | 0.2 |

M	H
F	0.7
T	0.8

H   Z   R

V	R
F	0.1
T	0.3

M	V	Z
F	F	0.01
F	T	0.4
T	F	0.5
T	T	0.6

6.034 Artificial Intelligence

Fall 2010