Exam 2 Solutions: CS 6364
Fall 2015

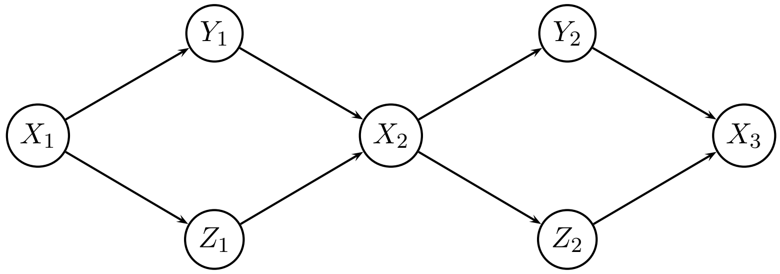
The exam is open book (Russell and Norvig, third edition) but not open notes/tablet/laptop. Answer the questions in the space provided on the question sheets. If you run out of room for an answer, use an additional sheet (available from the instructor) and staple it to your exam.

- **NAME** ________________________________
- **UTD-ID if known** ____________________________

Question	Points	Score
Bayesian Networks: Representation	10	
Bayesian Networks: Inference	10	
Particle Filtering and Bayes net Sampling	10	
Naive Bayes	10	
Perceptron	10	
Total:	50	

Question 1: Bayesian Networks: Representation (10 points)

Consider the Bayesian network given below:



(a) (5 points) Is X_3 conditionally independent of X_2 given Y_2? Explain your answer. No credit without a correct explanation.

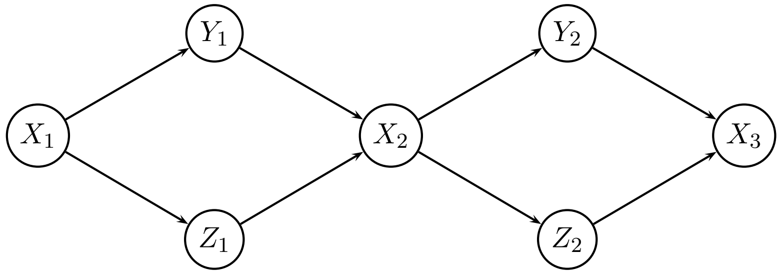
> **Solution:** ***No.*** Based on the test we discussed in class, the path $X_2 \rightarrow Z_2 \rightarrow X_3$ is active since Z_2 is not observed. Students can also use the advanced version of the test described in class.

(b) (5 points) Is X_1 conditionally independent of X_3 given X_2? Explain your answer. No credit without a correct explanation.

> **Solution:** ***Yes.*** All paths from X_1 to X_3 go through X_2 and have a sub-path of the form "$\rightarrow X_2 \rightarrow$" which is not active since X_2 is observed. As a result there are no active paths between X_1 and X_3 given X_2. Students can also use the advanced version of the test described in class.

Question 2: Bayesian Networks: Inference (10 points)

Consider the Bayesian network given below:



Assuming that all variables take values from the domain $\{0, 1\}$, the conditional probability tables are given below:

X_1	$P(X_1)$
0	0.8

Y_i	X_i	$P(Y_i \mid X_i)$
0	0	0.3
0	1	0.6

Z_i	X_i	$P(Z_i \mid X_i)$
0	0	0.2
0	1	0.5

X_{i+1}	Y_i	Z_i	$P(X_{i+1} \mid Y_i, Z_i)$
0	0	0	0.4
0	0	1	0.1
0	1	0	0.8
0	1	1	0.2

where $i \in \{1, 2\}$, which means that the CPT of $P(Y_2|X_2)$ is qualitatively equivalent to $P(Y_1|X_1)$, $P(Z_2|X_2)$ is qualitatively equivalent to $P(Z_1|X_1)$ and $P(X_3|Y_2, Z_2)$ is qualitatively equivalent to $P(X_2|Y_1, Z_1)$.

(a) (5 points) Compute $P(X_1 = 1|X_2 = 0)$ using the variable elimination algorithm. Show your work.

Solution: It turns out that the sub-network over X_3, Y_2 and Z_2 is not relevant since eliminating them in that order will yield a 1.

Thus, we have three factors remaining: $\phi_1(X_1) = P(X_1)$; $\phi_2(X_1, Y_1) = P(Y_1|X_1)$; $\phi_3(X_1, Z_1) = P(Z_1|X_1)$; and $\phi_4(Y_1, Z_1) = P(X_2 = 0|Y_1, Z_1)$

Eliminating $Y_1 : \sum_{Y_1} \phi_2(X_1, Y_1)\phi_4(Y_1, Z_1) = \phi_5(X_1, Z_1)$

$\sum_{Y_1}$

X_1	Y_1	$\phi_2(X_1, Y_1)$
0	0	0.3
0	1	0.7
1	0	0.6
1	1	0.4

$\times$

Y_1	Z_1	$\phi_4(Y_1, Z_1)$
0	0	0.4
0	1	0.1
1	0	0.8
1	1	0.2

$=$

X_1	Z_1	$\phi_5(X_1, Z_1)$
0	0	0.68
0	1	0.17
1	0	0.56
1	1	0.14

Eliminating $Z_1 : \sum_{Z_1} \phi_5(X_1, Z_1)\phi_3(X_1, Z_1) = \phi_6(X_1)$

$\sum_{Z_1}$

X_1	Z_1	$\phi_5(X_1, Y_1)$
0	0	0.68
0	1	0.17
1	0	0.56
1	1	0.14

$\times$

X_1	Z_1	$\phi_3(X_1, Z_1)$
0	0	0.2
0	1	0.8
1	0	0.5
1	1	0.5

$=$

X_1	$\phi_6(X_1)$
0	$0.68 * 0.2 + 0.17 * 0.8 = 0.272$
1	$0.56 * 0.5 + 0.14 * 0.5 = 0.35$

Calculations at X_1

X_1	$\phi_6(X_1)$
0	0.272
1	0.35

$\times$

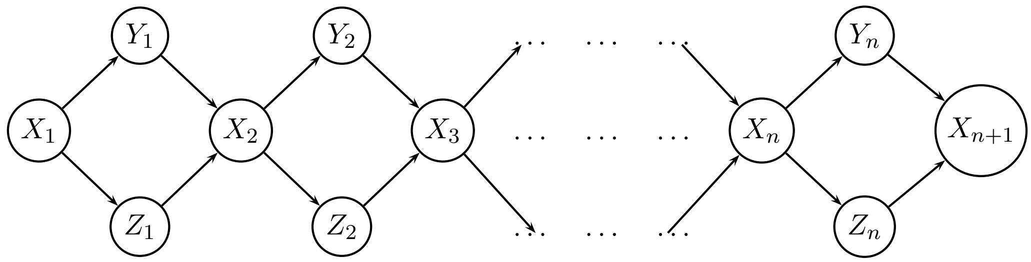
X_1	$\phi_1(X_1)$
0	0.8
1	0.2

$=$

X_1	$\phi_7(X_1)$
0	0.2176
1	0.07

Normalizing ϕ_7 yields $P(X_1 = 1|X_2 = 0) = 0.243$.

(b) (3 points) Consider the following generalization of the Bayesian network. For this question, assume that each variable can take d values.



What is the **optimal elimination order** for computing $P(X_1|X_{n+1} = x_{n+1})$ (here X_{n+1} is an evidence variable). Recall that an elimination order o is optimal if there does not exist an order o' such that running variable elimination along o' has smaller time and space complexity than running it along o.

Solution: The optimal elimination ordering is $(Y_n, Z_n, X_n, Y_{n-1}, Z_{n-1}, \ldots, X_2, Y_1, Z_1)$.

(c) (2 points) What is the **time and space complexity** of variable elimination along the order you provided in your answer to the previous question. Assume that there are d values in the domain of each variable.

Solution: Time and Space complexity is $O(nd^3)$ since each new factor generated using variable elimination will be defined over at most 3 variables using the optimal ordering.

Question 3: Particle Filtering and Bayes net Sampling (10 points)

Consider the Bayesian network given below. The domains of the variables X, Y, Z and W are $\{+x, -x\}$, $\{+y, -y\}$, $\{+z, -z\}$ and $\{+w, -w\}$ respectively.

$$X \longrightarrow Y \longrightarrow Z \longrightarrow W$$

Let the CPTs be defined as follows:

X	$P(X)$
$-x$	$3/4$

X	Y	$P(Y\|X)$
$-x$	$-y$	$2/3$
$+x$	$-y$	$4/5$

Y	Z	$P(Z\|Y)$
$-y$	$-z$	$1/4$
$+y$	$-z$	$1/2$

W	Z	$P(W\|Z)$
$-z$	$-w$	$1/8$
$+z$	$-w$	$5/6$

You are given the following 8 samples:

$+x$	$+y$	$-z$	$-w$
$+x$	$-y$	$+z$	$-w$
$-x$	$+y$	$+z$	$-w$
$-x$	$-y$	$+z$	$-w$

$+x$	$-y$	$-z$	$+w$
$+x$	$+y$	$+z$	$-w$
$-x$	$+y$	$-z$	$+w$
$-x$	$-y$	$+z$	$-w$

(a) (2 points) Assume that these samples came from performing **Prior Sampling**, and calculate the sample estimate of $P(+w)$.

Solution: Estimate is $2/8$ (two $+w$ samples out of the eight samples)

(b) (2 points) Now we will estimate $P(+z|+x, -w)$. In the samples given above, clearly cross out the samples that would not be used when doing **Rejection Sampling** for this task, and write down the sample estimate of $P(+z|+x, -w)$ below.

Solution: Cross out samples which do not have $(+x, -w)$, which is samples 3, 4, 5 , 7 and 8. Out of these three samples, $+z$ appears 2 times. Thus, the estimate is $2/3$.

(c) (2 points) Now, we will use likelihood weighting to estimate $P(-x|+y,-w)$. Suppose that the following samples are generated. Fill in the weight of each sample in the corresponding row.

1. Sample: $-x,+y,+z,-w$ Weight = $P(+y|-x)P(-w|+z) = 1/3 * 5/6 = 5/18$

2. Sample: $+x,+y,+z,-w$ Weight = $P(+y|+x)P(-w|+z) = 1/5 * 5/6 = 1/6$

3. Sample: $+x,+y,-z,-w$ Weight = $P(+y|+x)P(-w|-z) = 1/5 * 1/8 = 1/40$

4. Sample: $-x,+y,-z,-w$ Weight = $P(+y|-x)P(-w|-z) = 1/3 * 1/8 = 1/24$

(d) (2 points) From the weighted samples in the previous question, estimate $P(-x|+y,-w)$.

Solution:

$$P(-x|+y,-w) = \frac{5/18 + 1/24}{5/18 + 1/6 + 1/40 + 1/24} = 0.625$$

(e) (2 points) Which query is better suited for likelihood weighting, $P(W|X)$ or $P(X|W)$? Justify your answer in one sentence.

Solution: $P(W|X)$ is better than $P(X|W)$ because in the former we are sampling from the posterior distribution (namely $P(Y,Z,W|X)$) while in the latter the sampling distribution has no knowledge of evidence.

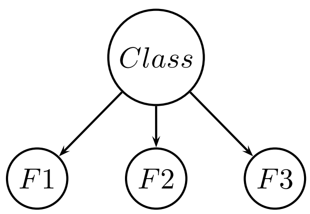
Question 4: Naive Bayes (10 points)

Consider the training data given below. Each feature $F1, F2, F3$ can take three values: $\{a, b, c\}$. The class variable (that you have to predict) can take two values $\{+, -\}$.

F1	F2	F3	Class
a	c	a	+
c	a	c	+
a	a	c	−
b	c	a	−
c	c	b	−

(a) (7 points) Build a Naive Bayes classifier for the above training data. **Use 1-laplace smoothing**. Precisely, show the structure of the network and the CPTs.

Solution:



$P(Class = +) = 2/5$
$P(F1|Class = +) = (2/5, 1/5, 2/5)$
$P(F1|Class = -) = (1/3, 1/3, 1/3)$
$P(F2|Class = +) = (2/5, 1/5, 2/5)$
$P(F2|Class = -) = (1/3, 1/6, 1/2)$
$P(F3|Class = +) = (2/5, 1/5, 2/5)$
$P(F3|Class = -) = (1/3, 1/3, 1/3)$

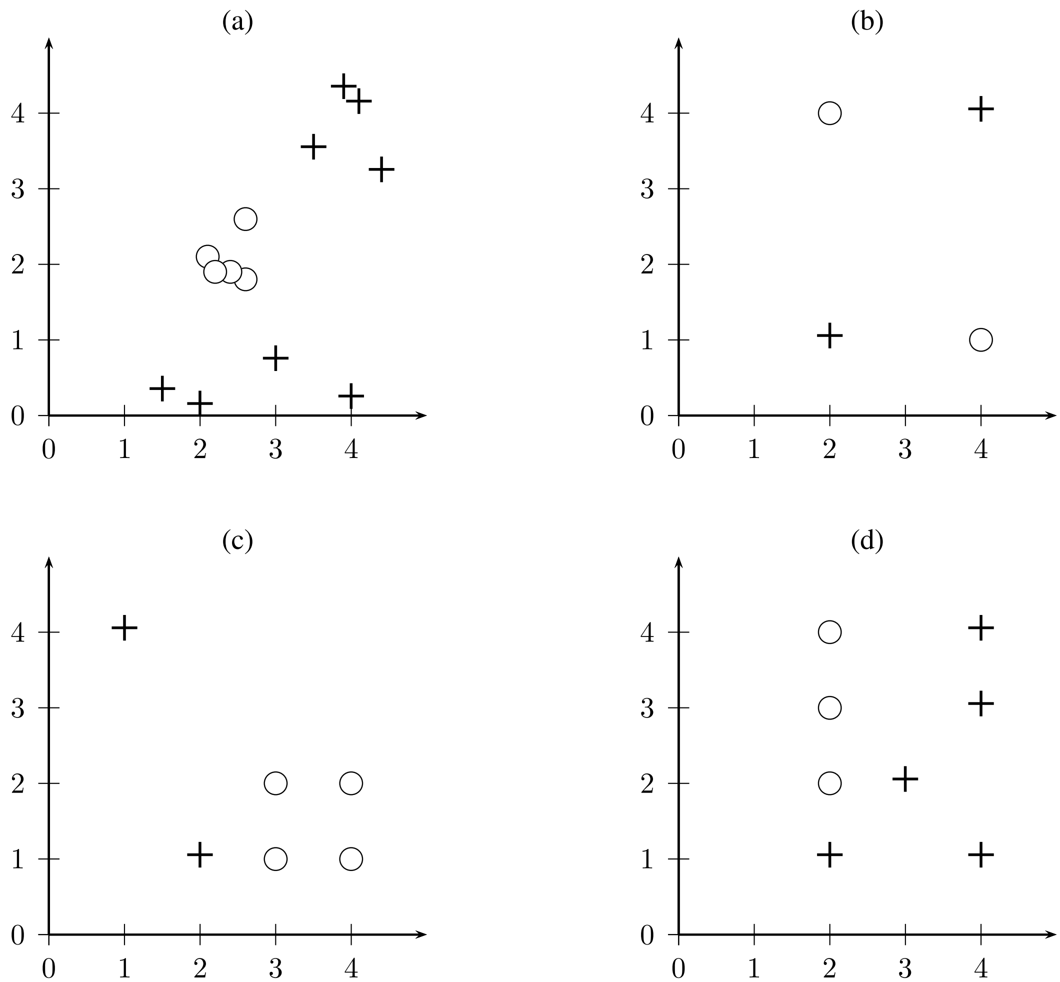
(b) (3 points) How would your Naive Bayes classifier classify the test example: $(F1 = a, F2 = c, F3 = b)$. Be sure to show your work.

Solution:

For $Class = +$: $P(Class = +)P(F1 = a|Class = +)P(F2 = c|Class = +)P(F3 = b|Class = +)$
$= 2/5 * 2/5 * 2/5 * 1/5 = 8/625$
For $Class = -$: $P(Class = -)P(F1 = a|Class = -)P(F2 = c|Class = -)P(F3 = b|Class = -)$
$= 3/5 * 1/3 * 1/2 * 1/3 = 3/90 = 1/30$
Thus, $P(Class = +|example) < P(Class = -|example)$; therefore class is "−".

Question 5: Perceptron (10 points)

(a) (8 points) For each of the following four datasets, indicate whether a perceptron will have zero training error on the dataset. Also, explain why in one sentence or explain by drawing a decision boundary. No credit if the explanation is incorrect. The features are the x and y co-ordinates and the classes are indicated by "$\circ$" (negative class) and "$+$" (positive class) respectively.



- Explanation for (a): **Not linearly separable**.
- Explanation for (b): **Not linearly separable**
- Explanation for (c): **Linearly separable. Draw a line separating positive from negative examples.**
- Explanation for (d): **Linearly separable. Again, just draw a line as before.**

(b) (2 points) Recall that the perceptron is an error-driven classifier while Naive Bayes is a probabilistic classifier. Given this, will the perceptron always have smaller training error as compared with the Naive Bayes classifier. True or False. Explain your answer.

Solution: No. If the data is not linearly separable, Naive Bayes can have smaller error than the perceptron. However, if the data is linearly separable, then the statement is true.