UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Department of Electrical and Computer Engineering

CS 440/ECE 448 Artificial Intelligence
Spring 2020

EXAM 2 SOLUTIONS

Monday, March 30, 2020

Problem 1, Version 1 (5 points)

Consider three binary events, A, B, and C, with probabilities given by $P(A) = 0.7$, $P(B) = 0.4$, and $P(C) = 0.3$.

(a) What's the largest possible $P(B \wedge C)$?

Solution:

$$\max P(B \wedge C) = \min\left(P(B), P(C)\right) = 0.3$$

(b) If A and B are independent, what's $P(A \wedge B)$?

Solution:

$$P(A)P(B) = 0.28$$

Problem 1, Version 2 (5 points)

Consider three binary events, A, B, and C, with probabilities given by $P(A) = 0.7$, $P(B) = 0.4$, and $P(C) = 0.3$.

(a) What's the smallest possible $P(A \wedge B)$?

Solution:

$$\begin{aligned}
\min P(A \wedge B) &= 1 - \max P(\neg(A \wedge B)) \\
&= 1 - \max\left(P(\neg A \vee \neg B)\right) \\
&= 1 - P(\neg A) - P(\neg B) \\
&= 1 - 0.3 - 0.6 = 0.1
\end{aligned}$$

(b) If A and B are independent, what's $P(A \vee B)$?

Solution:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B) = 0.7 + 0.4 - (0.7)(0.4) = 1.1 - 0.28 = 0.82$$

Problem 1, Version 3 (5 points)

Consider three binary events, A, B, and C, with probabilities given by $P(A) = 0.7$, $P(B) = 0.4$, and $P(C) = 0.3$.

(a) What's the largest possible $P(A \vee B)$?

Solution:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B) \leq P(A) + P(B) = 1.1$$

Since $1.1 > 1$, we conclude that $\max P(A \vee B) = 1$.

(b) If A and B are independent, what's $P(A \wedge \neg B)$?

Solution:

$$P(A \wedge \neg B) = P(A)P(\neg B) = (0.7)(0.6) = 0.42$$

Problem 2, Version 1 (5 points)

Consider three binary events, A, B, and C, with probabilities given by $P(A) = 0.7$, $P(B) = 0.4$, and $P(C) = 0.3$.

(a) If B and C are mutually exclusive, what's $P(B \wedge (\neg C))$?

Solution: If B and C are mutually exclusive, then $P(B \wedge C) = 0$, so

$$P(B \wedge (\neg C)) = P(B) = 0.4$$

(b) What's the largest possible $P(\neg(B \wedge C))$?

Solution: $\neg(B \wedge C)$ is the same as $(\neg B \vee \neg C)$.

$$P(\neg B \vee \neg C) = P(\neg B) + P(\neg C) - P(\neg B \wedge \neg C) \leq P(\neg B) + P(\neg C) = 1.3$$

Since $1.3 > 1$, we conclude that the largest possible $P(\neg(B \wedge C)) = 1$, i.e., this would occur if B and C are mutually exclusive.

Problem 2, Version 2 (5 points)

Consider three binary events, A, B, and C, with probabilities given by $P(A) = 0.7$, $P(B) = 0.4$, and $P(C) = 0.3$.

(a) If B and C are mutually exclusive, what's $P(B \wedge C))$?

Solution: If B and C are mutually exclusive, then

$$P(B \wedge C)) = 0$$

(b) What's the smallest possible $P\left(\neg(B \wedge C)\right)$?

Solution:

$$P\left(\neg(B \wedge C)\right) = 1 - P(B \wedge C) \geq 1 - \max P(B \wedge C) = 0.7$$

Problem 2, Version 3 (5 points)

Consider three binary events, A, B, and C, with probabilities given by $P(A) = 0.7$, $P(B) = 0.4$, and $P(C) = 0.3$.

(a) If B and C are mutually exclusive, what's $P((\neg B) \wedge C))$?

Solution: If B and C are mutually exclusive, then

$$P((\neg B) \wedge C) = P(C) = 0.3$$

(b) What's the smallest possible $P\left(\neg(A \vee B)\right)$?

Solution: $\neg(A \vee B)$ is $(\neg A \wedge \neg B)$:

$$\min P\left(\neg(A \vee B)\right) = 1 - \max P(A \vee B) = 0$$

Problem 3, Version 1 (5 points)

You're trying to determine whether a particular newspaper article is of class $Y = 0$ or $Y = 1$. The prior probability of class $Y = 1$ is $P(Y = 1) = 0.4$. The newspaper is written in a language that only has four words, so that the i^{th} word in the article must be $W_i \in \{0, 1, 2, 3\}$, with probabilities given by:

$$\begin{array}{ll} P(W_i=0|Y=0)=0.3 & P(W_i=0|Y=1)=0.1 \\ P(W_i=1|Y=0)=0.1 & P(W_i=1|Y=1)=0.1 \\ P(W_i=2|Y=0)=0.1 & P(W_i=2|Y=1)=0.3 \end{array}$$

The article is only three words long; it contains the words

$$A=(W_1=3,W_2=2,W_3=0)$$

What is $P(Y=1,A)$?

Solution:

$$P(Y=1,A)=P(Y=1)P(W_1=3|Y=1)P(W_2=2|Y=1)P(W_3=0|Y=1)=(0.4)(0.5)(0.3)(0.1)$$

Problem 3, Version 2 (5 points)

You're trying to determine whether a particular newspaper article is of class $Y=0$ or $Y=1$. The prior probability of class $Y=1$ is $P(Y=1)=0.4$. The newspaper is written in a language that only has four words, so that the i^{th} word in the article must be $W_i\in\{0,1,2,3\}$, with probabilities given by:

$$\begin{array}{ll} P(W_i=0|Y=0)=0.1 & P(W_i=0|Y=1)=0.2 \\ P(W_i=1|Y=0)=0.3 & P(W_i=1|Y=1)=0.4 \\ P(W_i=2|Y=0)=0.4 & P(W_i=2|Y=1)=0.1 \end{array}$$

The article is only three words long; it contains the words

$$A=(W_1=1,W_2=1,W_3=1)$$

What is $P(Y=1,A)$?

Solution:

$$P(Y=1,A)=P(Y=1)P(W_1=1|Y=1)P(W_2=1|Y=1)P(W_3=1|Y=1)=(0.4)(0.4)(0.4)(0.4)$$

Problem 3, Version 3 (5 points)

You're trying to determine whether a particular newspaper article is of class $Y=0$ or $Y=1$. The prior probability of class $Y=1$ is $P(Y=1)=0.4$. The newspaper is written in a language that only has four words, so that the i^{th} word in the article must be $W_i\in\{0,1,2,3\}$, with probabilities given by:

$$\begin{array}{ll} P(W_i=0|Y=0)=0.3 & P(W_i=0|Y=1)=0.4 \\ P(W_i=1|Y=0)=0.3 & P(W_i=1|Y=1)=0.4 \\ P(W_i=2|Y=0)=0.1 & P(W_i=2|Y=1)=0.1 \end{array}$$

The article is only three words long; it contains the words

$$A = (W_1 = 2, W_2 = 1, W_3 = 0)$$

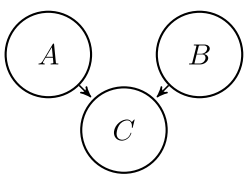
What is $P(Y = 1, A)$?

Solution:

$$P(Y = 1, A) = P(Y = 1)P(W_1 = 2|Y = 1)P(W_2 = 1|Y = 1)P(W_3 = 0|Y = 1) = (0.4)(0.1)(0.4)(0.4)$$

Problem 4, Version 1 (5 points)

Consider the following Bayes network (all variables are binary):



$$P(A) = 0.4, P(B) = 0.1$$

A, B	$P(C\|A, B)$
False,False	0.7
False,True	0.7
True,False	0.1
True,True	0.9

(a) What is $P(C)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(C) &= P(\neg A, \neg B, C) + P(\neg A, B, C) + P(A, \neg B, C) + P(A, B, C) \\ &= (0.6)(0.9)(0.7) + (0.6)(0.1)(0.7) + (0.4)(0.9)(0.1) + (0.4)(0.1)(0.9) \end{aligned}$$

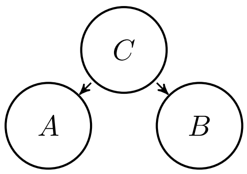
(b) What is $P(A|B = True, C = True)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(A|B, C) &= \frac{P(A, B, C)}{P(A, B, C) + P(\neg A, B, C)} \\ &= \frac{(0.4)(0.1)(0.9)}{(0.4)(0.1)(0.9) + (0.6)(0.1)(0.7)} \end{aligned}$$

Problem 4, Version 2 (5 points)

Consider the following Bayes network (all variables are binary):



$P(C) = 0.1$

C	$P(A\|C)$	$P(B\|C)$
False	0.8	0.7
True	0.4	0.7

(a) What is $P(A)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(A) &= P(\neg C, A) + P(C, A) \\ &= (0.9)(0.8) + (0.1)(0.4) \end{aligned}$$

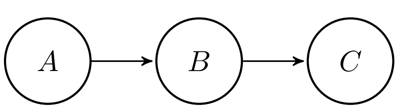
(b) What is $P(C|A = True, B = True)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(C|A, B) &= \frac{P(A, B, C)}{P(A, B, C) + P(A, B, \neg C)} \\ &= \frac{(0.1)(0.4)(0.7)}{(0.1)(0.4)(0.7) + (0.9)(0.8)(0.7)} \end{aligned}$$

Problem 4, Version 3 (5 points)

Consider the following Bayes network (all variables are binary):



$P(A) = 0.8$

A	$P(B\|A)$
False	0.7
True	0.3

B	$P(C\|B)$
False	0.5
True	0.7

(a) What is $P(C)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(C) &= P(\neg A, \neg B, C) + P(\neg A, B, C) + P(A, \neg B, C) + P(A, B, C) \\ &= (0.2)(0.3)(0.5) + (0.2)(0.7)(0.7) + (0.8)(0.7)(0.5) + (0.8)(0.3)(0.7) \end{aligned}$$

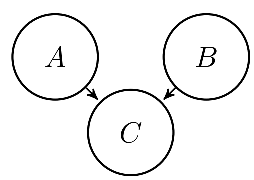
(b) What is $P(A|B = True, C = True)$? Write your answer in numerical form, but you don't need to simplify.

Solution:

$$\begin{aligned} P(A|B,C) &= \frac{P(A,B,C)}{P(A,B,C) + P(\neg A, B, C)} \\ &= \frac{(0.8)(0.3)(0.7)}{(0.8)(0.3)(0.7) + (0.2)(0.7)(0.7)} \end{aligned}$$

Problem 5, Version 1 (5 points)

Consider the following Bayes network (all variables are binary):



You've been asked to re-estimate the parameters of the network based on the following observations:

Observation	A	B	C
1	True	False	False
2	False	False	True
3	True	True	False
4	False	False	False

(a) Given the data in the table, what are the maximum likelihood estimates of the model parameters? If there is a model parameter that cannot be estimated from these data, mark it "UNKNOWN."

Solution: $P(A) = 2/4$, $P(B) = 1/4$, and

A, B	$P(C\|A, B)$
F, F	1/2
F, T	UNKNOWN
T, F	0/1
T, T	0/1

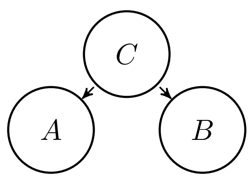
(b) Use the table of data, but this time, estimate the data using Laplace smoothing, with a smoothing parameter of $k = 1$.

Solution: $P(A) = 3/6$, $P(B) = 1/3$, and

A, B	$P(C\|A, B)$
F, F	2/4
F, T	1/2
T, F	1/3
T, T	1/3

Problem 5, Version 2 (5 points)

Consider the following Bayes network (all variables are binary):



You've been asked to re-estimate the parameters of the network based on the following observations:

Observation	A	B	C
1	False	True	False
2	True	True	False
3	False	False	True
4	False	False	True

(a) Given the data in the table, what are the maximum likelihood estimates of the model parameters? If there is a model parameter that cannot be estimated from these data, mark it "UNKNOWN."

Solution: $P(C) = 2/4$, and

C	$P(A\|C)$	$P(B\|C)$
F	1/2	2/2
T	0/2	0/2

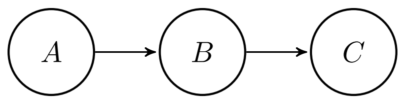
(b) Use the table of data, but this time, estimate the data using Laplace smoothing, with a smoothing parameter of $k = 1$.

Solution: $P(C) = 3/6$, and

C	$P(A\|C)$	$P(B\|C)$
F	2/4	3/4
T	1/4	1/4

Problem 5, Version 3 (5 points)

Consider the following Bayes network (all variables are binary):



You've been asked to re-estimate the parameters of the network based on the following observations:

Observation	A	B	C
1	True	False	False
2	False	False	True
3	True	True	False
4	False	False	False

(a) Given the data in the table, what are the maximum likelihood estimates of the model parameters? If there is a model parameter that cannot be estimated from these data, mark it "UNKNOWN."

Solution:

$$P(A) = 2/4$$

A	$P(B\|A)$
False	0/2
True	1/2

B	$P(C\|B)$
False	1/3
True	0/1

(b) Use the table of data, but this time, estimate the data using Laplace smoothing, with a smoothing parameter of $k = 1$.

Solution: $P(A) = 3/6$, $P(B) = 1/3$, and

$$P(A) = 3/6$$

A	$P(B\|A)$
False	1/4
True	2/4

B	$P(C\|B)$
False	2/5
True	1/3

Problem 6, Version 1 (5 points)

Consider the following probabilistic context-free grammar:

S→	NP VP	$P = 1.0$
NP→	N	$P = 0.9$
NP→	J N	$P = 0.1$
VP→	V	$P = 0.3$
VP→	V NP	$P = 0.7$
J→	beautiful	$P = 0.4$
J→	complicated	$P = 0.6$
N→	birds	$P = 0.8$
N→	flowers	$P = 0.2$
V→	enjoy	$P = 0.5$
V→	grow	$P = 0.5$

Consider the sentence

"Complicated flowers enjoy birds."

What is the probability that this sentence would be generated by the grammar shown above? (Ignore capitalization and punctuation.)

Solution: The rules that fire are

S→	NP VP	$P = 1.0$
NP→	J N	$P = 0.1$
J→	complicated	$P = 0.6$
N→	flowers	$P = 0.2$
VP→	V NP	$P = 0.7$
V→	enjoy	$P = 0.5$
NP→	N	$P = 0.9$
N→	birds	$P = 0.8$

The product of these probabilities is

$$P = (1.0)(0.1)(0.6)(0.2)(0.7)(0.5)(0.9)(0.8)$$

Problem 6, Version 2 (5 points)

Consider the following probabilistic context-free grammar:

$$
\begin{array}{rll}
\text{S}\rightarrow & \text{NP VP} & P = 1.0 \\
\text{NP}\rightarrow & \text{N} & P = 0.9 \\
\text{NP}\rightarrow & \text{J N} & P = 0.1 \\
\text{VP}\rightarrow & \text{V} & P = 0.3 \\
\text{VP}\rightarrow & \text{V NP} & P = 0.7 \\
\text{J}\rightarrow & \text{beautiful} & P = 0.4 \\
\text{J}\rightarrow & \text{complicated} & P = 0.6 \\
\text{N}\rightarrow & \text{birds} & P = 0.8 \\
\text{N}\rightarrow & \text{flowers} & P = 0.2 \\
\text{V}\rightarrow & \text{enjoy} & P = 0.5 \\
\text{V}\rightarrow & \text{grow} & P = 0.5
\end{array}
$$

Consider the sentence

"Birds grow flowers."

What is the probability that this sentence would be generated by the grammar shown above? (Ignore capitalization and punctuation.)

Solution: The rules that fire are

$$
\begin{array}{rll}
\text{S}\rightarrow & \text{NP VP} & P = 1.0 \\
\text{NP}\rightarrow & \text{N} & P = 0.9 \\
\text{N}\rightarrow & \text{birds} & P = 0.8 \\
\text{VP}\rightarrow & \text{V NP} & P = 0.7 \\
\text{V}\rightarrow & \text{grow} & P = 0.5 \\
\text{NP}\rightarrow & \text{N} & P = 0.9 \\
\text{N}\rightarrow & \text{flowers} & P = 0.2
\end{array}
$$

The product of these probabilities is

$$P = (1.0)(0.9)(0.8)(0.7)(0.5)(0.9)(0.2)$$

Problem 6, Version 3 (5 points)

Consider the following probabilistic context-free grammar:

S→	NP VP	$P = 1.0$
NP→	N	$P = 0.9$
NP→	J N	$P = 0.1$
VP→	V	$P = 0.3$
VP→	V NP	$P = 0.7$
J→	beautiful	$P = 0.4$
J→	complicated	$P = 0.6$
N→	birds	$P = 0.8$
N→	flowers	$P = 0.2$
V→	enjoy	$P = 0.5$
V→	grow	$P = 0.5$

Consider the sentence

"Flowers grow complicated flowers."

What is the probability that this sentence would be generated by the grammar shown above? (Ignore capitalization and punctuation.)

Solution: The rules that fire are

S→	NP VP	$P = 1.0$
NP→	N	$P = 0.9$
N→	flowers	$P = 0.2$
VP→	V NP	$P = 0.7$
V→	grow	$P = 0.5$
NP→	J N	$P = 0.1$
J→	complicated	$P = 0.6$
N→	flowers	$P = 0.2$

The product of these probabilities is

$$P = (1.0)(0.9)(0.2)(0.7)(0.5)(0.1)(0.6)(0.2)$$

Problem 7, Version 1 (5 points)

A particular hidden Markov model (HMM) has state variable X_t, and observation variables E_t, where t denotes time. Suppose that this HMM has two states, $X_t \in \{0, 1\}$, and three possible observations, $E_t \in \{0, 1, 2\}$. The initial state probability is $P(X_1 = 1) = 0.3$. The transition and observation probability matrices are

X_{t-1}	$P(X_t = 1 \mid X_{t-1})$	X_t	$P(E_t = 0 \mid X_t)$	$P(E_t = 1 \mid X_t)$
0	0.6	0	0.4	0.1
1	0.4	1	0.1	0.6

Suppose that, in a particular test of the HMM, the observation sequence is

$$\{E_1, E_2\} = \{2, 1\}$$

(a) What is the joint probability $P(X_1 = 1, E_1 = 2, X_2 = 0)$?

Solution:

$$\begin{aligned} P(X_1 = 1, E_1 = 2, X_2 = 0) &= P(X_1 = 1)P(E_1 = 2|X_1 = 1)P(X_2 = 0|X_1 = 1) \\ &= (0.3)(0.3)(0.6) \end{aligned}$$

(b) What is the probability of the most likely state sequence ending in $X_2 = 0$? In other words, what is $\max_{X_1} P(X_1, E_1 = 2, X_2 = 0, E_2 = 1)$?

$$\begin{aligned} \max_{X_1} P(X_1, E_1 = 2, X_2 = 0, E_2 = 1) &= \max_{X_1} P(X_1)P(E_1 = 2|X_1)P(X_2 = 0|X_1)P(E_2 = 1|X_2 = 0) \\ &= \max\left((0.7)(0.5)(0.4)(0.1), (0.3)(0.3)(0.6)(0.1)\right) \\ &= (0.7)(0.5)(0.4)(0.1) \end{aligned}$$

Problem 7, Version 2 (5 points)

A particular hidden Markov model (HMM) has state variable X_t, and observation variables E_t, where t denotes time. Suppose that this HMM has two states, $X_t \in \{0, 1\}$, and three possible observations, $E_t \in \{0, 1, 2\}$. The initial state probability is $P(X_1 = 1) = 0.3$. The transition and observation probability matrices are

X_{t-1}	$P(X_t = 1\|X_{t-1})$	X_t	$P(E_t = 0\|X_t)$	$P(E_t = 1\|X_t)$
0	0.6	0	0.4	0.1
1	0.4	1	0.1	0.6

Suppose that, in a particular test of the HMM, the observation sequence is

$$\{E_1, E_2\} = \{2, 1\}$$

(a) What is the joint probability $P(X_1 = 0, E_1 = 2, E_2 = 1)$?

Solution:

$$\begin{aligned} P(X_1 = 0, E_1 = 2, E_2 = 1) &= \sum_{X_2} P(X_1 = 0)P(E_1 = 2|X_1 = 0)P(X_2|X_1 = 0)P(E_2 = 1|X_2) \\ &= (0.7)(0.5)(0.4)(0.1) + (0.7)(0.5)(0.6)(0.6) \end{aligned}$$

(b) What is the probability of the most likely state sequence ending in $X_2 = 1$? In other words, what is $\max_{X_1} P(X_1, E_1 = 2, X_2 = 1, E_2 = 1)$?

$$\begin{aligned} \max_{X_1} P(X_1, E_1 = 2, X_2 = 1, E_2 = 1) &= \max_{X_1} P(X_1)P(E_1 = 2|X_1)P(X_2 = 1|X_1)P(E_2 = 1|X_2 = 1) \\ &= \max\left((0.7)(0.5)(0.6)(0.6), (0.3)(0.3)(0.4)(0.6)\right) \\ &= (0.7)(0.5)(0.6)(0.6) \end{aligned}$$

Problem 7, Version 3 (5 points)

A particular hidden Markov model (HMM) has state variable X_t, and observation variables E_t, where t denotes time. Suppose that this HMM has two states, $X_t \in \{0,1\}$, and three possible observations, $E_t \in \{0,1,2\}$. The initial state probability is $P(X_1 = 1) = 0.3$. The transition and observation probability matrices are

X_{t-1}	$P(X_t = 1 \vert X_{t-1})$	X_t	$P(E_t = 0 \vert X_t)$	$P(E_t = 1 \vert X_t)$
0	0.6	0	0.4	0.1
1	0.4	1	0.1	0.6

Suppose that, in a particular test of the HMM, the observation sequence is

$$\{E_1, E_2\} = \{2, 1\}$$

(a) What is the joint probability $P(E_1 = 2, X_2 = 1, E_2 = 1)$?

Solution:

$$\begin{aligned} P(E_1 = 2, X_2 = 1, E_2 = 1) &= \sum_{X_1} P(X_1)P(E_1 = 2|X_1)P(X_2 = 1|X_1 = 0)P(E_2 = 1|X_2 = 1) \\ &= (0.7)(0.5)(0.6)(0.6) + (0.3)(0.3)(0.4)(0.6) \end{aligned}$$

(b) If it is observed that $X_2 = 0$, what is the most likely value of X_1? In other words, what is $\arg\max_{X_1} P(X_1, E_1 = 2, X_2 = 0, E_2 = 1)$?

$$\begin{aligned} \arg\max_{X_1} P(X_1, E_1 = 2, X_2 = 0, E_2 = 1) &= \arg\max_{X_1} P(X_1)P(E_1 = 2|X_1)P(X_2 = 0|X_1) \\ &= \arg\max\left((0.7)(0.5)(0.4), (0.3)(0.3)(0.6)\right) \\ &= 0 \end{aligned}$$

Problem 8, Version 1 (5 points)

The horizontal difference-of-Gaussians (DoG) filter is given by

$$h(m,n) = \frac{1}{2\pi\sigma^2} \exp\left(-n^2/2\sigma^2\right) \left(\exp\left(-(m+1)^2/2\sigma^2\right) - \exp\left(-(m-1)^2/2\sigma^2\right)\right)$$

where σ is chosen by the designer. What is the advantage of a larger σ? What is the disadvantage of a larger σ?

Solution: The advantage of a larger σ is that it reduces noise, if the noise is uncorrelated from pixel to pixel. The disadvantage is that it might blur important image features, if the image features have a width less than σ pixels.

Problem 8, Version 2 (5 points)

A particular filter is given by

$$Z(x',y') = \sum_m \sum_n h(m,n) Y(x'-m, y'-n)$$

where $Y(x',y')$ is the input image, $Z(x',y')$ is the filtered image, and

$$h(m,n) = \begin{cases} \frac{1}{25} & -2 \le m \le 2, \ -2 \le n \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Would this filter be more useful for smoothing, or for edge detection? Why?

Solution: This filter would not be useful for edge detection, because all of the coefficients are positive, which means that it doesn't compute any differences between neighboring pixels. It would be more useful for smoothing, because the output, $Z(x',y')$, is the average of 25 pixels in the neighborhood of $Y(x',y')$.

Problem 8, Version 3 (5 points)

If your camera is tilted, then the ground will not appear flat in your image. We can model this by saying that the ground plane is no longer at a constant value of y: instead, the ground plane is given by the equation

$$y = ax + bz + c$$

for some coefficients a, b, and c. The image of the ground plane ends at a horizon line, as $z \to \infty$. Find the equation for the horizon line, as a function of the pixel coordinates (x',y'), of the image focal length f, and of the coefficients a, b, and/or c.

Solution: The pinhole camera equations are

$$\frac{x'}{x} = \frac{y'}{y} = \frac{-f}{z}$$

Combining this with the equation given in the problem, we find that

$$\frac{-y'z}{f} = \frac{-ax'z}{f} + bz + c$$

Taking the limit as $z \to \infty$ gives

$$-\frac{y'}{f} = \frac{-ax'}{f} + b$$

or

$$ax' - y' = bf$$