CS 361 (Fall 2019) Final Exam Cover Sheet

Last Name: | **UIN:** | **Seat Number:**

- **Do not start the exam or turn any pages until we instruct you to do so**.
- **Verify your last name initials and the last 4 digits of your UIN on this cover sheet**. if this booklet does not belong to you, notify us and **do not turn any pages**.
- If this is your booklet, verify that it has **7** sheets of paper including this cover sheet once the exam starts.
- **Show your work unless it's a multiple choice question**.
- This is a closed book exam, and the duration is **150** minutes. **Do not lose the paper clip**.
- You are allowed one 8.5x11-inch page (both sides) of computer typed notes or two 8.5x11-inch pages (both sides) of handwritten notes. Note sheet(s) should be turned in with the exam book.
- **Write your name on your notes sheet(s)**. Failure to write your name on the note sheet(s) will cause a deduction of 2 points from your score.
- You may **not use calculators** or any other electronic devices. Turn off your phone and store it in your bag/backpack.
- **No interaction** between students is allowed during the exam.
- **Put your i-card on the table**. We will check your i-cards during the exam.
- **Do not transfer/attach solutions**. Solutions should only be written on the exam book.
- **Do not write your final answers in the scratch/empty spaces**. Scratch spaces **will not** be scanned.
- **Once you finish, order your exam sheets, put your note sheets at the end, and attach them using the paper clip**.
- If you finish early, **do not move to the podium to submit your exam**. Just raise your hand and do not move from your seat. One of the staff will take it from you shortly, and let you leave.
- **When time is up, do not leave the room while we collect your booklet**. Please sit **silent** while we collect all students' sheets. We will give everyone the green light to leave shortly.

Sign below to verify that you have fully read the instructions, and this booklet belongs to you

Student Signature and Date:_____

Name: | **Netid:** | **UIN:**

Scratch Page 1

This will not be scanned. Do not write any solutions here.

CS 361 (Fall 2019) Final Exam Sheet A Front

Name: | | **Netid:** | | **UIN:** | |

1 First Problem (30 points, 3 points each)

Draw a box around the correct answer. You do not need to show work for this problem.
Question 1 Which is true of the histogram of this data set {-3, -2, -1, 0, 1, 2, 3, 5, 7} ?

a) Left-skewed b) Right-skewed c) Symmetric

Question 2

$$P(A|B^c)P(B^c) + P(A^c|B)P(B) = P(A \cup B) - P(A \cap B)$$

a) True b) False

Question 3
If two random variables are independent, then they are uncorrelated.

a) True b) False

Question 4
If a continuous random variable X is normally distributed with a mean of 0 and a variance of 9, what is $P(|X| < 6)$? Suppose you are given a CDF of the standard normal distribution.

a) 1-CDF(2) b) 2CDF(2)-1 c) CDF(2) d) 2CDF(1)-1

Question 5
Given two independent identically distributed standard normal random variables X and Y, let random variable Z be the difference between them: $Z = X - Y$, the variance of Z is zero.

a) True b) False

CS 361 (Fall 2019) Final Exam Sheet A Back

Name: | **Netid:** | **UIN:**

Question 6
The sample mean of N i.i.d. samples converges to the population mean as N turns to $+\infty$.

a) True b) False

Question 7
A covariance matrix is always symmetric and so always positive definite.

a) True b) False

Question 8
The accuracy of this classifier is better than random pick with equal probability for each class given the following confusion matrix

$$C = \begin{bmatrix} 50 & 0 & 0 \\ 0 & 49 & 5 \\ 0 & 1 & 45 \end{bmatrix}$$

a) True b) False

Question 9
Which are true about the difference between a hierarchical clustering (HC) and k-means clustering?
a) HC produces dendrogram while k-means results in only flat clusters.
b) HC doesn't need to choose the number of clusters while k-means needs that step.
c) HC has higher order time complexity than k-means.
d) All the above.

Question 10
Which of the following is an unsupervised learning method?
a) Principal component analysis.
b) Naive Bayesian classifier.
c) Support vector machine.
d) Random forest classifier.

CS 361 (Fall 2019) Final Exam Sheet B Front

Name: | **Netid:** | **UIN:**

2 Second Problem (30 pts)

1. **(15 points)** For this problem you can leave the answers with choose and fraction. Draw boxes around your answers.

   A drawer contains 4 black, 6 red, and 8 yellow balls.
   (a) **(10 points)** Two balls are selected at random from the drawer without replacement. What is the probability the two balls are of the same color? What is the conditional probability that both balls are black given they are of the same color?

   (b) **(5 points)** If one ball at a time is selected randomly with replacement, what is the expected number of selection (including the last time) for selecting a black ball for the first time?

CS 361 (Fall 2019) Final Exam Sheet B Back

Name: | **Netid:** | **UIN:**

2. **(15 points)** Draw boxes around your answers. If X is a binomial random variable with N=4 and p=1/4.
   (a) **(3 points)** Write down the probability distribution function P(X=k).

   (b) **(12 points)** What are the mean and variance of the random variable Y= 2+4X? Simplify the answers.

CS 361 (Fall 2019) Final Exam Sheet C Front

Name: | **Netid:** | **UIN:**

3 Third Problem (30 pts)

1. **(15 points)** Suppose you are receiving emails with a rate of 10 emails per hour according to a Poisson process. I offer to give you a free movie ticket if you receive exactly 7 emails in a given duration. What duration should you choose to maximize your chance to win the ticket? What is your maximum probability of winning? Show work and draw boxes around your answers.

2. **(15 points)** Suppose you are counting how many people pass by the Einstein Bros' in Siebel center each day. When you complete 36 days, you calculate the 68% confidence interval of the average number of people passing by to be [400, 600]. After you count 64 days more, you realize the mean and the sample unbiased standard deviation haven't changed from the earlier calculation. What is the 95% confidence interval for the average number of people passing by Einstein Bros' each day? Simplify and draw a box around your answer.

CS 361 (Fall 2019) Final Exam Sheet C Back

Name: | **Netid:** | **UIN:**

4 Fourth Problem (30 pts)

1. **(15 points)** You hypothesize that on average a student of UIUC uses one roll of kitchen tissue a month. After counting for 49 students you found a sample mean of 1.5 roll of kitchen tissue per month and the standard error is 0.1 roll. Assess the evidence against the claim. Draw a box around your answer.

2. **(15 points)** You find a 5-sided die and want to estimate its probability θ of coming up 5, you decided to roll it 12 times and then roll it until it comes up 5. You rolled 15 times altogether and found there were 3 times when the die came up 5. Write down the likelihood function $L(\theta)$. Draw a box around your answer.

CS 361 (Fall 2019) Final Exam Sheet D Front

Name: | **Netid:** | **UIN:**

5 Fifth Problem (30 pts)

1. **(15 points)** Show work and simplify to simplest fraction. Draw boxes around your answers. Suppose data set {x} has covariance matrix Covmat({x}) that can be diagonalized as

$$\Lambda = U^T Covmat(\{x\})U.$$

The matrices Λ and U are given as below.

$$\Lambda = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad U = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0.5 & -0.5 & 0.5 & -0.5 \\ 0 & 0.5 & -0.5 & -0.5 & 0.5 \\ 0 & 0.5 & 0.5 & -0.5 & -0.5 \end{bmatrix}$$

In a principal component analysis of {x},
(a) **(5 points)** what is the 2nd principal component of {x}?

(b) **(5 points)** what's the standard deviation of the data {x} corresponding to the 2nd principal component?

(c) **(5 points)** what is the mean square error if we drop the fifth principal component?

CS 361 (Fall 2019) Final Exam Sheet D Back

Name: | **Netid:** | **UIN:**

2. **(15 points)** Show work and draw boxes around your answers. Consider you are clustering a data set {(0,0), (1,0), (1, -1), (3,0), (3,1), (4, 1)}.
a. Find the eventual cluster centers after convergence of the k-means clustering algorithm with initial cluster centers c_1= (1,0) and c_2=(3,0).

b. Also find the eventual cluster centers after convergence if the initial cluster centers are c_1= (0,0) and c_2= (1,0).

CS 361 (Fall 2019) Final Exam Sheet E Front

Name: | **Netid:** | **UIN:**

6 Sixth Problem (30 pts)

1. **(15 points)** Suppose you want to train a linear regression model $y = \beta_0 + \beta_1 x + \beta_2 x^2$ using the training data below. Write down X and $\mathbf{y}$ so that the least-squares estimate for the coefficients is

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = (X^T X)^{-1} X^T \mathbf{y}$$

Draw boxes around your answers.

x	y
-1	-3
0	-2
1	-1
2	3
3	3

2. **(15 points)** Suppose you have two fair dice: a red 5-sided die and a blue 4-sided die. You put the dice in a bag and randomly (with equal probability) pick one to begin the rolling. Whenever the die you have rolled comes up 1, you roll the other die. As you go, you write down the color of the die you have rolled, generating a sequence of symbols Red and Blue.
(a) **(5 points)** If you model this process with Markov chain, what are the initial distribution π?
(b)**(10 points)** What is the transition matrix P for this Markov chain?
Assume that an N-sided die has sides numbered 1 through N. You can leave the numerical answers as fractions. Draw boxes around your answers.

CS 361 (Fall 2019) Final Exam Sheet E Back

Name: | **Netid:** | **UIN:**

7 Seventh Problem (20 pts)

1. **(20 points)** For the training data below, model each conditional probability of the form $P(\mathbf{x}^{(i)}|y)$ as a Poisson distribution. Then use the naïve Bayes assumption to write an expression for

$$\frac{P(y=0|\mathbf{x})}{P(y=1|\mathbf{x})}$$

Draw a box around your answer.

$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	y
4	4	1
6	10	1
5	7	1
2	7	0
0	5	0