CS 361 Sample Final Exam

NAME:

NETID:

CIRCLE YOUR DISCUSSION:

Thu 2-3 **Thu 4-5** **Fri 10-11** **Fri 11-12**

- Be sure that your exam booklet has 8 pages including this cover page
- Make sure to write your name exactly as it appears on your i-card
- Write your netid and circle your discussion section on this page
- **Show your work**
- This is a closed book exam
- You are allowed one handwritten 8.5 x 11-inch sheet of notes (both sides)
- You may **not** use a calculator or any other electronic device
- Turn off your phone and store it in your backpack
- Store away any other electronic devices including earphones and smartwatches
- Absolutely no interaction between students is allowed
- Use backs of pages for scratch work if needed
- Show your i-card when handing in your exam

Problem	1	2	3	4	5	6	7	Total
Possible	30	30	30	30	30	30	20	200
Score								

Problem 1 (30 pts)

1. (15 points) You buy a carton of a dozen eggs and find that exactly one has a double yolk. What is the variance of the number of yolks per egg in this carton? Draw a box around your answer.

2. (15 points) Suppose dataset $\{\mathbf{x}\}$ has the covariance matrix shown below. Calculate the correlation coefficient $\text{corr}(\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}\})$. Draw a box around your answer.

$$\text{Covmat}(\{\mathbf{x}\}) = \begin{bmatrix} 9 & -3 \\ -3 & 4 \end{bmatrix}$$

Problem 2 (30 pts)

1. (15 points) The word game Scrabble contains 100 tiles, of which there are exactly 12 E, 4 U, 4 S and 2 C tiles. If you draw 7 tiles randomly from the original 100 without replacement, what is the probability that you can arrange the tiles to form the word SUCCESS (without using blank tiles)? You may use factorials and choose notation in your answer. Draw a box around your answer.

2. (15 points) You have been invited to a fishing game with a twist. If you catch no fish in an hour, you will pay \$100. If you catch 1 fish in that hour, you will win \$40. If you catch 2 or more fish, you will win \$80. Some research tells you that catching fish in this pond is a Poisson process with intensity of 3 fish per hour. Should you play the game? Justify your answer with calculations. You may approximate e^{-3} as 0.05.

Problem 3 (30 pts)

1. (15 points) Suppose I offer to give you a prize if I toss a fair coin n times and it comes up heads k of those times. If I require that $n \geq 2$, what values of n and k should you choose to maximize your chance of winning? What is your maximum probability of winning? Draw a box around your answer.

2. (15 points) You asked 5 of your classmates what their scores were on Midterm 2 and then used Student's t-distribution to calculate that the 95% confidence interval for the population mean score is $[96, 144]$. Why is it wrong to conclude that the population mean score falls in the interval $[96, 144]$ with probability 95%?

Problem 4 (30 pts)

1. (15 points) You hypothesize that the average product rating on Amazon is 4 stars. From a sample of 100 items, you find a sample mean of 4.4 stars and a sample (unbiased) standard deviation of 0.1 stars. Assess the evidence against the claim.

2. (15 points) Last week you flipped a coin 10 times. The coin came up on the same side all 10 times, but you've forgotten if it was all heads or all tails. Write down the likelihood function $L(\theta)$ for the coin's probability of coming up heads. Draw a box around your answer.

Problem 5 (30 pts)

1. (15 points) Suppose you want to train a classifier using the training data below. Explain why you should **not** start by projecting the dataset $\{\mathbf{x}\}$ on to its first principal component.

$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	label
5	1	Y
-5	1	Y
5	-1	N
-5	-1	N

2. (15 points) Given a dataset $\{0, 2, 4, 6, 24, 26\}$, initialize the k-means clustering algorithm with 2 cluster centers $c_1 = 3$ and $c_2 = 4$. What are the values of c_1 and c_2 after one iteration of k-means? What are the values of c_1 and c_2 after the second iteration of k-means? Draw a box around your answer.

Problem 6 (30 pts)

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

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (X^T X)^{-1} X^T \mathbf{y}$$

Draw a box around your answer.

x	y
1	17
2	15
3	10
4	2

2. (15 points) Draw the state transition diagram (i.e. the directed graph) for the Markov chain with transition probability matrix P given below. Make sure to label the edges of the graph with the appropriate probabilities.

$$P = \begin{bmatrix} 0.1 & 0.6 & 0.3 \\ 0.9 & 0 & 0.1 \\ 0 & 1 & 0 \end{bmatrix}$$

Problem 7 (20 pts)

1. (20 points) For the training data below, model each conditional probability of the form $P(\mathbf{x}^{(i)}|y)$ as a normal 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	7	0
-4	5	0
2	10	1
10	4	1