Final examination

Name: PRASHANT KOLKUR

Be clear and concise. Write your answers in the space provided. Use the backs of pages for scratchwork.

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	

TOTAL POINTS: 80

1. **(10 points)** You are dealt two cards at random from a standard deck. What is the probability that:

(a) The first card is an ace?

$$\frac{1}{13}$$

(b) The first and second cards are both aces?

$$\frac{1}{221}$$

(c) The second card is an ace?

$$\frac{1}{13}$$

(d) The first card is an ace, given that it is a heart?

$$\frac{1}{13}$$

(e) The second card is an ace, given that the first card is an ace?

$$\frac{1}{17}$$

2. **(3 points)** Ten cards are chosen at random from a standard deck. Which of the following pairs of events A, B are independent? Circle them.

- A: first card is a ten, B: tenth card is a nine   Dependent

- A: first card is a ten, B: second card is a heart   Independent

- A: second card is a heart, B: fifth card is a club   Dependent

3. **(10 points)** Short answer questions.

(a) The letters G, H, I, R, T are randomly permuted. What is the probability that the result is the word R, I, G, H, T?

$$\frac{1}{120}$$

(b) Three fair dice are rolled. What is the probability that they all have the same value?

$$\frac{1}{36}$$

(c) Each time you go to the gym, you have a 20% chance of running into your worst enemy. What is the expected number of trips to the gym before you meet this person?

$$5$$

(d) A certain population consists of 40% men and 60% women. Of the men, 20% are left-handed, and of the women, 10% are left-handed. A person is picked at random from this population and is found to be left-handed. What is the probability that this person is female?

$$\frac{3}{7}$$

(e) A man has a bottle containing ten identical-looking pills. Two of them contain medicine while the other 8 are placebos. Upon taking a pill, the man feels either **good** or **not good**, with the following probabilities:

$$\text{Pr}(\text{feel good} \mid \text{medicine}) = \frac{3}{4}$$
$$\text{Pr}(\text{feel good} \mid \text{placebo}) = \frac{1}{2}$$

Today, the man picks a pill at random and finds that he feels good. What is the probability that the pill contained medicine?

$$\frac{3}{11}$$

4. **(8 points)** A die has six sides that come up with different probabilities.

$$\Pr(1) = \Pr(2) = \Pr(3) = \frac{1}{12}, \quad \Pr(4) = \Pr(5) = \Pr(6) = \frac{1}{4}.$$

(a) You roll the die; let X denote the outcome. What is $\mathbb{E}(X)$?

4.25

(b) What is $\text{var}(X)$?

2.354

(c) Now you roll this die a hundred times, and let Z be the sum of all the rolls. What is $\mathbb{E}(Z)$?

425

(d) What is $\text{var}(Z)$?

235.4

5. **(3 points)** A pair of random variables X_1 and X_2 have the following properties:

- They both take values in $\{-1, 1\}$
- X_1 has mean 0 while X_2 has mean 0.5
- The correlation between X_1 and X_2 is 0.25

Suppose we fit a (bivariate) Gaussian to (X_1, X_2). Give the mean and covariance matrix of this Gaussian.

$$\text{Mean} = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix}$$

$$\text{Cov} = \begin{bmatrix} 1 & 0.2165 \\ 0.2165 & 0.75 \end{bmatrix}$$

6. (10 points) A certain random variable $X \in \mathbb{R}^3$ has mean and covariance as follows

$$EX = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \text{cov}(X) = \begin{pmatrix} 5 & 3 & 0 \\ 3 & 5 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

(a) The eigenvectors of cov(X) can be found in the following list:

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \frac{1}{\sqrt 2}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \frac{1}{\sqrt 2}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt 2}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

Circle them.

(b) Find the eigenvalues corresponding to each of the eigenvectors in part (a). Make it clear which eigenvalue belongs to which eigenvector.

$\lambda = 2$, Eigen vector $= \frac{1}{\sqrt 2}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ $\qquad \lambda = 8$, Eigen vector $= \frac{1}{\sqrt 2}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$

$\lambda = 4$, Eigen vector $= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

(c) Suppose we used principal component analysis (PCA) to project points X into *two* dimensions. Which directions would it project onto?

$\frac{1}{\sqrt 2}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

(d) Continuing from part (c), what would be the resulting two-dimensional projection of the point $x = (4, 0, 2)$?

$\left(2\sqrt 2, 2 \right)$

(e) Continuing from part (d), suppose that starting from the 2-d projection, we tried to reconstruct the original x. What would the three-dimensional reconstruction be, exactly?
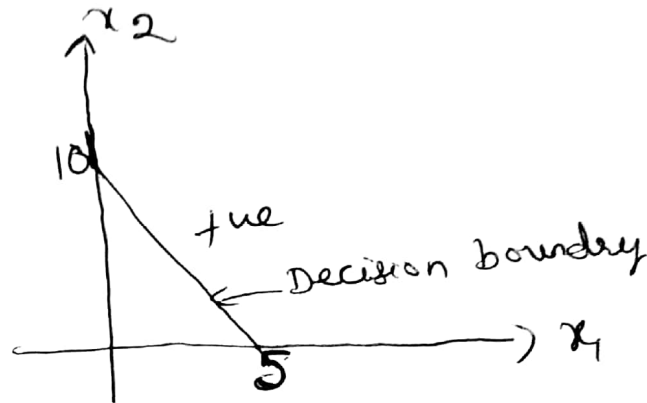
$\begin{bmatrix} 2, & -2, & 2 \end{bmatrix}$

7. **(4 points)** Consider the linear classifier $w \cdot x \geq \theta$, where

$$w = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \theta = 10.$$

Sketch the decision boundary in $\mathbb{R}^2$. Make sure to indicate where the boundary intersects the two axes, and which side of the boundary is the positive side.

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \geq 10$$

$$2x_1 + x_2 \geq 10$$

8. **(4 points)** A survey is taken to determine what fraction of freshman computer science majors have prior programming experience. Call this unknown fraction p. Out of the nationwide pool of computer science freshmen, 100 are chosen at random. Of them, 40% had prior programming experience.

   (a) The natural estimate of p is 0.4. Give a 95% confidence interval for the estimate.

$$[0.302, \ 0.498]$$
$$[30.2\%, \ 49.8\%]$$

   (b) Suppose we now want to estimate p more accurately, to within a 95% confidence interval of ± 0.01. What sample size should we use?

$$9600$$

9. **(2 points)** A school wants to determine the average number of hours that the students spend on homework; call this unknown number μ. 100 students are chosen at random, and each of them is asked to report the typical number of hours per week that he or she spends on homework. The reported numbers have a mean of 12.2 and a standard deviation of 5.4. Give a 95% confidence interval for μ.

$$[11.12, 13.28]$$

10. **(6 points)** Genius Academy is a high school that claims to prepare its students exceptionally well for the SAT exam. A random sample is taken of 100 Genius Academy students, and their SAT scores turn out to have a mean of 1930 and a standard deviation of 150. A random sample is also taken of 100 students from the other local high school, and their scores have a lower mean, of 1860, with a standard deviation of 200.

We wish to determine whether the difference between these observed averages is significant.

(a) State the null hypothesis.

Mean SAT Scores of Grenius academy students and local high school should be same

(b) Compute a suitable z-statistic for this situation.

$$Z = -2.8$$

(c) What is the p value, and what conclusion would you draw?

P value = 0.002555 (For significance level = 0.05)

Conclusion:— This is a strong evidence against null hypothesis

11. **(20 points)** For this last problem, you should turn in an iPython notebook.

Download the IRIS data set from:

`https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data`

This is a data set of 150 points in $\mathbb{R}^4$, with three classes; refer to the website for more details of the features and classes.

(a) Use a PCA projection to 2-d to visualize the entire data set. You should plot different classes using different colors/shapes. Do the classes seem well-separated from each other?

(b) Now build a classifier for this data set, based on a generative model (you can choose whichever you like).

- Split the data set into training/test data as follows: use the first 35 points in each class for training, and use the remaining 15 points for testing.
- What error rate do you achieve?