HOMEWORK 1

AKASH SHARMA

9081731771

Instructions: This is a background self-test on the type of math we will encounter in class. If you find many questions intimidating, we suggest you drop 760 and take it again in the future when you are more prepared.

Use this latex file as a template to develop your homework. Submit your homework on time as a single pdf file to Canvas. There is no need to submit the latex source or any code. Please check Piazza for updates about the homework.

1 Logistics [5 pts]

Read the course webpage at http://pages.cs.wisc.edu/~jerryzhu/cs760.html and answer the following questions:

1. Where do we post announcements and clarifications?
   Piazza

2. What time of day (hour:minute) are all homeworks due?
   10:59 AM

3. Will late homework be accepted?
   No, late homework will not be accepted.

4. Tom received the following scores on his 7 homeworks: 0, 59, 92, 95, 98, 100, 100. According to the homework policies, what is Tom's final average homework score?
   Lowest score 0 will be dropped from final average calculation. Final average is 90.67.

5. How can you discuss homework questions with fellow students while avoiding the impression of cheating?

   We can discuss the homework approaches and techniques broadly with other fellow students. However, all examinations, programming assignments, and written homeworks must be written up individually. Also, for programming assignments the code must not be developed in groups, nor should code be shared. We need to make sure to work independently on all the problems and write our own final writeups.

2 Vectors and Matrices [2 pts]

Consider the matrix X and the vectors $\mathbf{y}$ and $\mathbf{z}$ below:

$$X = \begin{pmatrix} 6 & 7 \\ 8 & 9 \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \qquad \mathbf{z} = \begin{pmatrix} 7 \\ 6 \end{pmatrix}$$

1. Computer $\mathbf{y}^\top X \mathbf{z}$

   $\mathbf{y}^T = \begin{pmatrix} 2 & 3 \end{pmatrix}$

   $\mathbf{y}^T X = \begin{pmatrix} 36 & 41 \end{pmatrix}$

   $\mathbf{y}^\top X \mathbf{z} = 36 * 7 + 41 * 6 = 498$

2. Is X invertible? If so, give the inverse, and if no, explain why not.

Yes, X is Invertible as its determinant is non-zero. The determinant is: 6*9 – 7*8 = 54 – 56 = -2. Inverse of X is,

$$\begin{pmatrix} -4.5 & 3.5 \\ 4 & -3 \end{pmatrix}$$

3   Calculus [1 pts]

1. If $y = e^x + \tan(z)x^{6z} - \ln(\frac{7x+z}{x^4})$, what is the partial derivative of y with respect to x?

Partial derivative of the above expression is:
$$e^x + 6z \tan(z)x^{6z-1} + (4/x)–(7/(7x + z))$$

4   Probability and Statistics [4 pts]

Consider a sequence of data $S = (0, 1, 1, 0, 1, 1, 1)$ created by flipping a coin x seven times, where 0 denotes that the coin turned up heads and 1 denotes that it turned up tails.

1. What is the probability of observing this data, assuming it was generated by flipping a biased coin with $p(x = 1) = 0.7$?

   $p(x = 1) = 0.7$ $p(x = 0) = 0.3$

   Probablility of observing this data = 0.3 * 0.7 * 0.7 * 0.3 * 0.7 * 0.7 * 0.7 = 0.0151263

2. Note that the probability of this data sample could be greater if the value of $p(x = 1)$ was not 0.7, but instead some other value. What is the value that maximizes the probability of S? Please justify your answer.

   Let $p(x = 1) = t$, hence, the Probablility of S becomes: $t^5 * (1 - t)^2$.
   Taking partial derivative to get the value of t to maximize the above, we get

   $$5 * (1 - t) = 2 * t$$

   Hence, $t = p(x = 1) = 5/7$

3. Consider the following joint probability table where both A and B are binary random variables:

A	B	$P(A, B)$
0	0	0.4
0	1	0.3
1	0	0.2
1	1	0.1

   (a) What is $P(A = 0|B = 1)$?

   Conditional Probablity of A = 0,given that B = 1 is:
   $P(A = 0 \cap B = 1)/P(B = 1) = 0.3/0.4 = = 0.75$

   (b) What is $P(A = 0 \lor B = 0)$?

   $P(A = 0 \lor B = 0)$ = 1 - $P(A = 1 \cap B = 1) = 1 - 0.1 = 0.9$

5 Big-O Notation [3 pts]

For each pair (f, g) of functions below, list which of the following are true: $f(n) = O(g(n))$, $g(n) = O(f(n))$, both, or neither. Briefly justify your answers.

1. $f(n) = \frac{n}{2}$, $g(n) = \log_2(n)$.
   f(n) = O(g(n)) if there exists a positive integer n0 and a positive constant c, such that f(n) is less than or equal to cg(n) for all n greater than or equal to n0.

   $\log_2(n)$ is less than $c(n/2)$ for all n greater than 4. Therefore, $g(n) = O(f(n))$

2. $f(n) = \ln(n)$, $g(n) = \log_2(n)$.

   Saying g (n) = O(f (n)) means g (n) / f (n) = O(1) or c as n tends to infinity. Here, both f(n)/g(n) and g(n)/f(n) tends to a constant. Hence here both $f(n) = O(g(n))$ and $g(n) = O(f(n))$ are true.

3. $f(n) = n^{100}$, $g(n) = 100^n$.

   Here, for values of n $\geq$ 1000, the value of g(n) > f(n). Also, as n tends to infinity, the value of f(n)/g(n) tends to a constant, while g(n)/f(n) goes to undefined. Hence, $f(n) = O(g(n))$

6 Probability and Random Variables

6.1 Probability [5 pts]

State true or false. Here Ω denotes the sample space and A^c denotes the complement of the event A.

1. For any $A, B \subseteq \Omega$, $P(A|B)P(B) = P(B|A)P(A)$.

   TRUE

2. For any $A, B \subseteq \Omega$, $P(A \cup B) = P(A) + P(B) - P(A|B)$.

   FALSE

3. For any $A, B, C \subseteq \Omega$ such that $P(B \cup C) > 0$, $\frac{P(A \cup B \cup C)}{P(B \cup C)} \geq P(A|B \cup C)P(B \cup C)$.

   TRUE

4. For any $A, B \subseteq \Omega$ such that $P(B) > 0, P(A^c) > 0$, $P(B|A^C) + P(B|A) = 1$.

   FALSE

5. For any n events $\{A_i\}_{i=1}^n$, if $P(\bigcap_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$, then $\{A_i\}_{i=1}^n$ are mutually independent.

   FALSE

6.2 Discrete and Continuous Distributions [5 pts]

Match the distribution name to its probability density / mass function. Below, $|\boldsymbol{x}| = k$.

(f) $f(\boldsymbol{x}; \boldsymbol{\Sigma}, \boldsymbol{\mu}) = \frac{1}{\sqrt{(2\pi)^k \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right)$

(g) $f(x; n, \alpha) = \binom{n}{x}\alpha^x(1 - \alpha)^{n-x}$ for $x \in \{0, \ldots, n\}$; 0 otherwise

(a) Laplace h

(b) Multinomial i

(c) Poisson l

(d) Dirichlet k

(e) Gamma j

(h) $f(x; b, \mu) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$

(i) $f(\boldsymbol{x}; n, \boldsymbol{\alpha}) = \frac{n!}{\Pi_{i=1}^k x_i!}\Pi_{i=1}^k \alpha_i^{x_i}$ for $x_i \in \{0, \ldots, n\}$ and $\sum_{i=1}^k x_i = n$; 0 otherwise

(j) $f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ for $x \in (0, +\infty)$; 0 otherwise

(k) $f(\boldsymbol{x}; \boldsymbol{\alpha}) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k x_i^{\alpha_i-1}$ for $x_i \in (0, 1)$ and $\sum_{i=1}^k x_i = 1$; 0 otherwise

(l) $f(x; \lambda) = \lambda^x \frac{e^{-\lambda}}{x!}$ for all $x \in Z^+$; 0 otherwise

6.3 Mean and Variance [5 pts]

1. Consider a random variable which follows a Binomial distribution: $X \sim \text{Binomial}(n, p)$.

   (a) What is the mean of the random variable?
   $n * p$

   (b) What is the variance of the random variable?
   $n * p * (1 - p)$

2. Let X be a random variable and $\mathbb{E}[X] = 1, \text{Var}(X) = 1$. Compute the following values:

   (a) $\mathbb{E}[3X]$
   $E[aX + b] = aE[X] + b \; E[3X] = 3 * 1 = 3$

   (b) $\text{Var}(3X)$
   $Var[aX + b] = a^2 Var(X)$
   $Var[3X] = 9 * Var[X] = 9$

   (c) $\text{Var}(X + 3)$
   $Var(X + 3) = 1$

6.4 Mutual and Conditional Independence [4 pts]

1. If X and Y are independent random variables, show that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
   $E(X, Y) = \sum_i \sum_j x_i y_j f_{x,y}(x_i, y_j)$
   Since, the events X and Y are independent, $f_{xy}(xi, yj) = f_x(xi)f_y(yj)$
   $E(X, Y) = \sum_i \sum_j x_i y_j f_x(xi)f_y(yj)$
   $E(X, Y) = \sum_i x_i f_x(xi) \sum_j y_j f_y(yj) = E(X)E(Y)$
   Hence, proved

2. If X and Y are independent random variables, show that $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.
   Hint: $\text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y)$

   Since, X and Y are independent random variables, the covariance of X and Y is 0, that is $Cov(X, Y) = 0$
   Since, $Var(X + Y) = Var(X) + 2Cov(X, Y) + Var(Y)$
   Using above, $Var(X + Y) = Var(X) + 2 * (0) + Var(Y) = Var(X) + Var(Y)$
   Hence, proved

3. If we roll two dice that behave independently of each other, will the result of the first die tell us something about the result of the second die?

   Since both the dices are independent of each other, the outcome of the first does not tell us anything about that of the second dice.

   If, however, the first die's result is a 1, and someone tells you about a third event — that the sum of the two results is even — then given this information is the result of the second die independent of the first die?

   Since, the sum of numbers is even, the events are not conditionally independent. But the result of the first dice is independent of the first die.

6.5   Central Limit Theorem [1 pts]

Provide one line explanation.

1. Let $X_i \sim \mathcal{N}(0, 1)$ and $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, then the distribution of $\bar{X}$ satisfies

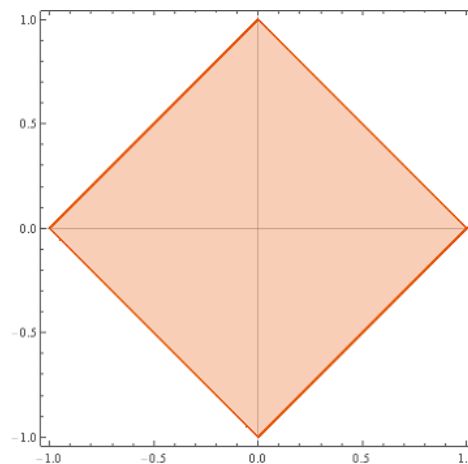$$\sqrt{n}\bar{X} \overset{n \to \infty}{\Longrightarrow} \mathcal{N}(0, 1)$$

The central limit theorem states that the distribution of sum of independent random variables approch a normal distribution,when the sample size gets very large.

7   Linear algebra

7.1   Norms [3 pts]

Draw the regions corresponding to vectors $\mathbf{x} \in \mathbb{R}^2$ with the following norms:

1. $||\mathbf{x}||_1 \leq 1$ (Recall that $||\mathbf{x}||_1 = \sum_i |x_i|$)

2. $||\mathbf{x}||_2 \leq 1$ (Recall that $||\mathbf{x}||_2 = \sqrt{\sum_i x_i^2}$)

3. $||\mathbf{x}||_\infty \leq 1$ (Recall that $||\mathbf{x}||_\infty = \max_i |x_i|$)


1. Norm1

7.2   Geometry [2 pts]

Prove the following. Provide all steps.

1. The smallest Euclidean distance from the origin to some point $\mathbf{x}$ in the hyperplane $\mathbf{w}^\top \mathbf{x} + b = 0$ is $\frac{|b|}{||\mathbf{w}||_2}$. You may assume $\mathbf{w} \neq 0$.

The equation of the shortest euclidean distance from a point to the plane is along the normal to the plane. unit normal vector = $w^T/(||w||)$. Shortest distance is the absolute value of the dot product vn (projection on normal vector), where v is the equation of external point to a point on hyperplane, that is, $d = (|w^T x1 - w^T x0|)/||w||$ Putting the value of origin as x0, we get the smallest distance as the above.

2. The Euclidean distance between two parallel hyperplane $\mathbf{w}^\top \mathbf{x} + b_1 = 0$ and $\mathbf{w}^\top \mathbf{x} + b_2 = 0$ is $\frac{|b_1 - b_2|}{||\mathbf{w}||_2}$ (Hint: you can use the result from the last question to help you prove this one).

The smallest Euclidean distance from origin to plane1 (as proved above) = $\frac{|b1|}{||\mathbf{w}||_2}$. Similarly, the smallest Euclidean distance from origin to plane2 = $\frac{|b2|}{||\mathbf{w}||_2}$. Since, the planes are parallel, distance between planes is
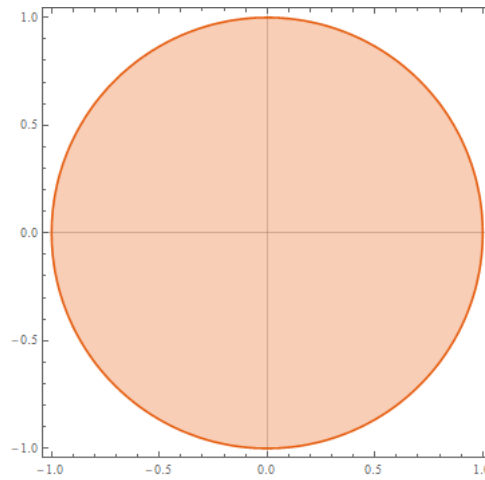
2. Norm2
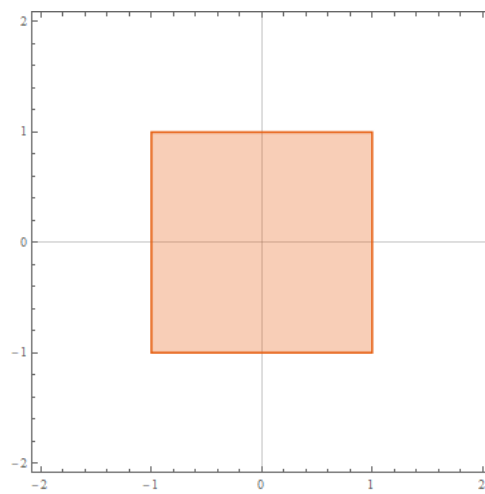
3. Norm Infinity

equal to the difference of smallest distances from origin to each of the planes, that is,
$$\frac{|b2|}{||\mathbf{w}||_2} - \frac{|b1|}{||\mathbf{w}||_2} = \frac{|b2-b1|}{||\mathbf{w}||_2}.$$

8 Programming Skills [3 pts]

Sampling from a distribution. For each question, submit a scatter plot (you will have 5 plots in total). Make sure the axes for all plots have the same ranges.

1. Draw 100 items $\mathbf{x} = [x_1, x_2]^\top$ from a 2-dimensional Gaussian distribution $N(\mu, \Sigma)$ with mean $\mu = (0,0)^T$ and identity covariance matrix $\Sigma = I$, i.e., $p(\mathbf{x}) = \frac{1}{2\pi} \exp\left(-\frac{||\mathbf{x}||^2}{2}\right)$, and make a scatter plot (x_1 vs. x_2).

2. Make a scatter plot by drawing 100 items from $N(\mu + (1, -1)^\top, 2\Sigma)$.

3. Make a scatter plot by drawing 100 items from a mixture distribution $0.3N\left((1,0)^\top, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}\right) + 0.7N\left((-1,0)^\top, \begin{pmatrix} 1 & -0.2 \\ -0.2 & 1 \end{pmatrix}\right)$.

Figure: 1

Figure: 2

Figure: 3