HOMEWORK 1

>>Sean(Xiaoyu) Sun<<
>>9078202463<<

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?

   The instructors and TAs will post announcements, clarifications on Piazza. And also http://pages.cs.wisc.edu/ jerryzhu/cs760.html

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

   Homework is always due the minute before class starts on the due date.

3. Will late homework be accepted?

   NO

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?

   The lowest score will be dropped. So it will be (59+92+95+98+100+100)/6 = 90.667

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

   Declare it in the solution

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 $\mathbf{y}^\top X \mathbf{z}$
   498

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

   It is veritble because its determinant is non-zero. $X^{-1} = \begin{pmatrix} -4.5 & 3.5 \\ 4 & -3 \end{pmatrix}$

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?
   $e^x + 6z\tan(z)x^{6z-1} + \frac{21x+4z}{7x^2+zx}$

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 $p(x = 1) = 0.7$?
   $0.7^5 * 0.3^2 = 0.015$

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.
   $f(p) = p^5(1 - p)^2$
   $f'(p) = p^5 + p^7 - 2p^6$
   given p is in the range of (0,1), when p = 5/7, f'(p)=0, which means f(p) reaches the maximum.

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 $P(A = 0|B = 1)$?
       $P(A = 0|B = 1) = \frac{0.3}{0.3+0.1} = \frac{3}{4}$
   (b) What is $P(A = 0 \lor B = 0)$?
       $P(A = 0 \lor B = 0) = 1 - 0.1 = 0.9$

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. $f(n) = \frac{n}{2}$, $g(n) = \log_2(n)$.
   $g(n) = O(f(n))$
   beacue when $n > 2, |g(n)| < |f(n)|$

2. $f(n) = \ln(n)$, $g(n) = \log_2(n)$.
   $f(n) = O(g(n))$
   beacue when $n > 1, |f(n)| < |g(n)|$

3. $f(n) = n^{100}$, $g(n) = 100^n$.
   $f(n) = O(g(n))$
   beacue there is an N that when $n > N, |f(n)| < |g(n)|$

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. It should be $P(A \cup B) = P(A) + P(B) - P(AB)$.

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 G 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) $f(\boldsymbol{x}; n, \boldsymbol{\alpha}) = \frac{n!}{\prod_{i=1}^k x_i!}\Pi_{i=1}^k \alpha_i^{x_i}$ 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?
      np

   (b) What is the variance of the random variable?
      np(1-p)

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

   (a) $\mathbb{E}[3X]$
      3.

   (b) $\text{Var}(3X)$
      9

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

6.4 Mutual and Conditional Independence [4 pts]

1. If X and Y are independent random variables, show that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

   $\mathbb{E}[XY] = P(XY)XY$
   since X and Y are independent, so $P(XY) = P(X)P(Y)$
   so $\mathbb{E}[XY] = P(X)XP(Y)Y = \mathbb{E}[X]\mathbb{E}[Y]$

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 $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$
   so $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

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?

   If they are same dice, the variance and mean value of the first die should be consistent with the second one.

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?

No

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

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

According to CLT, when enough observations is made, the average will be closely approximated by a normal distribution.
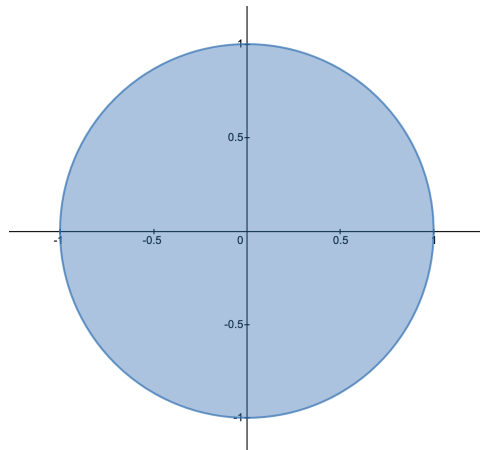
7 Linear algebra

7.1 Norms [3 pts]

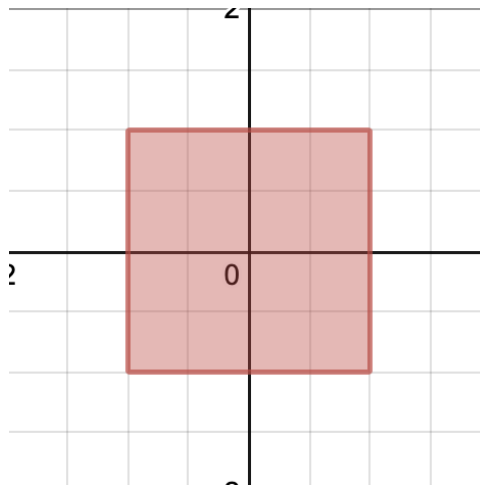Draw the regions corresponding to vectors $\mathbf{x} \in \mathbb{R}^2$ with the following norms:

1. $||\mathbf{x}||_1 \leq 1$ (Recall that $||\mathbf{x}||_1 = \sum_i |x_i|$)
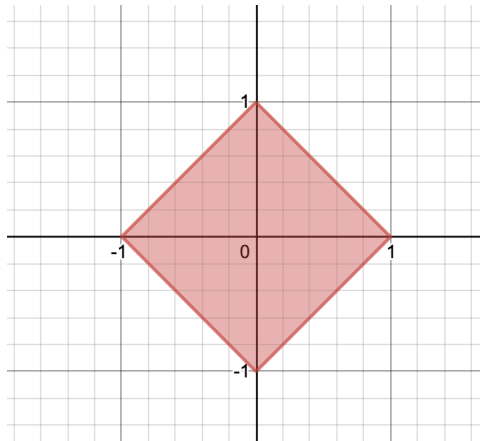
2. $||\mathbf{x}||_2 \leq 1$ (Recall that $||\mathbf{x}||_2 = \sqrt{\sum_i x_i^2}$)

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

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 smallest Euclidean distance from a point P to a hyper plane is $\frac{|\mathbf{w}^\top\mathbf{P}+b|}{\sqrt{|w|^2}}$

   so, when P is the orgin, the distance is $\frac{|b|}{||\mathbf{w}||_2}$.

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).

   According to question 1, since these two hyperplanes are parallel, every point P1 in plane 1 will have same smallest distance to plane 2 and this distance is also the smallest distance b/w two planes. So we can get the

smallest distance as $\frac{|\mathbf{w}^\top \mathbf{P1} + b_2|}{\sqrt{|w|^2}}$. According to the equations of plane 1, we can get $\mathbf{w}^\top \mathbf{P1} = -b_1$. So we can write the distance as $\frac{|-b_1 + b_2|}{\sqrt{|w|^2}} = \frac{|b_1 - b_2|}{\sqrt{|w|^2}} = \frac{|b_1 - b_2|}{||\mathbf{w}||_2}$

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).
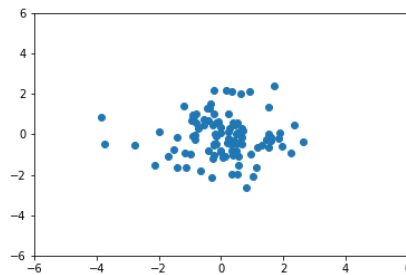
2. Make a scatter plot by drawing 100 items from $N(\mu + (1, -1)^\top, 2\Sigma)$.
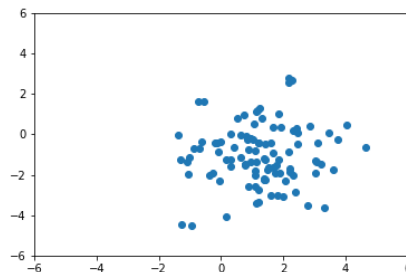
3. Make a scatter plot by drawing 100 items from a mixture distribution $0.3N\left((1, 0)^\top, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}\right) +$ $0.7N\left((-1, 0)^\top, \begin{pmatrix} 1 & -0.2 \\ -0.2 & 1 \end{pmatrix}\right)$.