DSB205, Winter 2025
Problem Set 0: Math prerequisites
Due Jan 17, 2025 at 11:59pm

Submission instructions

- Submit your solutions electronically on the course Gradescope site as PDF files.
- If you plan to typeset your solutions, please use the LaTeX solution template. If you must submit scanned handwritten solutions, please use a black pen on blank white paper and a high-quality scanner app.

This course assumes that you have a basic familiarity with several types of math (Calculus, Linear algebra, Probability and Statistics). Before taking the class, you should evaluate whether you have the mathematical background the class depends upon.

If you cannot pass this test, we suggest you fill in your math background before taking the class.

You may find the following resources helpful:

- Andrew Ng's CS229 Course (Stanford)
  - Linear Algebra Review (http://cs229.stanford.edu/section/cs229-linalg.pdf)
  - Probability Theory Review (http://cs229.stanford.edu/section/cs229-prob.pdf)

1 Multivariate Calculus [2 pts]

Consider $y = (ax_1 + bx_2)^2$. What is the partial derivative of y with respect to x_2?

$$\frac{\partial f}{\partial x_2}(ax_1 + bx_2)^2$$

$$= 2(ax_1 + bx_2) \times \frac{\partial f}{\partial x_2}(ax_1 + bx_2)$$

$$= 2(ax_1 + bx_2) \times (b)$$

$$= 2b(ax_1 + bx_2)$$

2 Linear Algebra [8 pts]

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

$$\boldsymbol{X} = \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} \qquad \boldsymbol{y} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \qquad \boldsymbol{z} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

(a) What is the inner product $\boldsymbol{y}^T\boldsymbol{z}$?

$$\begin{pmatrix} 1 & 3 \end{pmatrix} \times \begin{pmatrix} 2 \\ 3 \end{pmatrix} = 11$$

(b) What is the product $\boldsymbol{X}\boldsymbol{y}$?

$$\begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} \times \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 * 1 + 4 * 3 \\ 1 * 1 + 3 * 3 \end{pmatrix} = \begin{pmatrix} 14 \\ 10 \end{pmatrix}$$

(c) Is $\boldsymbol{X}$ invertible? If so, give the inverse; if not, explain why not.

if det(X) == 0, it's not invertible

$$det(X) = (2 * 3) - (1 * 4) = 2 \implies$$

$$X^{-1} = \frac{Adj X}{det(X)} = \frac{\begin{pmatrix} 3 & -4 \\ -1 & 2 \end{pmatrix}}{2} = \begin{pmatrix} 1.5 & -2 \\ -0.5 & 1 \end{pmatrix}$$

(d) What is the rank of $\boldsymbol{X}$?

The rank is 2, all columns are linearly independent

3 Probability and Statistics [10 pts]

Consider a sample of data S obtained by flipping a coin X, where 0 denotes the coin turned up heads, and 1 denotes that it turned up tails: $S = \{1, 1, 0, 1, 0\}$.

(a) What is the sample mean for this data?

$$\overline{x} = 3/5 = 0.6$$

(b) What is the sample variance ?

$$s^2 = \frac{\sum(x_i - \overline{x})^2}{n-1} \implies \frac{(1-0.6)^2 + (1-0.6)^2 + ...}{4} = 0.3$$

(c) What is the probability of observing this data assuming that a coin with an equal probability of heads and tails was used? (*i.e.*, The probability distribution of X is $P(X = 1) = 0.5$, $P(X = 0) = 0.5$.)

$$P = 0.5^5 = 0.03125$$

(d) Note the probability of this data sample would be greater if the value of $P(X = 1)$ was not 0.5 but some other value. What is the value that maximizes the probability of sample S? [Optional: Can you prove your answer is correct?]

P(X=1) = 0.6 should maximize the probability of sample S

(e) Given the following joint distribution between X and Y, what is $P(X = T|Y = b)$?

$P(X,Y)$		Y: a	b	c
X	T	0.2	0.1	0.2
	F	0.05	0.15	0.3

$P(X = T|Y = b) = 0.1$

4 Probability axioms [5 pts]

Let A and B be two discrete random variables. In general, are the following true or false? (Here A^c denotes complement of the event A.)

(a) $P(A \cup B) = P(A \cap (B \cap A^c)) = \boldsymbol{false}$

(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \boldsymbol{true}$

(c) $P(A) = P(A \cap B) + P(A^c \cap B) = \boldsymbol{false}$

(d) $P(A|B) = P(B|A) = \boldsymbol{false}$

(e) $P(A_1 \cap A_2 \cap A_3) = P(A_3|(A_2 \cap A_1))P(A_2|A_1)P(A_1) = \boldsymbol{true}$

5 Discrete and Continuous Distributions[5 pts]

(a) Match the distribution name to its formula.

(a) Gaussian	(i) $p^x(1-p)^{1-x}$
(b) Exponential	(ii) $\frac{1}{b-a}$ when $a \leq x \leq b; 0$ otherwise
(c) Uniform	(iii) $\binom{n}{x} p^x(1-p)^{n-x}$
(d) Bernoulli	(iv) $\lambda e^{-\lambda x}$ when $x \geq 0; 0$ otherwise
(e) Binomial	(v) $\frac{1}{\sqrt{(2\pi)\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$

$$\boldsymbol{a = v, b = iv, c = ii, d = i, e = iii}$$

(b) Let X be a random variable distributed according to a Gaussian or normal distribution with probability density function:

$$P(x) = \frac{1}{2\sqrt{\pi}\sigma} \exp\left(-\frac{x^2}{4\sigma^2}\right)$$

What is the mean and variance of X?

$$\boldsymbol{\mu = 0}$$

$$\boldsymbol{\sigma^2 = 2\sigma^2}$$

6 Mean and Variance[5 pts]

(a) What is the mean and variance of a *Bernoulli*(p) random variable?

$$\boldsymbol{E[X] = \sum xP(x)}$$

$$\boldsymbol{E[X] = 0 * (1-p) + 1p \implies E[X] = p}$$

$$\boldsymbol{Var[X] = E[X^2] - E[X]^2}$$

$$\boldsymbol{E[X^2] = \sum x^2 P(x) = 0^2 * (1-p) + 1^2 p = p}$$

$$\boldsymbol{Var[X] = p - p^2 \implies p(1-p)}$$

(b) If the variance of a zero-mean random variable X is σ^2, what is the variance of $2X$? What about the variance of $X + 2$?

$$\boldsymbol{Var[2X] = 2^2 Var[X] = 4Var[X]}$$

$$\boldsymbol{Var[X+2] = Var[X]}$$

7 Algorithms[6 pts]

(a) **Big-O notation**

For each pair (f, g) of functions below, list which of the following are true: $f(n) = O(g(n))$, $g(n) = O(f(n))$, or both. Briefly justify your answers.

i. $f(n) = ln(n), g(n) = lg(n)$. Note that ln denotes log to the base e and lg denotes log to the base 2.

$$\boldsymbol{f(n) = O(g(n)) = True}$$

$$\boldsymbol{g(n) = O(f(n)) = False}$$

This is because ln(n) grows slower than lg(n)

8 Probability and Random Variables [5 pts]

(a) **Independence**

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

$$\boldsymbol{E[XY] = \sum\sum xy * P(X = x) * P(Y = y) \implies}$$

$$\boldsymbol{\sum x * P(X = x) * \sum y * P(Y = y) \implies E[X]E[Y]}$$

(b) **Law of Large Numbers and Central Limit Theorem**
Provide one line justifications.

i. If a fair die is rolled 6000 times, the number of times 3 shows up is close to 1000.

Law of Large Numbers

ii. If a fair coin is tossed n times and $\bar{X}$ denotes the average number of heads, then the distribution of $\bar{X}$ satisfies

CLT

$$\sqrt{n}(\bar{X} - \tfrac{1}{2}) \xrightarrow{n\to\infty} \mathcal{N}(0, \tfrac{1}{4})$$

9 Eigendecomposition [10 pts]

(a) Give the definition of the eigenvalues and the eigenvectors of a square matrix.

An eigenvector of a matrix is a non-zero vector such that $Av = \lambda v$. They are vectors that only change by a scalar value (the eigenvalue) after a linear transformation is applied to them

(b) Find the eigenvalues and eigenvectors of

$$\boldsymbol{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

$$det(\lambda I - A) = 0$$

$$\lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \implies det(\begin{pmatrix} \lambda - 2 & -1 \\ -1 & \lambda - 2 \end{pmatrix}$$

$$det = (\lambda - 2)^2 - 1 = \lambda^2 - 4\lambda + 3 \implies \lambda = 1, 3$$

$$\vec{v} : (A - \lambda I)\vec{X} = \vec{0} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - 1 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

after row-reducing, this becomes

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies x = -y; \vec{v}(\lambda = 1) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

same process for lambda=3 results in vector [1,1]

(c) For any positive integer k, show that the eigenvalues of $\boldsymbol{A}^k$ are $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$, the k^{th} powers of the eigenvalues of matrix $\boldsymbol{A}$, and that each eigenvector of $\boldsymbol{A}$ is still an eigenvector of $\boldsymbol{A}^k$.

let's multiply both sides by A to simulate A^2

$$Av = \lambda v \rightarrow A(Av) = A(\lambda v)$$

$$= A^2 v = \lambda(Av)$$

plugging in λv for Av results in

$$A^2 v = \lambda^2 v$$

by induction, we could keep showing that $A^k v = \lambda^k v$, resulting in same vector and values

10 Vector and Matrix Calculus [10 pts]

Consider the vectors $\boldsymbol{x}$ and $\boldsymbol{a}$ and the symmetric matrix $\boldsymbol{A}$.

(a) What is the first derivative of $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ with respect to $\boldsymbol{x}$? What is the second derivative?

the first derivative would be 2Ax and the second derivative would be 2A