Homework #0

CSE 446: Machine Learning
Ananya Shreya Soni
Date: **Wednesday** April 9, 2025 11:59pm

Probability and Statistics

A1. *[2 points]* (From Murphy Exercise 2.4.) After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested positive for a serious disease, and that the test is 99% accurate (i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative given that you don't have the disease). The good news is that this is a rare disease, striking only one in 10,000 people. What are the chances that you actually have the disease?

A1

- Let P be the event of testing positive.

- Let D be the event of having the disease.

Given:
$$\mathbb{P}(P \mid D) = 0.99, \quad \mathbb{P}(\neg P \mid \neg D) = 0.99, \quad \mathbb{P}(D) = \frac{1}{10,000}$$

$$\mathbb{P}(P \mid \neg D) = 1 - \mathbb{P}(\neg P \mid \neg D) = 0.01$$

We want to find: $\mathbb{P}(D \mid P)$ By Bayes' Theorem:

$$\mathbb{P}(D \mid P) = \frac{\mathbb{P}(P \mid D) \cdot \mathbb{P}(D)}{\mathbb{P}(P)}$$

Using the Law of Total Probability:

$$\mathbb{P}(P) = \mathbb{P}(P \mid D) \cdot \mathbb{P}(D) + \mathbb{P}(P \mid \neg D) \cdot \mathbb{P}(\neg D)$$

$$\mathbb{P}(D \mid P) = \frac{0.99 \cdot \frac{1}{10,000}}{0.99 \cdot \frac{1}{10,000} + 0.01 \cdot \left(1 - \frac{1}{10,000}\right)}$$

$$\approx 0.0098$$

Therefore, the probability that you actually have the disease given a positive test result is approximately 0.98%.

A2. For any two random variables X, Y the *covariance* is defined as $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$. You may assume X and Y take on a discrete values if you find that is easier to work with.

  a. *[1 point]* If $\mathbb{E}[Y \mid X = x] = x$ show that $\text{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$.

  b. *[1 point]* If X, Y are independent show that $\text{Cov}(X, Y) = 0$.

A2.a

Lemma 1: $\mathbb{E}[Y|X=x] = \sum_{y\in\Omega_Y} y\cdot \mathbb{P}_{Y|X}(y|x) = x$ (by given and definition of conditional expectation)

Lemma 2: $\mathbb{E}[X]=\mathbb{E}[Y]$:

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{x\in\Omega_X} \mathbb{E}[Y|X=x]\cdot \mathbb{P}_X(x) \quad \text{(by Law of Total Expectation)} \\
&= \sum_{x\in\Omega_X} x\cdot \mathbb{P}_X(x) \quad \text{(by Lemma 1)} \\
&= \mathbb{E}[X]
\end{aligned}$$

Lemma 3: $\mathbb{E}[XY]=\mathbb{E}[X^2]$

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{x,y\in\Omega_{X,Y}} xy\cdot \mathbb{P}_{X,Y}(x,y) \quad \text{(by definintion of expectation)} \\
&= \sum_{x,y\in\Omega_{X,Y}} xy\cdot \mathbb{P}_{Y|X}(y|x)\cdot \mathbb{P}_X(x) \\
&= \sum_{x\in\Omega_X} x^2\cdot \mathbb{P}_X(x) \quad \text{(by Lemma 1)} \\
&= \mathbb{E}[X^2] \quad \text{(by definintion of expectation)}
\end{aligned}$$

Lemma 4: $\mathbb{E}[Y\mathbb{E}[X]]=\mathbb{E}[X\mathbb{E}[X]]$

$$\begin{aligned}
\mathbb{E}[Y\mathbb{E}[X]] &= \sum_{x\in\Omega_X} \mathbb{E}[Y|X=x]\cdot \mathbb{P}_X(x)\cdot \mathbb{E}[X] \quad \text{(by Law of Total Expectation)} \\
&= \sum_{x\in\Omega_X} x\cdot \mathbb{E}[X]\cdot_X(x) \quad \text{(by Lemma 1)} \\
&= \mathbb{E}[X\mathbb{E}[X]] \quad \text{(by definition of expectation)}
\end{aligned}$$

$$\begin{aligned}
\mathbf{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] \quad \text{(by definintion of covariance)} \\
&= \mathbb{E}[XY - Y\mathbb{E}[X] - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]] \\
&= \mathbb{E}[XY] - \mathbb{E}[Y\mathbb{E}[X]] - \mathbb{E}[X\mathbb{E}[Y]] + \mathbb{E}[X]\mathbb{E}[Y] \quad \text{(by Linearity of Expectation)} \\
&= \mathbb{E}[X^2] - \mathbb{E}[X\mathbb{E}[X]] - \mathbb{E}[X\mathbb{E}[X]] + \mathbb{E}[X]\mathbb{E}[X] \quad \text{(by Lemmas 2, 3, and 4)} \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X\mathbb{E}[X]] + \mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] \quad \text{(by Linearity of Expectation)} \\
&= \mathbb{E}[(X-\mathbb{E}[X])^2]
\end{aligned}$$

A2.b

If X and Y are independent, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\cdot \mathbb{E}[Y] \quad \text{(by Independence)}$$

$$\begin{aligned}
\text{Cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\cdot \mathbb{E}[Y] \quad \text{(by definintion of covariance)} \\
&= \mathbb{E}[X]\cdot \mathbb{E}[Y] - \mathbb{E}[X]\cdot \mathbb{E}[Y] \\
&= 0
\end{aligned}$$

A3. Let X and Y be independent random variables with PDFs given by f and g, respectively. Let h be the PDF of the random variable $Z = X + Y$.

a. *[1 point]* Show that $h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)\,dx$. (If you are more comfortable with discrete probabilities, you can instead derive an analogous expression for the discrete case, and then you should give a one sentence explanation as to why your expression is analogous to the continuous case.)

b. *[Optional]* If X and Y are both independent and uniformly distributed on $[0,1]$ (i.e. $f(x) = g(x) = 1$ for $x \in [0,1]$ and 0 otherwise) what is h, the PDF of $Z = X + Y$?

A3

Determine the PDF of Z:
Let $F_Z(z) = \mathbb{P}(Z \leq z)$ be the CDF of $Z = X + Y$. Then:

$F_Z(z) = \mathbb{P}(X + Y \leq z)$ (by definintion of Z)

$\quad = \int_{x \in \Omega_X} \mathbb{P}(x + Y \leq z \mid X = x) f_X(x)\,dx$ (by the Law of Total Probability and given $X = x$)

$\quad = \int_{x \in \Omega_X} \mathbb{P}(Y \leq z - x \mid X = x) f_X(x)\,dx$

$\quad = \int_{x \in \Omega_X} \mathbb{P}(Y \leq z - x) f_X(x)\,dx$ (since X and Y are Independent)

$\quad = \int_{x \in \Omega_X} F_Y(z - x) f_X(x)\,dx$ (by definition of CDF of Y)

Differentiate the CDF $F_Z(z)$ with respect to z to get the PDF of Z:

$$h(z) = f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{x \in \Omega_X} f_X(x) f_Y(z - x)\,dx$$

Therefore the following holds:

$$h(z) = \int_{-\infty}^{\infty} f(x) g(z - x)\,dx$$

A4. Let $X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2)$ be i.i.d random variables. Compute the following:

a. *[1 point]* $a \in \mathbb{R}, b \in \mathbb{R}$ such that $aX_1 + b \sim \mathcal{N}(0, 1)$.

b. *[1 point]* $\mathbb{E}[X_1 + 2X_2], \mathrm{Var}[X_1 + 2X_2]$.

c. *[2 points]* Setting $\widehat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, the mean and variance of $\sqrt{n}(\widehat{\mu}_n - \mu)$.

A4.a

We are given $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ and $aX_1 + b \sim \mathcal{N}(0, 1)$
Solve for a and b:

$\mathbb{E}[aX_1 + b] = a\mathbb{E}[X_1] + b$ (by Linearity of Expectation)
$= a\mu + b = 0$

$\mathrm{Var}(aX_1 + b) = a^2\mathrm{Var}(X_1)$
$= a^2\sigma^2 = 1$

So we have: $a\mu + b = 0$ and $a^2\sigma^2 = 1$

This gives us:

$$a^2\sigma^2 = 1 \quad \Rightarrow \quad a = \frac{1}{\sigma}$$

$$a\mu + b = 0 \quad \Rightarrow \quad b = -a\mu \quad \Rightarrow \quad b = -\frac{\mu}{\sigma}$$

Therefore: $a = \frac{1}{\sigma}, b = -\frac{\mu}{\sigma}$

A4.b

$$\mathbb{E}[X_1 + 2X_2] = \mathbb{E}[X_1 + 2X_2]$$
$$= \mathbb{E}[X_1] + 2\mathbb{E}[X_2] \quad \text{(by Linearity of Expectation)}$$
$$= \mu + 2\mu$$
$$= 3\mu$$

$$\mathrm{Var}(X_1 + 2X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(2X_2) \quad \text{(since } X_1 \text{ and } X_2 \text{ are independent)}$$
$$= \mathrm{Var}(X_1) + 4\mathrm{Var}(X_2)$$
$$= \sigma^2 + 4\sigma^2$$
$$= 5\sigma^2$$

A4.c

Find $\mathbb{E}[\sqrt{n}(\hat{\mu}_n - \mu)]$ and $\mathrm{Var}[\sqrt{n}(\hat{\mu}_n - \mu)]$ and we are given $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$\mathbb{E}[\sqrt{n}(\hat{\mu}_n - \mu)] = \mathbb{E}\left[\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)\right] = \sqrt{n}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right]$$

$$= \sqrt{n}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right] = \sqrt{n}\left(\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] - \mu\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mu - \mu\right) = \sqrt{n}(\mu - \mu)$$

$$= \sqrt{n}(0) = 0$$

$$\mathrm{Var}[\sqrt{n}(\hat{\mu}_n - \mu)] = \mathrm{Var}\left(\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)\right) = n\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) = n\left(\frac{1}{n}\right)^2\sum_{i=1}^{n}\mathrm{Var}(X_i)$$

$$= \frac{1}{n}\cdot n\sigma^2 = \sigma^2$$

Linear Algebra and Vector Calculus

A5. Let $A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. For each matrix A and B:

a. *[2 points]* What is its rank?

b. *[2 points]* What is a (minimal size) basis for its column span?

Find EF of A:

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & -1 & 1 \end{bmatrix} \quad R_2 = R_2 - R_1 \quad R_3 = R_3 - R_1$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & -2 & 2 \end{bmatrix} \quad R_3 = 2R_3$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & 0 & 0 \end{bmatrix} \quad R_3 = R_3 - R_2$$

basis for column span $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}$

rank = 2 since 2 non-zero rows in EF of A

Find EF of B:

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -2 \\ 0 & -1 & -1 \end{bmatrix} \quad R_2 = R_2 - R_1 \quad R_3 = R_3 - R_1$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -2 \\ 0 & -2 & -2 \end{bmatrix} \quad R_3 = 2R_3$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -2 \\ 0 & 0 & 0 \end{bmatrix} \quad R_3 = R_3 - R_2$$

basis for column span $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}$

rank = 2 since 2 non-zero rows in EF of B

a) Rank of A and B = 2

b) Minimal basis for column span of A and B = $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}$

A6. Let $A = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix}$, $b = \begin{bmatrix} -2 & -2 & -4 \end{bmatrix}^\top$, and $c = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\top$.

   a. *[1 point]* What is Ac?

   b. *[2 points]* What is the solution to the linear system $Ax = b$?

A6.a

$$A\vec{c} = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}$$

A6.b

$$\begin{bmatrix} 0 & 2 & 4 & | & -2 \\ 2 & 4 & 2 & | & -2 \\ 3 & 3 & 1 & | & -4 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 2 & | & -1 \\ 1 & 2 & 1 & | & -1 \\ 1 & 1 & \frac{1}{3} & | & -\frac{4}{3} \end{bmatrix} \quad R_1 = \tfrac{1}{2}R_1 \quad R_2 = \tfrac{1}{2}R_2 \quad R_3 = \tfrac{1}{3}R_3$$

$$\begin{bmatrix} 1 & 1 & \frac{1}{3} & | & -\frac{4}{3} \\ 1 & 2 & 1 & | & -1 \\ 0 & 1 & 2 & | & -1 \end{bmatrix} \quad \text{swap } R_1 \text{ with } R_3$$

$$\begin{bmatrix} 1 & 1 & \frac{1}{3} & | & -\frac{4}{3} \\ 0 & 1 & \frac{2}{3} & | & \frac{1}{3} \\ 0 & 1 & 2 & | & -1 \end{bmatrix} \quad R_2 = R_2 - R_1$$

$$\begin{bmatrix} 1 & 1 & \frac{1}{3} & | & -\frac{4}{3} \\ 0 & 1 & \frac{2}{3} & | & \frac{1}{3} \\ 0 & 0 & \frac{4}{3} & | & -\frac{4}{3} \end{bmatrix} \quad R_3 = R_3 - R_2$$

$$\begin{bmatrix} 1 & 1 & \frac{1}{3} & | & -\frac{4}{3} \\ 0 & 1 & 0 & | & 1 \\ 0 & 0 & \frac{4}{3} & | & -\frac{4}{3} \end{bmatrix} \quad R_2 = -\tfrac{1}{2}R_3 + R_2$$

$$\begin{bmatrix} 1 & 1 & 0 & | & -1 \\ 0 & 1 & 0 & | & 1 \\ 0 & 0 & \frac{4}{3} & | & -\frac{4}{3} \end{bmatrix} \quad R_1 = -\tfrac{1}{4}R_3 + R_1$$

$$\begin{bmatrix} 1 & 0 & 0 & | & -2 \\ 0 & 1 & 0 & | & 1 \\ 0 & 0 & 1 & | & -1 \end{bmatrix} \quad R_1 = -R_2 + R_1 \quad R_3 = \tfrac{3}{4}R_3$$

Final Answer:
$x_1 = -2$
$x_2 = 1$
$x_3 = -1$

A7. For possibly non-symmetric $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ and $c \in \mathbb{R}$, let $f(x, y) = x^\top \mathbf{A} x + y^\top \mathbf{B} x + c$. Define

$$\nabla_z f(x, y) = \begin{bmatrix} \frac{\partial f}{\partial z_1}(x, y) & \frac{\partial f}{\partial z_2}(x, y) & \cdots & \frac{\partial f}{\partial z_n}(x, y) \end{bmatrix}^\top \in \mathbb{R}^n .$$

Note: If you are unfamiliar with gradients, you may find the resources available on the course website useful. Section 4 of Zico Kolter and Chuong Do's Linear Algebra Review and Reference may be particularly helpful.

a. *[2 points]* Explicitly write out the function $f(x, y)$ in terms of the components $A_{i,j}$ and $B_{i,j}$ using appropriate summations over the indices.

b. *[2 points]* What is $\nabla_x f(x, y)$ in terms of the summations over indices *and* vector notation?

c. *[2 points]* What is $\nabla_y f(x, y)$ in terms of the summations over indices *and* vector notation?

A7.a

Write $f(x, y)$ in terms of the components A_{ij} and B_{ij}. Note: A and B are both $n \times n$ matrices.

$$f(x, y) = x^T A x + y^T B x + c$$

$$= x^T \sum_{j=1}^{n} A_j x_j + y^T \sum_{j=1}^{n} B_j x_j + c$$

$$= \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} A_{ij} x_j \right) + \sum_{i=1}^{n} y_i \left(\sum_{j=1}^{n} B_{ij} x_j \right) + c$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} B_{ij} y_i x_j + c$$

A7.b

Find $\nabla_x f(x, y)$
Summation notation: For all k, $\nabla_x f(x, y)_k =$

$$\frac{\partial}{\partial x_k} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} B_{ij} y_i x_j + c \right)$$

$$= \frac{\partial}{\partial x_k} \left(\sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{j \neq k} A_{kj} x_k x_j + \sum_{i \neq k} A_{ik} x_i x_k + A_{kk} x_k^2 + \sum_{i=1}^{n} \sum_{j \neq k} B_{ij} y_i x_j + \sum_{i=1}^{n} B_{ik} y_i x_k + c \right)$$

$$= \sum_{j \neq k} A_{kj} x_j + \sum_{i \neq k} A_{ik} x_i + 2 A_{kk} x_k + \sum_{i=1}^{n} B_{ik} y_i$$

$$= \sum_{j=1}^{n} A_{kj} x_j + \sum_{i=1}^{n} A_{ik} x_i + \sum_{i=1}^{n} B_{ik} y_i$$

Vector Notation:

$$\frac{\partial}{\partial x} (x^T A x + y^T B x + c) = (A + A^T) x + B^T y$$

A7.c

Find $\nabla_y f(x, y)$

Summation notation: For all k, $\nabla_y (f(x, y))_k =$

$$\frac{\partial}{\partial y_k} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} B_{ij} y_i x_j + c \right)$$

$$= \frac{\partial}{\partial y_k} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j + \sum_{i \neq k} \sum_{j=1}^{n} B_{ij} y_i x_j + \sum_{j=1}^{n} B_{kj} y_k x_j + c \right)$$

$$= \sum_{j=1}^{n} B_{kj} x_j$$

Vector Notation:

$$\frac{\partial}{\partial y} (x^T A x + y^T B x + c) = Bx$$

A8. Show the following:

a. *[2 points]* Let $g \colon \mathbb{R} \to \mathbb{R}$ and $v, w \in \mathbb{R}^n$ such that $g(v_i) = w_i$ for $i \in [n]$. Find an expression for g such that $\mathrm{diag}(v)^{-1} = \mathrm{diag}(w)$.

b. *[2 points]* Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be orthonormal and $x \in \mathbb{R}^n$. An orthonormal matrix is a square matrix whose columns and rows are orthonormal vectors, such that $\mathbf{A}\mathbf{A}^\top = \mathbf{A}^\top \mathbf{A} = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. Show that $||\mathbf{A}x||_2^2 = ||x||_2^2$.

c. *[2 points]* Let $\mathbf{B} \in \mathbb{R}^{n \times n}$ be invertible and symmetric. A symmetric matrix is a square matrix satisfying $\mathbf{B} = \mathbf{B}^\top$. Show that $\mathbf{B}^{-1}$ is also symmetric.

d. *[2 points]* Let $\mathbf{C} \in \mathbb{R}^{n \times n}$ be positive semi-definite (PSD). A positive semi-definite matrix is a symmetric matrix satisfying $x^\top \mathbf{C} x \geq 0$ for any vector $x \in \mathbb{R}^n$. Show that its eigenvalues are non-negative.

A8.a

From $\text{diag}(v)^{-1} = \text{diag}(w)$, we know that the diagonal elements must satisfy $\frac{1}{v_i} = w_i$ for all $i \in [n]$ because:

$$\text{diag}(v) = \begin{bmatrix} v_1 & 0 & 0 & \cdots & 0 \\ 0 & v_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & v_n \end{bmatrix}$$

$$\text{diag}(w) = \begin{bmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w_n \end{bmatrix}$$

$$\text{diag}(v)^{-1} = \begin{bmatrix} \frac{1}{v_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{v_2} & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{v_n} \end{bmatrix}$$

By the property of diagonal matrix inversion, the inverse of a diagonal matrix is obtained by replacing each non-zero diagonal element with its reciprocal. Thus, $[\text{diag}(v)^{-1}]_i = \frac{1}{v_i}$. Since $g(v_i) = w_i$ and $\text{diag}(v)^{-1} = \text{diag}(w)$, we have $w_i = \frac{1}{v_i}$. Therefore, $g(x) = \frac{1}{x}$ for $x \neq 0$ is an expression for g that satisfies the condition.

A8.b

Show that $\|\mathbf{A}x\|_2^2 = \|x\|_2^2$:

$$\begin{aligned} \|\mathbf{A}x\|_2^2 &= (\mathbf{A}x)^\top (\mathbf{A}x) \\ &= x^\top \mathbf{A}^\top \mathbf{A}x \quad \text{(property of transpose)} \\ &= x^\top \mathbf{I}x \quad \text{(since } \mathbf{A} \text{ is orthonormal, } \mathbf{A}^\top \mathbf{A} = \mathbf{I}) \\ &= x^\top x \\ &= \|x\|_2^2 \quad \text{(by definition of } l_2 \text{ norm)} \end{aligned}$$

A8.c

Show $\mathbf{B}^{-1}$ is symmetric when $\mathbf{B}$ is symmetric and invertible:

$$\begin{aligned} (\mathbf{B}^{-1})^\top &= (\mathbf{B}^\top)^{-1} \quad \text{(property of transpose and inverse)} \\ &= \mathbf{B}^{-1} \quad \text{(since } \mathbf{B} \text{ is symmetric, } \mathbf{B}^\top = \mathbf{B}) \end{aligned}$$

Therefore, $(\mathbf{B}^{-1})^\top = \mathbf{B}^{-1}$, which means $\mathbf{B}^{-1}$ is symmetric.

> **A8.d**
>
> Show the eigenvalues of a positive semi-definite matrix $\mathbf{C}$ are non-negative:
> We can represent $\mathbf{C}$ as $\mathbf{C} = \mathbf{U}\Lambda\mathbf{U}^\top$, where $\mathbf{U}$ is the matrix of eigenvectors of $\mathbf{C}$ and Λ is the diagonal matrix of eigenvalues.
> For any vector $x \in \mathbb{R}^n$, we have:
>
> $$\begin{aligned} x^\top \mathbf{C} x &= x^\top \mathbf{U}\Lambda\mathbf{U}^\top x \\ &= y^\top \Lambda y \quad (\text{let } y = \mathbf{U}^\top x) \\ &= \sum_{i=1}^n \lambda_i y_i^2 \end{aligned}$$
>
> Since $\mathbf{C}$ is positive semi-definite, we know $x^\top \mathbf{C} x \geq 0$ for all $x \in \mathbb{R}^n$. For the sum $\sum_{i=1}^n \lambda_i y_i^2$ to be non-negative, each eigenvalue λ_i must be non-negative because y_i^2 is guarenteed to be positive. Therefore, all eigenvalues of $\mathbf{C}$ are non-negative ($>= 0$).

Programming

These problems are available in a .zip file, with some starter code. All coding questions in this class will have starter code. **Before attempting these problems, you will need to set up a Conda environment that you will use for every assignment in the course. Unzip the HW0-A.zip file and read the instructions in the README file to get started.**

A9. For $\nabla_x f(x, y)$ as solved for in Problem 7:

   a. *[1 point]* Using native Python, implement `vanilla_solution` using your `vanilla_matmul` and `vanilla_transpose` functions.

   b. *[1 point]* Now implement `numpy_version` using NumPy functions.

   c. *[1 point]* Report the difference in wall-clock time for parts a-b, and discuss reasons for the observed difference.
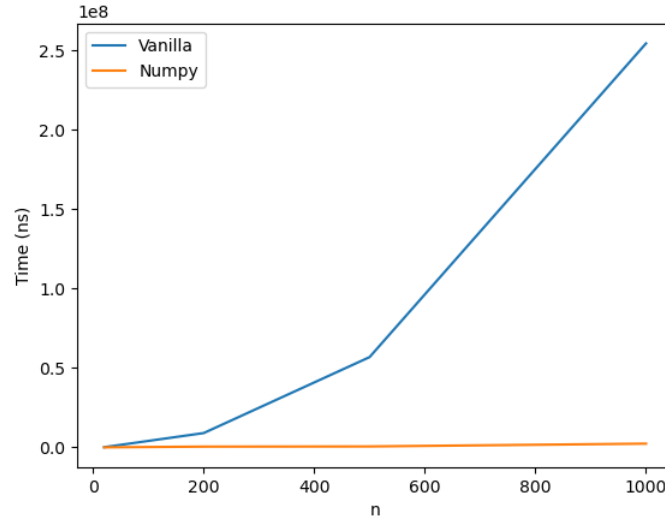
Time for vanilla implementation: 0.113ms - Time for numpy implementation: 0.018ms = 0.095ms
Time for vanilla implementation: 9.012ms - Time for numpy implementation: 0.468ms = 8.544ms
Time for vanilla implementation: 56.876ms - Time for numpy implementation: 0.574ms = 56.302ms
Time for vanilla implementation: 254.547ms - Time for numpy implementation: 2.337ms = 252.21ms

A reason for the observed difference could be that Numpy does matrix operations like
matrix multiplications much more efficiently than the vanilla nested loop operations.
Additionally, numpy arrays can only store one data type unlike Python lists and Numpy is built with C
which allows it to do computation much more faster and efficiently than native Python.

A10. Two random variables X and Y have equal distributions if their CDFs, F_X and F_Y, respectively, are equal, i.e. for all x, $|F_X(x) - F_Y(x)| = 0$. The central limit theorem says that the sum of k independent, zero-mean, variance $1/k$ random variables converges to a (standard) Normal distribution as k tends to infinity. We will study this phenomenon empirically (you will use the Python packages NumPy and Matplotlib). Each of the following subproblems includes a description of how the plots were generated; these have been coded for you. The code is available in the .zip file. In this problem, you will add to our implementation to explore **matplotlib** library, and how the solution depends on n and k.

a. *[2 points]* For $i = 1, \ldots, n$ let $Z_i \sim \mathcal{N}(0, 1)$. Let $x \mapsto F(x)$ denote the true CDF from which each Z_i is drawn (i.e., Gaussian). Define $\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{Z_i \leq x\}$ for $x \in \mathbb{R}$ and we will choose n large enough such that, for all $x \in \mathbb{R}$,

$$\sqrt{\mathbb{E}\left[\left(\widehat{F}_n(x) - F(x)\right)^2\right]} \leq 0.0025 \ .$$

Plot $x \mapsto \widehat{F}_n(x)$ for x ranging from -3 to 3.

b. *[2 points]* Define $Y^{(k)} = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} B_i$ where each B_i is equal to -1 and 1 with equal probability and the B_i's are independent. We know that each $\frac{1}{\sqrt{k}} B_i$ is zero-mean and has variance $1/k$. For each $k \in \{1, 8, 64, 512\}$ we will generate n (same as in part a) independent copies $Y^{(k)}$ and plot their empirical CDF on the same plot as part a.

Be sure to always label your axes. Your plot should look something like the following (up to styling) (Tip: checkout `seaborn` for instantly better looking plots.)
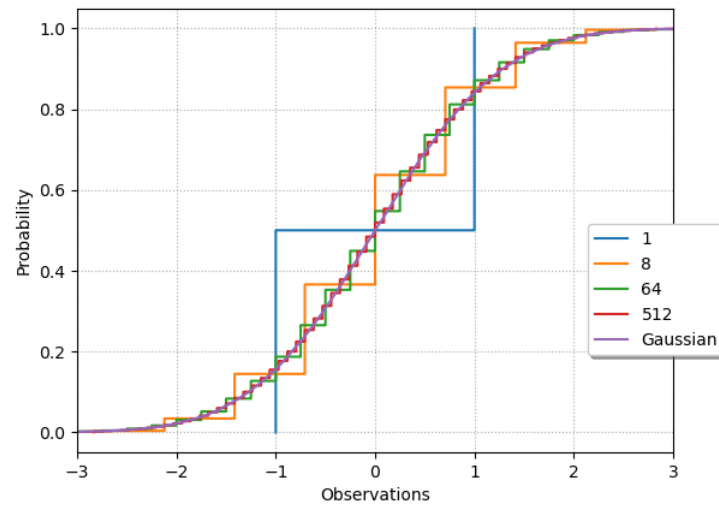
n = 40000

The empirical CDF increasingly matches the true CDF as k increases. The plot visually depicts how as k increases the empirical CDF looks more and more like the Gaussian.