Homework #0

CSE 446 / 546: Machine Learning
Professor Pang Wei Koh & Sewoong Oh
Due: **Wednesday** October 8, 2025 11:59pm
38 points

Jiayao Huang

Please review all homework guidance posted on the website before submitting to Gradescope. Reminders:

- All code must be written in Python and all written work must be typeset (e.g. LaTeX).

- Make sure to read the "What to Submit" section following each question and include all items.

- Please provide succinct answers and supporting reasoning for each question. Similarly, when discussing experimental results, concisely create tables and/or figures when appropriate to organize the experimental results. All explanations, tables, and figures for any particular part of a question must be grouped together.

- For every problem involving generating plots, please include the plots as part of your PDF submission.

- When submitting to Gradescope, please link each question from the homework in Gradescope to the location of its answer in your homework PDF. Failure to do so may result in deductions of up to 10% of the value of each question not properly linked. For instructions, see `https://www.gradescope.com/get_started#student-submission`.

- Make sure to submit your code by running `inv submit` from within the `cse446` conda environment and submitting the zip folder created. Submissions that don't follow this procedure will fail the autograder on Gradescope. After the due date, we will not accept resubmissions. **We suggest you wait for the autograder on Gradescope to complete and display your score before regarding the coding portion of assignments complete.**

Not adhering to these reminders may result in point deductions.

Important: By turning in this assignment (and all that follow), you acknowledge that you have read and understood the collaboration policy with humans and AI assistants alike: `https://courses.cs.washington.edu/courses/cse446/25au/assignments/`. Any questions about the policy should be raised at least 24 hours before the assignment is due. There are no warnings or second chances. If we suspect you have violated the collaboration policy, we will report it to the college of engineering who will complete an investigation.

Quick note about macro.tex

The `macro.tex` file provided on the course website provides the definitions for different macros that are referenced in the CSE 446 homework LaTeX files. For example, it includes the new command that makes the point values for problems pink and italicized on the homework documents, and also includes commands for things like set notation and writing matrices.

While not required to use the provided homework LaTeX files, if you would like to compile them, place the homework file in a directory and change the path to the macro file on line 3 (e.g. if the macro file is in the same directory as the homework file, the path should be `\subimport*{./}{macro}`). You can also choose to directly copy the contents of the macro file into your LaTeX code.

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?

What to Submit:

- Final Answer

- Corresponding Calculations

Solve.

`Final Answer`: About 0.98%

`Calculations`:

Let D be the probability of having the disease, and T_p be the probability of testing positive. Then, D^c would be the probability of not having the disease, and T_n is testing negative.

Now, we know

$$P(D) = \frac{1}{10000} = 0.0001, \quad P(D^c) = 0.9999$$

and

$$P(T_p \mid D) = 0.99, \quad P(T_n \mid D^c) = 0.99, \quad P(T_p \mid D^c) = 1 - P(T_n \mid D^c) = 0.01.$$

By Bayes' Formula, we have

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

Then, by law of total probability:

$$P(T_p) = P(T_p \mid D)P(D) + P(T_p \mid D^c)P(D^c) = 0.99 \times 0.0001 + 0.01 \times 0.9999 = 0.000099 + 0.009999 = 0.010098.$$

Thus,

$$P(D \mid T_p) = \frac{0.99 \times 0.0001}{0.010098} = \frac{0.000099}{0.010098} \approx 0.00980392 \approx 0.98\%.$$

which means the chances (probability) of actually having the disease given that you are tested positive is about 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 $\text{Cov}(X, Y) = 0$.

What to Submit:

- **Parts a-b:** Proofs

Proof. `Part a.`
Given that $\mathbb{E}[Y \mid X = x] = x$, we know $\mathbb{E}[Y \mid X] = X$, then use the law of iterated expectations,

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[X]$$

Note that

$$Y - \mathbb{E}[Y] = Y - \mathbb{E}[X] = (Y - X) + (X - \mathbb{E}[X])$$

Thus we have

$$(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = (X - \mathbb{E}[X])(Y - X) + (X - \mathbb{E}[X])^2$$

Taking expectations on both sides, we get

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - X)] + \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Use the law of iterated expectations again, we have

$$\mathbb{E}[(X - \mathbb{E}[X])(Y - X)] = \mathbb{E}[\mathbb{E}[(X - \mathbb{E}[X])(Y - X) \mid X]]$$

Conditional on X, $X - \mathbb{E}[X]$ is constant, so we take it out

$$\mathbb{E}[(X - \mathbb{E}[X])(Y - X) \mid X] = (X - \mathbb{E}[X])\mathbb{E}[Y - X \mid X] = (X - \mathbb{E}[X])(\mathbb{E}[Y \mid X] - X) = (X - \mathbb{E}[X])(X - X) = 0.$$

Thus

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - X)] + \mathbb{E}[(X - \mathbb{E}[X])^2]$$
$$= \mathbb{E}[0] + \mathbb{E}[(X - \mathbb{E}[X])^2]$$
$$= \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Therefore

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

`Part b.` Since X and Y are independent, we have

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Thus

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY - X\mathbb{E}[Y] - \mathbb{E}[X]Y + \mathbb{E}[X]\mathbb{E}[Y]]$$
$$= \mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$
$$= \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] = 0$$

Therefore, if X, Y are independent,
$$\text{Cov}(X, Y) = 0$$

A3. *[2 points]* Let $X \sim \mathcal{N}(0, \sigma^2)$ and $t \in \mathbb{R}$. Show that $\mathbb{E}\left[\exp\left(Xt - \frac{t^2\sigma^2}{2}\right)\right] = 1$.

What to Submit:

- Proof

Proof. Let $X \sim \mathcal{N}(0, \sigma^2)$, so its PDF is

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

We compute

$$\mathbb{E}\left[exp\left(Xt - \frac{\sigma^2 t^2}{2}\right)\right] = \int_{-\infty}^{\infty} exp\left(xt - \frac{\sigma^2 t^2}{2}\right) \cdot \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{x^2}{2\sigma^2}\right) dx.$$

Combine these exponents

$$xt - \frac{\sigma^2 t^2}{2} - \frac{x^2}{2\sigma^2} = -\frac{x^2}{2\sigma^2} + xt - \frac{\sigma^2 t^2}{2} = -\frac{1}{2\sigma^2}\left[x^2 - 2\sigma^2 tx\right] - \frac{\sigma^2 t^2}{2}.$$

$$= -\frac{(x - \sigma^2 t)^2}{2\sigma^2} + \frac{\sigma^2 t^2}{2} - \frac{\sigma^2 t^2}{2} = -\frac{(x - \sigma^2 t)^2}{2\sigma^2}.$$

Then

$$\mathbb{E}\left[exp\left(Xt - \frac{\sigma^2 t^2}{2}\right)\right] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \sigma^2 t)^2}{2\sigma^2}\right) dx$$

Note that the right hand side of this equation is the PDF of $\mathcal{N}(\sigma^2 t, \sigma^2)$. Thus the integral is

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \sigma^2 t)^2}{2\sigma^2}\right) dx = 1$$

Therefore

$$\mathbb{E}\left[exp\left(Xt - \frac{\sigma^2 t^2}{2}\right)\right] = 1$$

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]* $\mathbb{E}[X_1 + 2X_2], \text{Var}[X_1 + 2X_2]$.

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

What to Submit:

- **Part a:** a, b, and the corresponding calculations

- **Part b:** $\mathbb{E}[X_1 + 2X_2]$, $\text{Var}[X_1 + 2X_2]$

- **Part c:** $\mathbb{E}[\sqrt{n}(\widehat{\mu}_n - \mu)]$, $\text{Var}[\sqrt{n}(\widehat{\mu}_n - \mu)]$

- **Parts a-c** Corresponding calculations

Solve.
Part a.
$$a = \frac{1}{\sigma}, \quad b = -\frac{\mu}{\sigma} \quad \text{or} \quad a = -\frac{1}{\sigma}, \quad b = \frac{\mu}{\sigma}$$

Calculations:
$$\mathbb{E}[aX_1 + b] = a\mathbb{E}[X_1] + b = a\mu + b = 0 \quad \Rightarrow \quad b = -a\mu$$

$$\text{Var}(aX_1 + b) = a^2\text{Var}[X_1] = a^2\sigma^2 = 1 \quad \Rightarrow \quad a = \pm\frac{1}{\sigma} \quad \Rightarrow \quad b = \mp\frac{\mu}{\sigma}$$

Part b.
$$\mathbb{E}[X_1 + 2X_2] = 3\mu, \quad \text{Var}[X_1 + 2X_2] = 5\sigma^2$$

Calculations: Since $X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2)$ are i.i.d random variables, we have

$$\mathbb{E}[X_1 + 2X_2] = \mathbb{E}[X_1] + 2\mathbb{E}[X_2] = \mu + 2\mu = 3\mu$$

$$\text{Var}[X_1 + 2X_2] = \text{Var}[X_1] + 2^2\text{Var}[X_2] = \sigma^2 + 4\sigma^2 = 5\sigma^2$$

Part c.
$$\mathbb{E}[\sqrt{n}(\hat{\mu}_n - \mu)] = 0, \quad \text{Var}[\sqrt{n}(\hat{\mu}_n - \mu)] = \sigma^2$$

Calculations: Since $\widehat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, and $X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2)$ are i.i.d random variables, we have

$$\mathbb{E}[\hat{\mu}_n] = \frac{1}{n}\mathbb{E}[X_1 + X_2 + ... + X_n] = \frac{1}{n} \cdot n\mu = \mu$$

$$\text{Var}[\hat{\mu}_n] = \frac{1}{n^2}\text{Var}[X_1 + X_2 + ... + X_n] = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

Thus
$$\mathbb{E}[\hat{\mu}_n - \mu] = \mathbb{E}[\hat{\mu}_n] - \mu = 0 \quad \Rightarrow \quad \mathbb{E}[\sqrt{n}(\hat{\mu}_n - \mu)] = \sqrt{n} \cdot \mathbb{E}[\hat{\mu}_n - \mu] = \sqrt{n} \cdot 0 = 0$$

and

$$\text{Var}[\hat{\mu}_n - \mu] = \text{Var}[\hat{\mu}_n] - \text{Var}[\mu] = \frac{\sigma^2}{n} \quad \Rightarrow \quad \text{Var}[\sqrt{n}(\hat{\mu}_n - \mu)] = \text{Var}[\sqrt{n}]\text{Var}[\hat{\mu}_n - \mu] = n \cdot \frac{\sigma^2}{n} = \sigma^2$$

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?

What to Submit:

- **Parts a-b:** Solution and corresponding calculations

Solve.
Part a.
$$\text{rank}(A) = 2, \quad \text{rank}(B) = 2$$

Calculations: For matrix A,

$$\begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix} \xrightarrow[r_3 - r_1]{r_2 - r_1} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & -1 & 1 \end{bmatrix} \xrightarrow{r_3 - \frac{1}{2}r_2} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{-\frac{1}{2}r_2} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

Row reduction yields 2 pivot columns (since $c_3 = 3c_1 - c_2$) $\Rightarrow$ rank = 2.

For matrix B,

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \xrightarrow[r_3 - r_1]{r_2 - r_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -2 \\ 0 & -1 & -1 \end{bmatrix} \xrightarrow{r_3 - \frac{1}{2}r_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -2 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{-\frac{1}{2}r_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Row reduction yields 2 pivot columns (since $c_3 = c_1 + c_2$) $\Rightarrow$ rank = 2.

Part b.
$$\text{Basis for } A : \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}, \quad \text{Basis for } B : \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Calculations: For both matrices, since rank = 2, the column space has dimension 2. The pivot columns (columns 1 and 2) from the original matrices form a basis for the column space.
Thus we get the columns as:
Columns for A are $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$
Columns for B are $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$

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 $Ax = b$?

What to Submit:

- **Parts a-b:** Solution and corresponding calculations

Solve.
Part a.

$$Ac = \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}$$

Calculations:

$$Ac = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \cdot 1 + 2 \cdot 1 + 4 \cdot 1 \\ 2 \cdot 1 + 4 \cdot 1 + 2 \cdot 1 \\ 3 \cdot 1 + 3 \cdot 1 + 1 \cdot 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}$$

Part b.

$$x = \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$$

Calculations:

$$[A \mid b] = \left[\begin{array}{ccc|c} 0 & 2 & 4 & -2 \\ 2 & 4 & 2 & -2 \\ 3 & 3 & 1 & -4 \end{array} \right] \xrightarrow{\text{R1} \leftrightarrow \text{R2}} \left[\begin{array}{ccc|c} 2 & 4 & 2 & -2 \\ 0 & 2 & 4 & -2 \\ 3 & 3 & 1 & -4 \end{array} \right] \xrightarrow{\text{R1} \leftarrow \text{R1/2}} \left[\begin{array}{ccc|c} 1 & 2 & 1 & -1 \\ 0 & 2 & 4 & -2 \\ 3 & 3 & 1 & -4 \end{array} \right]$$

$$\xrightarrow{\text{R3} \leftarrow \text{R3 - 3R1}} \left[\begin{array}{ccc|c} 1 & 2 & 1 & -1 \\ 0 & 2 & 4 & -2 \\ 0 & -3 & -2 & -1 \end{array} \right] \xrightarrow{\text{R2} \leftarrow \text{R2/2}} \left[\begin{array}{ccc|c} 1 & 2 & 1 & -1 \\ 0 & 1 & 2 & -1 \\ 0 & -3 & -2 & -1 \end{array} \right]$$

$$\xrightarrow{\text{R3} \leftarrow \text{R3 + 3R2}} \left[\begin{array}{ccc|c} 1 & 2 & 1 & -1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 4 & -4 \end{array} \right] \xrightarrow{\text{R3} \leftarrow \text{R3/4}} \left[\begin{array}{ccc|c} 1 & 2 & 1 & -1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & -1 \end{array} \right]$$

Substitute back, we have

$$x_3 = -1$$
$$x_2 + 2 \times (-1) = -1 \quad \Rightarrow \quad x_2 = 1$$
$$x_1 + 2 \times 1 + (-1) = -1 \quad \Rightarrow \quad x_1 = -2$$

Thus

$$x = \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$$

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

$$\nabla_z f(x,y) = \left[\frac{\partial f}{\partial z_1}(x,y) \quad \frac{\partial f}{\partial z_2}(x,y) \quad \cdots \quad \frac{\partial f}{\partial z_n}(x,y) \right]^\top \in \mathbb{R}^n .$$

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?

What to Submit:

- **Part a:** Explicit formula for $f(x,y)$

- **Parts b-c:** Summation form and corresponding calculations. Summation form includes writing out what each component of the resultant vector is, where each component is expressed as a summation. Intermediate components may be indicated by ellipses, like in the equation given in the problem description.

- **Parts b-c:** Vector form and corresponding calculations. Vector form includes writing the final answer only in terms of products, sums (or differences), and/or transposes of the input matrices and vectors.

Solve.
Part a.

$$f(x,y) = \sum_{i=1}^n \sum_{j=1}^n x_i A_{i,j} x_j + \sum_{i=1}^n \sum_{j=1}^n y_i B_{i,j} x_j + c$$

Part b.

$$\nabla_x f(x,y) = (A + A^T)x + B^T y = \begin{bmatrix} \sum_{j=1}^n A_{1j} x_j + \sum_{i=1}^n A_{i1} x_i + \sum_{i=1}^n y_i B_{i1} \\ \vdots \\ \sum_{j=1}^n A_{nj} x_j + \sum_{i=1}^n A_{in} x_i + \sum_{i=1}^n y_i B_{in} \end{bmatrix}$$

Calculations: We take the derivative of f with respect to x and get

$$\frac{\partial f}{\partial x_k} = \frac{\partial}{\partial x_k} \left(\sum_{i,j} x_i A_{ij} x_j \right) + \frac{\partial}{\partial x_k} \left(\sum_{i,j} y_i B_{ij} x_j \right)$$

$$= \sum_j A_{kj} x_j + \sum_i A_{ik} x_i + \sum_i y_i B_{ik}$$

Then the summation form becomes

$$\nabla_x f(x,y) = \begin{bmatrix} \sum_{j=1}^n A_{1j} x_j + \sum_{i=1}^n A_{i1} x_i + \sum_{i=1}^n y_i B_{i1} \\ \vdots \\ \sum_{j=1}^n A_{nj} x_j + \sum_{i=1}^n A_{in} x_i + \sum_{i=1}^n y_i B_{in} \end{bmatrix}$$

Vector form:

$$\sum_j A_{kj} x_j = (Ax)_k \quad \sum_i A_{ik} x_i = (A^T x)_k \quad \sum_i y_i B_{ik} = (B^T y)_k$$

Thus

$$\nabla_x f(x,y) = Ax + A^T x + B^T y = (A + A^T)x + B^T y$$

Part c.

$$\nabla_y f(x,y) = Bx = \begin{bmatrix} \sum_{j=1}^{n} B_{1j} x_j \\ \vdots \\ \sum_{j=1}^{n} B_{nj} x_j \end{bmatrix}$$

Calculations: Since only the second term depends on y, we take the derivative of it with respect to y and get

$$\frac{\partial f}{\partial y_k} = \frac{\partial}{\partial y_k} \left(\sum_{i,j} y_i B_{ij} x_j \right) = \sum_{j=1}^{n} B_{kj} x_j$$

Thus the summation form becomes

$$\nabla_y f(x,y) = \begin{bmatrix} \sum_{j=1}^{n} B_{1j} x_j \\ \vdots \\ \sum_{j=1}^{n} B_{nj} x_j \end{bmatrix}$$

Since

$$\sum_{j} B_{kj} x_j = (Bx)_k$$

we have the vector form as

$$\nabla_y f(x,y) = Bx$$

A8. Show the following:

a. *[2 points]* Let $g\colon \mathbb{R} \setminus \{0\} \to \mathbb{R}$ 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.

What to Submit:

- **Part a:** Explicit formula for g

- **Parts a-d:** Proof

Solve.
Part a.

$$g(t) = \frac{1}{t}$$

Proof: We know that for a diagonal matrix,

$$\mathrm{diag}(v)^{-1} = \mathrm{diag}(1/v_1, \ldots, 1/v_n), \quad v_i \neq 0, \text{ for } i \in [n]$$

Now we need

$$\mathrm{diag}(w) = \mathrm{diag}(v)^{-1} = \mathrm{diag}(1/v_1, \ldots, 1/v_n)$$

which means

$$w_i = 1/v_i \quad \text{for all } i \in [n] \quad \Leftrightarrow \quad v_i = 1/w_i \quad \text{for all } i \in [n]$$

Given $g(v_i) = w_i$, we have the expressions for g should be

$$g(t) = 1/t$$

Part b.
Proof: Since A is orthonormal, $A^T A = I$. Then:

$$\|Ax\|_2^2 = (Ax)^T (Ax) = x^T A^T A x = x^T I x = x^T x = \|x\|_2^2$$

Part c.
Proof: Given $B^T = B$, we take inverse on both sides

$$(B^T)^{-1} = B^{-1}$$

Note that

$$(B^T)^{-1} = (B^{-1})^T$$

we have

$$(B^{-1})^T = B^{-1}$$

Thus, B^{-1} is symmetric.

```
Part d.
```
Proof: Let λ be eigenvalue with eigenvector $v \neq 0$: $Cv = \lambda v$. Then

$$v^T C v = v^T (\lambda v) = \lambda v^T v = \lambda \|v\|_2^2$$

Since C is positive semi-definite, and $\|v\|_2^2 > 0$. Thus

$$v^T C v \geq 0 \quad \Rightarrow \quad \lambda \|v\|_2^2 \geq 0 \quad \Rightarrow \quad \lambda \geq 0$$

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.

What to Submit:

- **Part c:** Plot that shows the difference in wall-clock time for parts a-b
- **Part c:** Explanation for the difference (1-2 sentences)
- **Code** on Gradescope through coding submission

Solve.
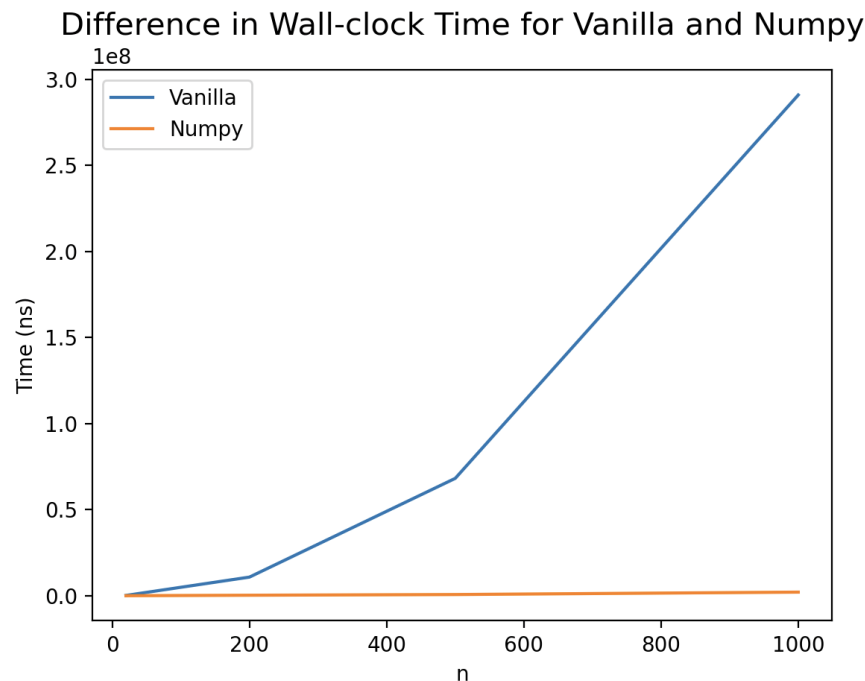Plot showing the difference:

Figure 1: Difference in Wall-clock Time for Vanilla and Numpy

Explanations:
NumPy shows near-constant time complexity because it may use optimized code underneath with parallel processing. Vanilla python exhibits exponential-like growth as it uses slow loops for matrix operations. As the matrix size grows, the Vanilla version gets much slower compared to NumPy.

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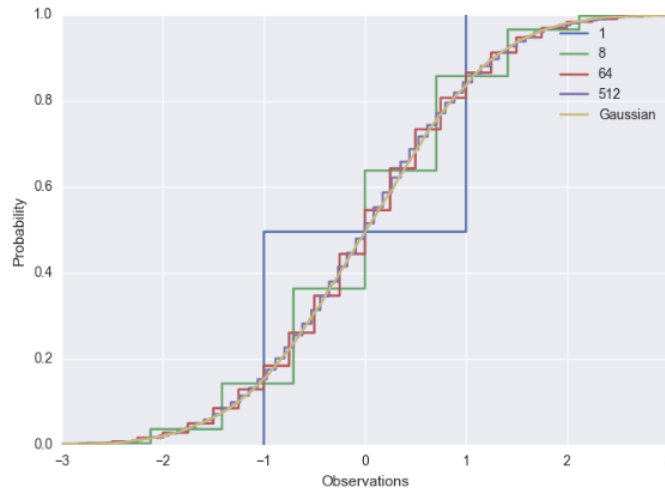
a. *[2 points]* For $i = 1, \ldots, n$ let $Z_i \sim \mathcal{N}(0, 1)$. Let $x \mapsto F(x)$ denote the true CDF from which each Z_i is drawn (i.e., Gaussian). Define $\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{Z_i \le x\}$ for $x \in \mathbb{R}$ and we will choose n large enough such that,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

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

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.)

What to Submit:

- **Part a:** Value for n (You can simply print the value determined by the code provided for you). **Part b:** In 1-2 sentences: How does the empirical CDF change with k?

- **Parts a and b:** Plot of $x \mapsto \widehat{F}_n(x)$ for $x \in [-3, 3]$

- **Code** on Gradescope through coding submission

13

> **Solve.**
> `Part a.` $n = 40000$
> `Part b.` As k increases, the empirical CDF becomes much smoother and converges to the Gaussian CDF. It can be inferred that as $k \to \infty$, the empirical CDF would converge to a standard Normal distribution, which is what the central limit theorem says.
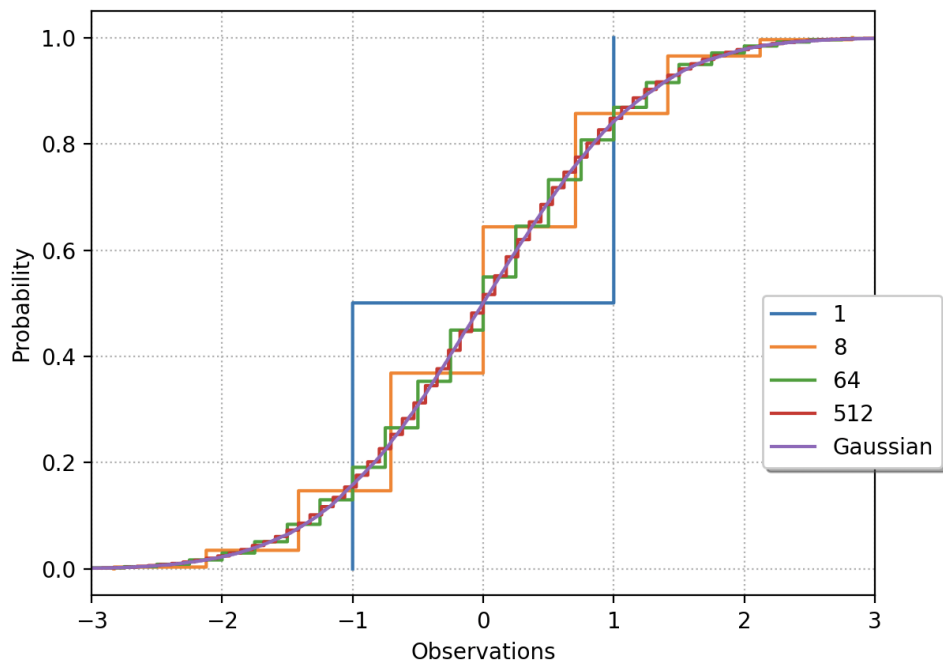> `Plot showing the center limit theorem with CDFs:`

Figure 2: The Center Limit Theorem with Empirical/Gaussian CDFs