Homework #0

CSE 446: Machine Learning
Professors Matt Golub & Pang Wei Koh
Due: **Wednesday** October 02, 2024 11:59pm
38 points

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`.

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/cse446/24au/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

Solution:
Let A be the event you test positive for the disease and B the event you actually have the disease.

$\Pr[A] = \Pr[A \cap B] + \Pr[A \cap B^c] = (0.0001)(0.99) + (0.9999)(0.01)$

$\Pr[B|A] = \frac{\Pr[B \cap A]}{\Pr[A]} = \frac{(0.0001)(0.99)}{(0.0001)(0.99)+(0.9999)(0.01)}$ or 0.9804%

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

Solution:
Using the law of total probability, we know that $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$.
If $\mathbb{E}[Y \mid X = x] = x$, we can then deduce that $\mathbb{E}[Y \mid X] = X$.
Substituting it in, we get $\mathbb{E}[Y] = \mathbb{E}[X]$.

Now, we can solve...
$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])|X]]$ by law of total probability
$= \mathbb{E}[(X - \mathbb{E}[X])\,\mathbb{E}[(Y - \mathbb{E}[Y])|X]]$ since $(X - \mathbb{E}[X])$ is a constant given X
$= \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[Y])]$ since $\mathbb{E}[(Y - \mathbb{E}[Y])|X] = \mathbb{E}[Y|X] - \mathbb{E}[\mathbb{E}[Y]|X] = X - \mathbb{E}[Y]$ $(\mathbb{E}[Y|X] = X)$
$= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$ since we deduced $\mathbb{E}[Y] = \mathbb{E}[X]$.

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

Solution:
Since X, Y are independent, ...
$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[(X - \mathbb{E}[X])]\,\mathbb{E}[(Y - \mathbb{E}[Y])]$
$= 0$ since intuitively $\mathbb{E}[(X - \mathbb{E}[X])]$ and $\mathbb{E}[(Y - \mathbb{E}[Y])]$ both equal 0

What to Submit:

- **Parts a-b:** Proofs

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

Solution:
CDF of Z: $H(z) = P(Z \leq z) = P(X + Y \leq z) = \int\int_{(x+y\leq z)} f(x)g(y)dydx$
$= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x)g(y)dydx$
$= \int_{-\infty}^{\infty} f(x)G(z-x)dx$

$h(z) = \frac{d}{dz}[H(Z)] = \frac{d}{dz}[\int_{-\infty}^{\infty} f(x)G(z-x)dx] = \int_{-\infty}^{\infty} f(x)g(z-x)dx$

b. *[1 point]* 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 $Z = X + Y$?

Solution:
We'll refer to $h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx$ from part a.
To start, we know $h(z) = 0$ for $z > 2$ and $z < 0$ since x, y $\in [0, 1]$ (0+0 = 0, 1+1 = 2).
Since $g(z-x)$ is nonzero for $0 \leq z - x \leq 1$, we find that $z - 1 \leq x \leq z$.
Also, since $f(x)$ is nonzero for $0 \leq x \leq 1$, we can simplify $h(z)$ to $\int_{-\infty}^{\infty} f(x)g(z-x)dx = \int_0^1 g(z-x)dx$.

Now, to get h, we to find the overlap between $0 \leq x \leq 1$ and $z - 1 \leq x \leq z$, which we can do by cases:

1) $0 \leq z \leq 1$
In this case, $z - 1 \leq x \leq z$ gives us $-1 \leq x \leq z$. Since $0 \leq x \leq 1$, the bounds of the integral becomes $0 \leq x \leq z$. Thus, ...
$h(z) = \int_0^z g(z-x)dx = \int_0^z (1)dx = z$

2) $1 < z \leq 2$
In this case, since $z - 1 > 1$, we simply integrate from $z - 1$ to 1 (the upper bound of x). Thus, ...
$h(z) = \int_{z-1}^1 g(z-x)dx = \int_{z-1}^1 (1)dx = 1 - (z-1) = 2 - z$
Altogether, we have ...

$$h(z) = \begin{cases} z & \text{for } 0 \leq z \leq 1 \\ 2 - z & \text{for } 1 < z \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

What to Submit:

- **Part a:** Proof
- **Part b:** Formula for PDF Z and corresponding calculations

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 $aX_1 + b \sim \mathcal{N}(0, 1)$.

> Solution:
> Mean of $aX_1 + b = a\mu + b$
> Variance of $aX_1 + b = a^2\sigma^2$
> $a^2\sigma^2 = 1 \to a^2 = \frac{1}{\sigma^2} \to a = \frac{1}{\sigma}$
> $a\mu + b = 0 \to (\frac{\mu}{\sigma}) + b = 0 \to b = \frac{\mu}{\sigma}$

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

> Solution:
> $\mathbb{E}[X_1 + 2X_2] = \mathbb{E}[X_1] + 2\,\mathbb{E}[X_2]$ by LOE
> $= \mu + 2(\mu) = 3\mu$
> $\mathrm{Var}[X_1 + 2X_2] = \mathrm{Var}[X_1] + \mathrm{Var}[2X_2]$ since X_1 and X_2 are independent
> $= \mathrm{Var}[X_1] + (2)^2\,\mathrm{Var}[X_2] = \sigma^2 + 4\sigma^2 = 5\sigma^2$

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

> Solution:
> Mean of $\sqrt{n}(\widehat{\mu}_n - \mu)$:
> $\mathbb{E}[\widehat{\mu}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n}(n\mu) = \mu$
> $\mathbb{E}[\sqrt{n}(\widehat{\mu}_n - \mu)] = \sqrt{n}(\mu - \mu) = 0$
>
> Variance of $\sqrt{n}(\widehat{\mu}_n - \mu)$:
> $\mathrm{Var}[\widehat{\mu}_n] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = (\frac{1}{n})^2\sum_{i=1}^n \mathrm{Var}[X_i] = (\frac{1}{n})^2(n\sigma^2) = \frac{\sigma^2}{n}$
> $\mathrm{Var}[\sqrt{n}(\widehat{\mu}_n - \mu)] = \mathrm{Var}[\sqrt{n}\widehat{\mu}_n]$ since μ is a constant
> $= n\,\mathrm{Var}[\widehat{\mu}_n] = n(\frac{\sigma^2}{n}) = \sigma^2$

What to Submit:

- **Part a:** a, b, and the corresponding calculations
- **Part b:** $\mathbb{E}[X_1 + 2X_2]$, $\mathrm{Var}[X_1 + 2X_2]$
- **Part c:** $\mathbb{E}[\sqrt{n}(\widehat{\mu}_n - \mu)]$, $\mathrm{Var}[\sqrt{n}(\widehat{\mu}_n - \mu)]$
- **Parts a-c** Corresponding calculations

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?

Solution:

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

Subtract a_1 from a_2 and a_3: $a_2 = a_2 - a_1$ and $a_3 = a_3 - a_1$

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

Subtract $\frac{1}{2}a_2$ from a_3: $a_3 = a_3 - \frac{1}{2}a_2$

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

Since there are two non-zero rows, the rank of A is 2.

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

Subtract b_1 from b_2 and b_3: $b_2 = b_2 - b_1$ and $b_3 = b_3 - b_1$

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

Subtract $\frac{1}{2}b_2$ from b_3: $b_3 = b_3 - \frac{1}{2}b_2$

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

Since there are two non-zero rows, the rank of B is 2.

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

Solution:

Referring to the row-echelon form matrices we found in part a, there are two pivot columns for both of the two matrices: the first and the second column. Thus, the minimal size basis for the column span of matrix A is ...

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

And, the minimal size basis for the column span of matrix B is also ...

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

What to Submit:

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

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?

Solution:
$$Ac = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$
$0 * 1 + 2 * 1 + 4 * 1 = 6$
$2 * 1 + 4 * 1 + 2 * 1 = 8$
$3 * 1 + 3 * 1 + 1 * 1 = 7$
Thus, ..
$$Ac = \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}$$

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

Solution:
$Ax = b$
1: $2x_2 + 4x_3 = -2 \rightarrow x_2 + 2x_3 = -1$
2: $2x_1 + 4x_2 + 2x_3 = -2 \rightarrow x_1 + 2x_2 + x_3 = -1$
3: $3x_1 + 3x_2 + x_3 = -4$

$3(x_1 + 2x_2 + x_3 = -1)$
- $3x_1 + 3x_2 + x_3 = -4$
$-3x_2 + 2x_3 = 1$

$3(x_2 + 2x_3 = -1)$
- $-3x_2 + 2x_3 = 1$
$4x_3 = -4 \rightarrow x_3 = -1$

$x_2 + 2(-1) = -1$
$x_2 = 1$

$x_1 + 2(1) + (-1) = -1$
$x_1 = -2$
$$x = \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$$

What to Submit:

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

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

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

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.

Solution:
$f(x,y) = x^\top \mathbf{A} x + y^\top \mathbf{B} x + c$
$= \begin{bmatrix} x_1, ... x_n \end{bmatrix} \begin{bmatrix} A_{11}, ... A_{1n} \\ ... \\ A_{n1}, ... A_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ ... \\ x_n \end{bmatrix} + \begin{bmatrix} y_1, ... y_n \end{bmatrix} \begin{bmatrix} B_{11}, ... B_{1n} \\ ... \\ B_{n1}, ... B_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ ... \\ x_n \end{bmatrix}$ + c
$= \sum_{j=1}^n \sum_{i=1}^n A_{ij} x_i x_j + \sum_{j=1}^n \sum_{i=1}^n B_{ij} y_i x_j + c$

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

Solution:
$\nabla_x f(x,y) = \frac{\partial}{\partial x_k}(\sum_{j=1}^n \sum_{i=1}^n A_{ij} x_i x_j + \sum_{j=1}^n \sum_{i=1}^n B_{ij} y_i x_j + c)$
$= \frac{\partial}{\partial x_k}((x_1 \sum_{i=1}^n A_{i1} x_i + ... x_n \sum_{i=1}^n A_{in} x_i) + (x_1 \sum_{i=1}^n B_{i1} y_i + ... x_n \sum_{i=1}^n B_{in} y_i))$
$= (\sum_{j=1}^n A_{kj} x_j + \sum_{i=1}^n A_{ik} x_i) + (\sum_{i=1}^n B_{ik} y_i)$
$= (\sum_{j=1}^n A_{kj} x_j + A_{jk} x_j + B_{jk} y_j)$ - Summation
$= (\sum_{j=1}^n A_{kj} + A_{jk}) x + B^\top y$
$= (\mathbf{A} + \mathbf{A}^\top) x + \mathbf{B}^\top y$ - Vector Notation

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

Solution:
$\nabla_y f(x,y) = \frac{\partial}{\partial y_k}(\sum_{j=1}^n \sum_{i=1}^n a_{ij} x_i x_j + \sum_{j=1}^n \sum_{i=1}^n b_{ij} y_i x_j + c)$
$= \frac{\partial}{\partial y_k}(\sum_{j=1}^n \sum_{i=1}^n b_{ij} y_i x_j))$
$= (\sum_{j=1}^n (b_{kj} x_j))$ - Summation
$= \mathbf{B} x$ - 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.

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 $\operatorname{diag}(v)^{-1} = \operatorname{diag}(w)$.

Solution:
$\text{diag}(v) = \begin{bmatrix} v_1, ...0 \\ 0, v_2, ... \\ 0, ...v_n \end{bmatrix}$
$\text{diag}(v) = \begin{bmatrix} \frac{1}{v_1}, ...0 \\ 0, \frac{1}{v_2}, ... \\ 0, ... \frac{1}{v_n} \end{bmatrix}$
$\text{diag}(w) = \begin{bmatrix} w_1, ...0 \\ 0, w_2, ... \\ 0, ...w_n \end{bmatrix}$
$w_i = \frac{1}{v_i}$
$g(v_i) = \frac{1}{v_i}$ for $i \in [n]$

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 $||\mathbf{A}x||_2^2 = ||x||_2^2$.

Solution:
$||\mathbf{A}x||_2^2 = (\mathbf{A}x)^\top(\mathbf{A}x) = x^\top\mathbf{A}^\top\mathbf{A}x = x^\top\mathbf{I}x = x^\top x = ||x||_2^2$.

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.

Solution:
$\mathbf{B}\mathbf{B}^{-1} = \mathbf{I} \to (\mathbf{B}\mathbf{B}^{-1})^\top = (\mathbf{I})^\top \to (\mathbf{B}^{-1})^\top\mathbf{B}^\top = \mathbf{I} \to (\mathbf{B}^{-1})^\top\mathbf{B} = \mathbf{I} \to (\mathbf{B}^{-1})^\top = \mathbf{I}\mathbf{B}^{-1}$
$\to (\mathbf{B}^{-1})^\top = \mathbf{B}^{-1}$

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.

Solution:
Suppose there exists some eigenvalue λ of $\mathbf{C}$. Then, there will be an eigenvector x such that $\mathbf{C}x = \lambda x$. The rest follows as ...
$x^\top\mathbf{C}x = \lambda x^\top x \geq 0$. Since $x^\top x$ will always be positive, λ must be non-negative.

What to Submit:

- **Part a:** Explicit formula for g
- **Parts a-d:** Proof

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.**

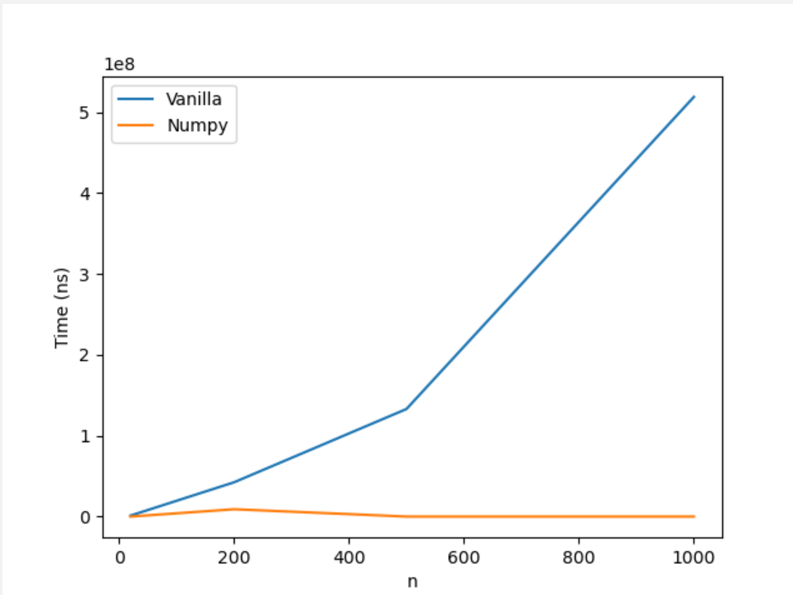
A9. For $\nabla_x f(x, y)$ as solved for in Problem 7:

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.

Solution:
Time for vanilla implementation: 1.047ms vs Time for numpy implementation: 0.231ms
Time for vanilla implementation: 36.6748ms vs Time for numpy implementation: 4.3226ms
Time for vanilla implementation: 221.3069ms vs Time for numpy implementation: 0.9481ms
Time for vanilla implementation: 890.9276ms vs Time for numpy implementation: 2.0036ms



Explanation:
Based on what I knew and read, the NumPy function is faster than native Python-built function due to three main reasons. The first is that NumPy arrays are stored in contiguous blocks of memory, which improves cache utilization and allows efficient access to elements. The second reason is that NumPy's core operations are implemented in low-level languanges such as C and Fortran, which makes them significantly faster native Python implementations. The final one is that NumPy supports vectorized operations, which allows operations on entire arrays without loops.

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

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 $x \in \mathbb{R}$,

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

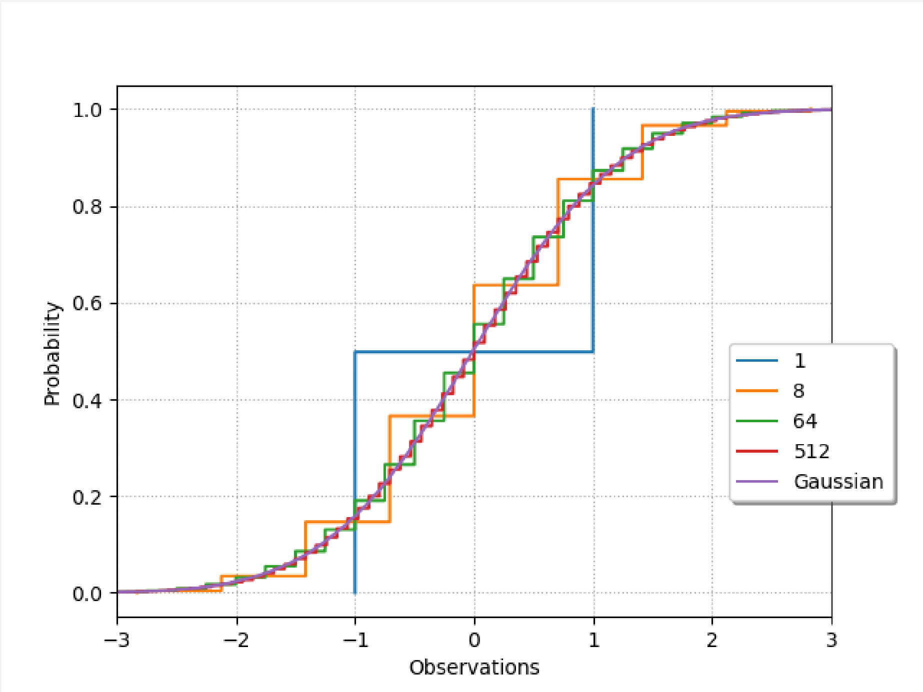
Plot $x \mapsto \widehat{F}_n(x)$ for x ranging from -3 to 3.

> Solution:
> $n = 40000$

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.

> Solution:
> The empirical CDF approaches the Gaussian Distribution as k becomes larger.

a. and b.)
Plot:



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