Homework #0

CSE 446/546: Machine Learning
Professors Matt Golub and Hunter Schafer
Due: **Wednesday** January 10, 2024 11:59pm
38 points

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

- 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`.
- For every problem involving code, please include all code you have written for the problem as part of your PDF submission *in addition to* submitting your code to the separate assignment on Gradescope created for code. Not submitting all code files will lead to a deduction of up to 10% of the value of each question missing code.

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/cse446/24wi/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 446/546 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.

If you would like to compile the provided homework files, place the file in a directory and change the path to the 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: As the Bayes theorem said,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}.$$

Thus, we have

$P(gettingdisease|testpositive)$

$$= \frac{P(\text{test positive}|\text{getting disease})P(\text{getting disease})}{P(\text{test positive}|\text{getting disease})P(\text{getting disease}) + P(\text{test positive}|\text{not getting disease})P(\text{not getting disease})}$$

$$= \frac{0.99 * 0.0001}{0.99 * 0.0001 + 0.01 * 0.9999}$$

$= 0.00980392156.$

Final Answer: The chances that you actually have the disease is 0.0098.

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

Solution:
Part a

$$\begin{aligned}\text{Cov}(X,Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - 2\,\mathbb{E}[X]\,\mathbb{E}[Y] + \mathbb{E}[X]EY = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] \\ &= \sum_{x,y} xyP(X = x, Y = y) - \left(\sum_x xP(X = x)\right)\left(\sum_y yP(Y = y)\right).\end{aligned}$$

We have $\mathbb{E}[Y \mid X = x] = x$, that is, $\sum_y yP(Y = y|X = x) = x$, and thus, based on the law of probability,

$$\begin{aligned}
\sum_y yP(Y = y) &= \sum_x \sum_y yP(Y = y|X = x)P(X = x) \\
&= \sum_x xP(X = x) \\
&= \mathbb{E}[X] \\
\sum_{x,y} xyP(X = x, Y = y) &= \sum_x x \sum_y yP(Y = y|X = x)P(X = x) \\
&= \sum_x x^2 P(X = x) \\
&= \mathbb{E}\left[X^2\right]
\end{aligned}$$

Combine the equations together, we have

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \sum_{x,y} xyP(X = x, Y = y) - \left(\sum_x xP(X = x)\right)\left(\sum_y yP(Y = y)\right) \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]\,\mathbb{E}[X] \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2
\end{aligned}$$

Given that

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

we have

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

Part b When X, Y are independent, we have $P(X, Y) = P(X)P(Y)$, and we know $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$. Thus,

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \sum_{x,y} xyP(X = x, Y = y) - \left(\sum_x xP(X = x)\right)\left(\sum_y yP(Y = y)\right) \\
&= \sum_{x,y} xyP(X = x)P(Y = y) - \left(\sum_x xP(X = x)\right)\left(\sum_y yP(Y = y)\right) \\
&= \left(\sum_x xP(X = x)\right)\left(\sum_y yP(Y = y)\right) - \left(\sum_x xP(X = x)\right)\left(\sum_y yP(Y = y)\right) \\
&= 0.
\end{aligned}$$

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

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 $Z = X + Y$?

What to Submit:

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

Solution:
Part a

$$
\begin{aligned}
H(z) &= \int\int_{x+y\leq z} f(x)g(y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x)g(y)dydx \\
h(z) &= \frac{\mathrm{d}}{\mathrm{d}z}H(Z) \\
&= \int_{-\infty}^{\infty} f(x)g(z-x)dx
\end{aligned}
$$

Part b We know $f(x) = 1$ and $g(y) = 1$ for $x \in [0,1]$ and $y \in [0,1]$. Thus, we just need to figure out the integration region for part b. For $0 \leq z \leq 1$,

$$h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx = \int_0^z 1dx = z.$$

For $1 < z \leq 2$,

$$h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx = \int_{z-1}^1 1dx = 2-z.$$

Thus, in the end, we have

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

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 $\sqrt{n}(\widehat{\mu}_n - \mu)$.

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

Solution:
Part a Since $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, we have $aX_1 + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$, meaning that

$$\begin{cases} a\mu + b = 0 \\ a^2\sigma^2 = 1 \end{cases}$$

Thus, $a = \frac{1}{\sigma}$ and $b = -\frac{\mu}{\sigma}$. Or, $a = -\frac{1}{\sigma}$ and $b = \frac{\mu}{\sigma}$.

Part b Since $X_1, X_2 \sim \mathcal{N}(\mu, \sigma^2)$, we have $2X_2 \sim \mathcal{N}(2\mu, 2^2\sigma^2)$. And thus, $X_1 + 2X_2 \sim \mathcal{N}(\mu + 2\mu, \sigma^2 + 2^2\sigma^2)$.

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

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

Part c Since $X_1, X_2, ..., X_n \sim \mathcal{N}(\mu, \sigma^2)$,we have $\sum_{i=1}^n X_i \sim \mathcal{N}(n\mu, n\sigma^2)$ and $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$. Thus, $\sqrt{n}(\widehat{\mu}_n - \mu) \sim \mathcal{N}(0, \sigma^2)$. And finally, we have

$$\mathbb{E}\big[\sqrt{n}(\widehat{\mu}_n - \mu)\big] = 0$$

$$\mathrm{Var}\big[\sqrt{n}(\widehat{\mu}_n - \mu)\big] = \sigma^2$$

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

Solution:
Part a

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

As shown above, we got the RREF of the matrix A, and by that, the rank of A is 2.

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

As shown above, we got the RREF of the matrix B, and by that, the rank of B is 2.

Part b Since the rank of A is 2, and the first two columns are independent, thus, they can be a basis of the column span. So as the matrix B. And thus, the basis for both A and B's column span is

$$\begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$

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

Solution:
Part a

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

Part b

$$\left[\begin{array}{ccc|c} 0 & 2 & 4 & -2 \\ 2 & 4 & 2 & -2 \\ 3 & 3 & 1 & -4 \end{array}\right] \to \left[\begin{array}{ccc|c} 0 & 1 & 2 & -1 \\ 1 & 2 & 1 & -1 \\ 3 & 3 & 1 & -4 \end{array}\right] \to \left[\begin{array}{ccc|c} 0 & 1 & 2 & -1 \\ 1 & 2 & 1 & -1 \\ 0 & -3 & -2 & -1 \end{array}\right] \to \left[\begin{array}{ccc|c} 0 & 1 & 2 & -1 \\ 1 & 2 & 1 & -1 \\ 0 & -2 & 0 & -2 \end{array}\right] \to \left[\begin{array}{ccc|c} 0 & 1 & 2 & -1 \\ 1 & 0 & 1 & -3 \\ 0 & -1 & 0 & -1 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|c} 0 & 0 & 1 & -1 \\ 1 & 0 & 1 & -3 \\ 0 & -1 & 0 & -1 \end{array}\right] \to \left[\begin{array}{ccc|c} 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & -2 \\ 0 & -1 & 0 & -1 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

It is clear that $x_1 = -2$, $x_2 = 1$, and $x_3 = -1$. 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

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

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?

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.

Solution:
Part a

$$\begin{aligned}
f(x, y) &= x^\top \mathbf{A} x + y^\top \mathbf{B} x + c \\
&= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + c \\
&= \begin{bmatrix} \sum_{i=1}^n a_{i1} x_i & \sum_{i=1}^n a_{i2} x_i & \cdots & \sum_{i=1}^n a_{in} x_i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^n b_{i1} y_i & \sum_{i=1}^n b_{i2} y_i & \cdots & \sum_{i=1}^n b_{in} y_i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + c \\
&= \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n b_{ij} y_i x_j + c
\end{aligned}$$

Part b

$$\nabla_x f(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x, y) & \frac{\partial f}{\partial x_2}(x, y) & \cdots & \frac{\partial f}{\partial x_n}(x, y) \end{bmatrix}$$

$$\begin{aligned}\frac{\partial f}{\partial x_1}(x,y) &= 2a_{11}x_1 + \sum_{i=2}^{n} a_{1i}x_i + \sum_{i=2}^{n} a_{i1}x_i + \sum_{i=1}^{n} b_{i1}y_i \\ &= \sum_{i=1}^{n} a_{1i}x_i + \sum_{i=1}^{n} a_{i1}x_i + \sum_{i=1}^{n} b_{i1}y_i\end{aligned}$$

Thus, we can get

$$\begin{aligned}\frac{\partial f}{\partial x_k}(x,y) &= \sum_{i=1}^{n} a_{ki}x_i + \sum_{i=1}^{n} a_{ik}x_i + \sum_{i=1}^{n} b_{ik}y_i \\ \nabla_x f(x,y) &= \begin{bmatrix} \sum_{i=1}^{n} a_{1i}x_i + \sum_{i=1}^{n} a_{i1}x_i + \sum_{i=1}^{n} b_{i1}y_i \\ \vdots \\ \sum_{i=1}^{n} a_{ki}x_i + \sum_{i=1}^{n} a_{ik}x_i + \sum_{i=1}^{n} b_{ik}y_i \\ \vdots \end{bmatrix} \\ &= Ax + A^\top x + B^\top y\end{aligned}$$

Part c

$$\begin{aligned}\nabla_y f(x,y) &= \begin{bmatrix} \frac{\partial f}{\partial y_1}(x,y) \\ \frac{\partial f}{\partial y_2}(x,y) \\ \vdots \\ \frac{\partial f}{\partial y_n}(x,y) \end{bmatrix} \\ &= \begin{bmatrix} \sum_{j=1}^{n} b_{1j}x_j \\ \sum_{j=1}^{n} b_{2j}x_j \\ \vdots \\ \sum_{j=1}^{n} b_{nj}x_j \end{bmatrix} \\ &= Bx\end{aligned}$$

A8. Show the following:

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

b. *[2 points]* Let $\mathbf{A} \in R^{n\times n}$ be orthonormal and $x \in \mathbb{R}^n$. An orthonormal matrix is a square martix 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 $\mathbf{B} \in R^{n\times n}$ 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 $\mathbf{C} \in R^{n\times n}$ 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

Solution:
Part a
Let denote the diag(v) as $\mathbf{A}$, and diag($g(v)$) as $\mathbf{B}$. When we have $\mathbf{A}^{-1} = \mathbf{B}$, then, $\mathbf{AB} = \mathbf{AA}^{-1} = \mathbf{I}$, meaning that for each element i of $\mathbf{AB}$, $v_i g(v_i) = 1$, and thus

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

Part b
$||\mathbf{A}x||_2^2 = (\mathbf{A}x)^\top \mathbf{A}x = x^\top \mathbf{A}^\top \mathbf{A}x$. Given that $\mathbf{A}^\top \mathbf{A} = \mathbf{I}$, we have

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

Part c
Given that $\mathbf{I} = \mathbf{B}^{-1}\mathbf{B}$ and $\mathbf{B}^\top = \mathbf{B}$, we have

$$(\mathbf{B}^{-1})^\top = (\mathbf{B}^{-1})^\top \mathbf{B}\mathbf{B}^{-1} = (\mathbf{B}^{-1})^\top \mathbf{B}^\top (\mathbf{B}^\top)^{-1} = (\mathbf{B}\mathbf{B}^{-1})^\top (\mathbf{B}^\top)^{-1} = (\mathbf{B}^\top)^{-1} = \mathbf{B}^{-1}$$

Thus, we have $(\mathbf{B}^{-1})^\top = \mathbf{B}^{-1}$, and thus, $\mathbf{B}^{-1}$ is symmetric.
Part d
For eigenvalue λ and eigenvector x, we have $\mathbf{C}x = \lambda x$, meaning that $x^\top \mathbf{C}x = x^\top \lambda x$. Given that for any vector $y \in \mathbb{R}^n$, $y^\top \mathbf{C}y \geq 0$, we have that for λ and x, $x^\top \lambda x \geq 0$. Thus, $\lambda x^\top x \geq 0$. Given that $x^\top x = ||x||_2^2 \geq 0$ always hold, we must have $\lambda \geq 0$.

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 the summation form. You can pass matrices/vectors to the `vanilla_matmul` and `vanilla_transpose` functions, but the actual operations should be done element-wise.

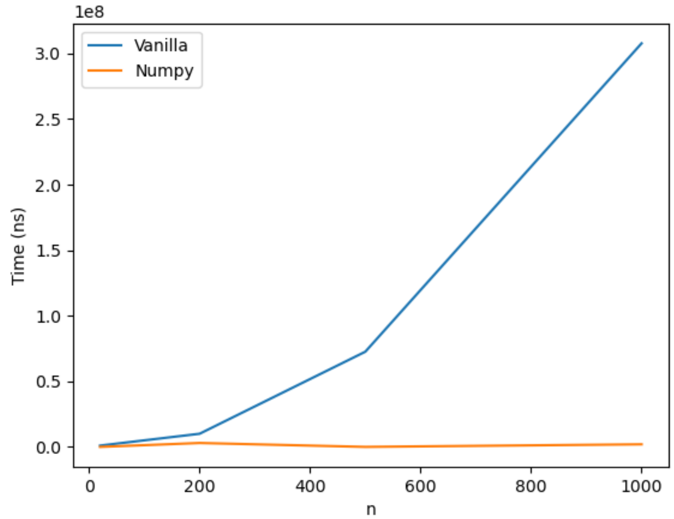
b. *[1 point]* Using NumPy, implement the vector form.

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

Solution:
Part a-b Please see the separately submitted code.
Part c Please see the figure below.
As we can see, the differences between NumPy and vanilla increase as the n increases. This is because NumPy operation adopts the vectorization, by which the "for" loop is operated in the C, and can be done in parallel.



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 $i = 1, \dots, n$ let $Z_i \sim \mathcal{N}(0, 1)$. Let $F(x)$ denote the true CDF from which each Z_i is drawn (i.e., Gaussian). Define $\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{Z_i \leq x\}$ 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 $\widehat{F}_n(x)$ 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 $\frac{1}{\sqrt{k}} B_i$ 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.)

../img/full.png

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 $\widehat{F}_n(x) \in [-3, 3]$
- **Code** on Gradescope through coding submission

Solution:
Part a The value of n is 40000, as printed by the python program.
Part b As k increases to infinity, the empirical CDF of $Y^{(k)}$, according to the central limit theorem should converge to the CDF of Z, that is, the CDF of (standard) normal distribution. The figure below does show such relation.