Homework 0

CSE 546: ML
Ray Chen
Collaborator: Kevin Shao

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 Answer:

- Corresponding Calculations
  We can set $T = 1$ as tested positive $T = 0$ as tested negative; $D = 1$ as have disease and $D = 0$ as not. So, we have the following

$$P(T = 1|D = 1) = P(T = 0|D = 0) = 0.99 \tag{1}$$
$$P(D = 1) = 0.0001 \tag{2}$$

  We need to compute the chances that actually have the disease, which is $P(D = 1|T = 1)$. We can use Bayes' rule and the Law of total probability to get the result.

$$\begin{aligned} P(D = 1|T = 1) &= \frac{P(T = 1|D = 1)P(D = 1)}{P(T = 1|D = 1)P(D = 1) + P(T = 1|D = 0)P(D = 0)} \\ &= \frac{0.99 \times 0.0001}{0.99 \times 0.0001 + 0.01 \times 0.9999} \\ &= 0.0098 \end{aligned} \tag{3}$$

A2. For any two random variables X, Y the *covariance* is defined as $\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$. 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 $\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$.

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

A2 Answer:

- **Parts a:**

  First, we simplify the $\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY - X\,\mathbb{E}[Y] - Y\,\mathbb{E}[X] + \mathbb{E}[X]\,\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[Y]\,\mathbb{E}[X] + \mathbb{E}[X]\,\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] \end{aligned} \tag{4}$$

  We apply the tower rule to $\mathbb{E}[Y|X = x] = x$ compute the value of $\mathbb{E}[Y] = \mathbb{E}[X]$.

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathbb{E}[XY|X = x] - \mathbb{E}[X]\,\mathbb{E}[Y|X = x] \\ &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \\ &= \mathbb{E}\left[(X - \mathbb{E}\left[X^2\right])^2\right] \end{aligned} \tag{5}$$

- **Parts b:**

  When X, Y are independent, it means $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$.

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] \qquad \text{(By Equation(4))} \\ &= \mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[X]\,\mathbb{E}[Y] \end{aligned} \tag{6}$$

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

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 $Z = X + Y$?

A3 Answer:

- **Part a:**

  Because X and Y are independent random variables, the PDF should just be f multiple g.

$$\mathbb{P}\{Z \leq z\} = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x)g(y)\,\mathrm{d}y\,\mathrm{d}x \tag{7}$$

  We plug Equation (7) in $h(z) = \frac{\partial}{\partial z}\mathbb{P}\{Z \leq z\}$, then we have

$$\begin{aligned} h(z) &= \frac{\partial}{\partial z}\mathbb{P}\{Z \leq z\} \\ &= \frac{\partial}{\partial z}\int_{-\infty}^{\infty} f(x)G(z-x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty}\frac{\partial}{\partial z} f(x)G(z-x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} f(x)g(z-x)\,\mathrm{d}x \end{aligned} \tag{8}$$

- **Part b:**

  For now, we set $s = z - x$, plug in Equation (8) we have

$$\begin{aligned} h(z) &= \int_{-\infty}^{\infty} f(x)g(z-x)\,\mathrm{d}x \\ &= \int_{0}^{1} f(x)g(x-z)\,\mathrm{d}x \\ &= \begin{cases} 1, z \in (-\infty, 1) \\ 2-z, z \in [0,1] \\ 0, other \end{cases} \end{aligned} \tag{9}$$

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 Answer:

- **Part a:**

  We have

$$\begin{aligned}\mathbb{E}[X] &= \mu, Var(X) = \sigma^2 \\ Y &= aX + b \\ \mathbb{E}[Y] &= 0, Var(Y) = 1\end{aligned}$$

  So, we can obtain

$$\mathbb{E}[Y] = a\,\mathbb{E}[X] + b \tag{10}$$

$$Var(Y) = (a^2)Var(X) \tag{11}$$

  From above Equation(10) and (11), we have

$$0 = a\mu + b \tag{12}$$

$$1 = a^2\sigma^2 \tag{13}$$

  Then, we have answer

$$a = \pm\frac{1}{\sigma} \tag{14}$$

$$b = \mp\frac{\mu}{\sigma} \tag{15}$$

- **Part b:**

$$\begin{aligned}\mathbb{E}[X_1 + 2X_2] &= \mathbb{E}[X_1] + \mathbb{E}[2X_2] \\ &= \mu + 2\mu \\ &= 3\mu\end{aligned} \tag{16}$$

  and

$$\begin{aligned}\mathrm{Var}[X_1 + 2X_2] &= \mathrm{Var}[X_1] + \mathrm{Var}[2X_1] \\ &= \mathrm{Var}[X_1] + 4\,\mathrm{Var}[X_1] \\ &= \sigma^2 + 4\sigma^2 \\ &= 5\sigma^2\end{aligned} \tag{17}$$

- **Part c:**

First, we need to compute the $\mathbb{E}[\hat{\mu}_n]$

$$\begin{aligned}
\mathbb{E}[\hat{\mu}_n] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mu \\
\mathbb{E}[\hat{\mu}_n] &= \mu \qquad (18)
\end{aligned}$$

Thanks to Equation(16), we can get the mean

$$\begin{aligned}
\mathbb{E}[\sqrt{n}(\hat{\mu}_n - \mu)] &= \sqrt{n}\big(\mathbb{E}[\hat{\mu}_n - \mathbb{E}[\mu]]\big) \\
&= \sqrt{n}(\mu - \mu) \qquad \text{(By Equation(16))} \\
&= 0 \qquad (19)
\end{aligned}$$

For the variance, we have

$$\begin{aligned}
\mathrm{Var}\big[\sqrt{n}(\hat{\mu}_n - \mu)\big] &= \mathbb{E}\Big[\big(\sqrt{n}(\hat{\mu}_n - \mu) - \mathbb{E}\big[\sqrt{n}(\hat{\mu}_n - \mu)\big]\big)^2\Big] \\
&= \mathbb{E}\big[(\sqrt{n}(\hat{\mu}_n - \mu))^2\big] \\
&= \mathbb{E}\Big[n(\hat{\mu}_n - \mu)^2\Big] \\
&= \mathbb{E}\left[n\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)^2\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - \mu)(X_j - \mu)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}[(X_i - \mu)(X_j - \mu)] \\
&= n(\mathbb{E}[\hat{\mu}_n^2] - 2\mu\mathbb{E}[\hat{\mu}_n] + \mathbb{E}[\mu^2]) \\
&= n((\mu^2 + \sigma^2/n) - 2\mu^2 + \mu^2) \\
&= \sigma^2 \qquad (20)
\end{aligned}$$

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?

A5 Answer:

- **Parts a:**

We need to find the Reduced Row Echelon Form of each matrix to obtain the basis.

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix} \xrightarrow{C_3 \leftarrow C_3 - C_1} \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix} \xrightarrow{C_2 \leftarrow C_2 - 2C_1} \begin{bmatrix} 1 & 0 & 0 \\ 1 & -2 & 2 \\ 1 & -1 & 1 \end{bmatrix} \xrightarrow{C_3 \leftarrow C_3 + C_2} \begin{bmatrix} 1 & 0 & 0 \\ 1 & -2 & 0 \\ 1 & -1 & 0 \end{bmatrix} \tag{21}$$

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \xrightarrow{C_3 \leftarrow C_3 - C_1 - C_2} \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \tag{22}$$

For now, we can find first and second columns are linearly independent. So, both rank of matrices A and B is 2.

- **Parts b:**

So, the basis for column-span of A is $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ -1 \end{bmatrix} \right\}$.

And, the basis for column-span of B is $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}$.

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 Answer:

- **Parts a:**

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

- **Parts b:**

  Base on the linear system, we can have the following.

$$Ax = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} x1 \\ x2 \\ x3 \end{bmatrix} = \begin{bmatrix} 0 \times x1 + 2 \times x2 + 4 \times x3 \\ 2 \times x1 + 4 \times x2 + 2 \times x3 \\ 3 \times x1 + 3 \times x2 + 1 \times x3 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ -4 \end{bmatrix} = b$$

  Then, we can get simultaneous equations

$$\begin{cases} 0 \times x1 + 2 \times x2 + 4 \times x3 = -2 \\ 2 \times x1 + 4 \times x2 + 2 \times x3 = -2 \\ 3 \times x1 + 3 \times x2 + 1 \times x3 = -4 \end{cases}$$

  The result of the equations is

$$\begin{cases} x1 = -2 \\ x2 = 1 \\ x3 = -1 \end{cases}$$

  Finally, we can obtain the solution is

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

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

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

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 Answer:

- **Part a:**

We set

$$x = [x_1, x_2, \cdots x_n]^\top; x^\top = [x_1, x_2, \cdots x_n]$$
$$y = [y_1, y_2, \cdots y_n]^\top; y^\top = [y_1, y_2, \cdots y_n]$$
$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,n} \end{bmatrix}$$
$$B = \begin{bmatrix} B_{1,1} & B_{1,2} & \cdots & B_{1,n} \\ B_{2,1} & B_{2,2} & \cdots & B_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ B_{n,1} & B_{n,2} & \cdots & B_{n,n} \end{bmatrix}$$

Recall the $f(x,y) = x^\top \mathbf{A} x + y^\top \mathbf{B} x + c$, then plug in previous equations, which yields

$$f(x,y) = [x_1, x_2, \cdots x_n] \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + [y_1, y_2, \cdots y_n] \begin{bmatrix} B_{1,1} & B_{1,2} & \cdots & B_{1,n} \\ B_{2,1} & B_{2,2} & \cdots & B_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ B_{n,1} & B_{n,2} & \cdots & B_{n,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + c$$

$$= \begin{bmatrix} \sum_{i=1}^n x_i A_{i,1} & \sum_{i=1}^n x_i A_{i,2} & \cdots & \sum_{i=1}^n x_i A_{i,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^n x_i B_{i,1} & \sum_{i=1}^n x_i B_{i,2} & \cdots & \sum_{i=1}^n x_i B_{i,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + c$$

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

$$= \sum_{j=1}^n (\sum_{i=1}^n (x_i A_{i,j} + x_i B_{i,j})) + c$$

So, we have the final answer as

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

- **Parts b:**

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

  and plug that in $\nabla_x f(x,y) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x,y) & \frac{\partial f}{\partial x_2}(x,y) & \dots & \frac{\partial f}{\partial x_n}(x,y) \end{bmatrix}^\top$

  Then, We have

$$
\begin{aligned}
\frac{\partial f(x,y)}{\partial x_k} &= \frac{\partial}{\partial x_k}(\sum_{j=1}^{n}(x_j)(\sum_{i=1}^{n}(x_i A_{i,j} + x_i B_{i,j})) + c) \\
&= \sum_{j=1}^{n}[\frac{\partial}{\partial x_k}(x_j)\sum_{i=1}^{n}(x_i A_{i,j} + y_i B_{i,j})] + \sum_{j=1}^{n}[x_j \frac{\partial}{\partial x_k}(\sum_{i=1}^{n}(x_i A_{i,j} + y_i B_{i,j}))] \\
&= \frac{\partial}{\partial x_k}(x_k)\sum_{i=1}^{n}(x_i A_{i,k} + y_i B_{i,k}) + \sum_{j=1}^{n}[x_j \sum_{i=1}^{n}(\frac{\partial}{\partial x_k}(x_i A_{i,j}))] \\
&= \sum_{i=1}^{n}(x_i A_{i,k} + y_i B_{i,k}) + \sum_{j=1}^{n} x_j [\frac{\partial}{\partial x_k}(x_k A_{k,j})] \\
&= \sum_{i=1}^{n}(x_i A_{i,k} + y_i B_{i,k}) + \sum_{j=1}^{n} x_j A_{k,j} \\
&= \sum_{i=1}^{n} x_i A_{i,k} + \sum_{i=1}^{n} y_i B_{i,k} + \sum_{i=1}^{n} x_i A_{k,i} \\
&= [Ax]_k + \left[A^\top x\right]_k + \left[B^\top y\right]_k \\
&= \left[Ax + A^\top x + B^\top y\right]_k \\
\nabla_x f(x,y) &= Ax + A^\top x + B^\top y \qquad (23)
\end{aligned}
$$

- **Parts c:**

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

  and plug that in $\nabla_y f(x,y) = \begin{bmatrix} \frac{\partial f}{\partial y_1}(x,y) & \frac{\partial f}{\partial y_2}(x,y) & \dots & \frac{\partial f}{\partial y_n}(x,y) \end{bmatrix}^\top$

  Then, We have

$$
\begin{aligned}
\frac{\partial f(x,y)}{\partial y_k} &= \frac{\partial}{\partial y_k}(\sum_{j=1}^{n}(x_j)(\sum_{i=1}^{n}(x_i A_{i,j} + y_i B_{i,j})) + c) \\
&= \sum_{j=1}^{n}(x_j)(\frac{\partial}{\partial y_k}(\sum_{i=1}^{n}(x_i A_{i,j} + y_i B_{i,j}))) \\
&= \sum_{j=1}^{n}(x_j)(\sum_{i=1}^{n}(\frac{\partial}{\partial y_k} y_i B_{i,j})) \\
&= \sum_{j=1}^{n} x_j B_{k,j} \\
&= [Bx]_k \\
\nabla_y f(x,y) &= Bx \qquad (24)
\end{aligned}
$$

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.

A8 Answer:

- **Part a:**

  For diag(v) and diag(w)

$$v = \begin{bmatrix} v_1 & 0 & ... & 0 \\ 0 & v_2 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & v_n \end{bmatrix}, v^{-1} = \begin{bmatrix} 1/v_1 & 0 & ... & 0 \\ 0 & 1/v_2 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & 1/v_n \end{bmatrix}, w = \begin{bmatrix} w_1 & 0 & ... & 0 \\ 0 & w_2 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & w_n \end{bmatrix}$$

  Since $diag(v)^{-1} = diag(w)$, we have

$$\begin{bmatrix} 1/v_1 & 0 & ... & 0 \\ 0 & 1/v_2 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & 1/v_n \end{bmatrix} = \begin{bmatrix} w_1 & 0 & ... & 0 \\ 0 & w_2 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & w_n \end{bmatrix}$$

  This gives us $\frac{1}{v_n} = w_n$ and recall to $g(v_n) = w_n$, then we have

$$\begin{aligned} g(v_n) &= w_n = \frac{1}{v_n} \\ g(v_n) &= \frac{1}{v_n} \end{aligned} \tag{25}$$

  Last, we can obtain the g, that is

$$g(m) = \frac{1}{m_i} \tag{26}$$

- **Parts b:**

$$
\begin{aligned}
||\mathbf{A}x||_2^2 &= (\mathbf{A}x)^\top(\mathbf{A}x) \\
&= \mathbf{A}^\top\mathbf{A}x^\top x \\
&= Ix^\top x \qquad \text{(Matrix times identity matrix, result is itself)} \\
&= x^\top x \\
&= ||x||_2^2
\end{aligned}
$$

- **Parts c:**

First, start with

$$
\begin{aligned}
\mathbf{B}\mathbf{B}^{-1} &= (\mathbf{B}\mathbf{B}^{-1})^\top \\
&= \mathbf{B}^\top(\mathbf{B}^{-1})^\top \\
&= \mathbf{B}(\mathbf{B}^{-1})^\top \qquad \text{(B is invertible)}
\end{aligned}
$$

Then we have

$$
\begin{aligned}
\mathbf{B}\mathbf{B}^{-1}\mathbf{B}^{-1} &= \mathbf{B}(\mathbf{B}^{-1})^\top\mathbf{B}^{-1} \\
\mathbf{B}^{-1}\mathbf{I} &= (\mathbf{B}^{-1})^\top\mathbf{I} \\
\mathbf{B}^{-1} &= (\mathbf{B}^{-1})^\top
\end{aligned}
$$

- **Parts d:**

We set λ as an eigenvalue of Matrix C, and there exist eigenvector v $\in$ V.

For example, we have $Cv = \lambda v$. So, $0 \leq v^\top Cv = \lambda vv^\top$. Since $vv^\top$ is always positive for any v, which means λ, the eigenvalue, is non-negative.

Programming

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

a. *[1 point]* Using native Python, implement the summation form.

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.

A9 Answer:

- **Part a:**

  Code:

```python
M1 = vanilla_matmul(A,x)
M2 = vanilla_matmul(vanilla_transpose(A),x)
M3 = vanilla_matmul(vanilla_transpose(B),y)
res = [0 for i in range(len(A))]
for i in range(len(A)):
    res[i] = M1[i]+ M2[i] +M3[i]
return res
```

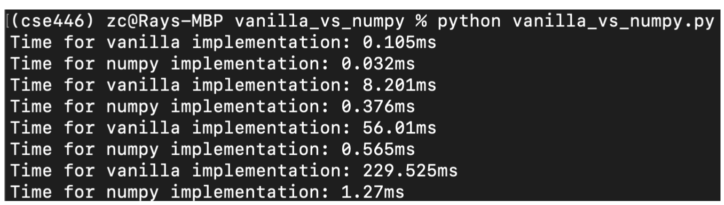
- **Part b:**

  Code:

```python
return (A @ x + np.transpose(A) @ x + np.transpose(B) @ y)
```

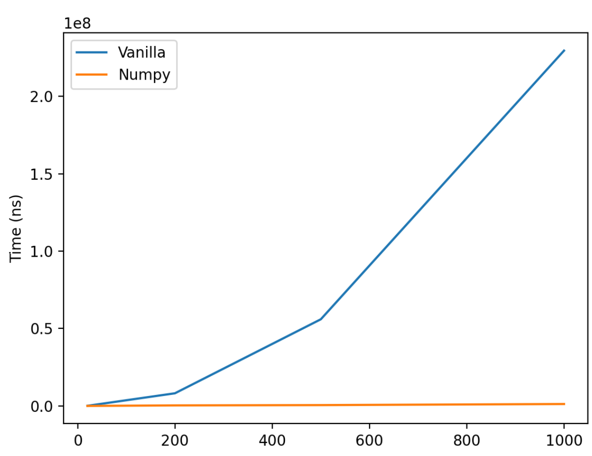
- **Part c:**

  Vanilla and Numpy Time:



```
(cse446) zc@Rays-MBP vanilla_vs_numpy % python vanilla_vs_numpy.py
Time for vanilla implementation: 0.105ms
Time for numpy implementation: 0.032ms
Time for vanilla implementation: 8.201ms
Time for numpy implementation: 0.376ms
Time for vanilla implementation: 56.01ms
Time for numpy implementation: 0.565ms
Time for vanilla implementation: 229.525ms
Time for numpy implementation: 1.27ms
```

  The time difference between vanilla and numpy is getting bigger with the increase of 'n.' Because the numpy has a lot of built-in functions, it can reduce the time complexity while the vanilla implementation has to do the matrix addition iteratively. Below is the graph shows n vs. time.



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.)

A10 Answer:

- **Part a:** Value for n (Hint: You will need to print it)

  $$n = int(np.ceil(1.0/(0.0025 * 2))) * *2 = 40,000$$

- **Part b:** How does empirical CDF change with k?

  As k increasing, the empirical CDF is more alike to $\widehat{F}_n(x)$.

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

```
plt.xlim([-3,3])
plt.xlabel("Observations"); plt.ylabel("Probability")
plt.show()
```