CSE6740 CDA Homework 0

Name:
GTID:

Deadline: Aug 31 st 11:59 pm ET

1 Linear Algebra

1.1 Rank

(a) If

$$A = \begin{pmatrix} 1 & 1 & -1 \\ 2 & -3 & 4 \\ 3 & -2 & 3 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & -2 & 3 \\ -2 & 4 & -6 \\ 5 & 1 & -1 \end{pmatrix},$$

then find the rank of AB and the rank of BA.

Solution:

We have

$$AB = \begin{bmatrix} -6 & 1 & -2 \\ 28 & -12 & 20 \\ 22 & -11 & 18 \end{bmatrix}, \quad BA = \begin{bmatrix} 6 & 1 & 0 \\ -12 & -2 & 0 \\ 4 & 4 & -4 \end{bmatrix}.$$

Let the rows of AB be r_1, r_2, r_3. We compute:

$$r_1 + r_2 = [-6 + 28,\ 1 - 12,\ -2 + 20] = [22,\ -11,\ 18] = r_3.$$

Thus, $r_3 = r_1 + r_2$, meaning the three rows are linearly dependent.

Perform the row operation

$$R_3 \leftarrow R_3 - R_1 - R_2,$$

which yields

$$R_3 \leftarrow r_3 - (r_1 + r_2) = 0.$$

Hence the third row becomes the zero row.

Now, look at the 2×2 minor formed by the first two rows and first two columns:

$$\begin{vmatrix} -6 & 1 \\ 28 & -12 \end{vmatrix} = (-6)(-12) - (1)(28) = 72 - 28 = 44 \neq 0.$$

A nonzero 2×2 minor means the first two rows are linearly independent, so there are exactly 2 independent rows left. Therefore,

$$\text{rank}(AB) = 2.$$

—

For BA, let the rows be s_1, s_2, s_3. Note that

$$s_2 = [-12,\ -2,\ 0] = -2[6,\ 1,\ 0] = -2s_1.$$

Thus, rows 1 and 2 are dependent.

Perform the row operation

$$R_2 \leftarrow R_2 + 2R_1,$$

which gives

$$R_2 = s_2 + 2s_1 = -2s_1 + 2s_1 = 0.$$

Now, check the 2×2 minor using rows 1 and 3 and columns 1 and 2:

$$\begin{vmatrix} 6 & 1 \\ 4 & 4 \end{vmatrix} = 6 \cdot 4 - 1 \cdot 4 = 24 - 4 = 20 \neq 0.$$

This shows s_1 and s_3 are independent, hence

$$\text{rank}(BA) = 2.$$

1.2 Eigen Vectors

Consider a small network of 3 web pages: A, B, and C. The probability that a user clicks from one page to another is given by the transition matrix M:

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

Here, each entry M_{ij} represents the probability that a user currently on page j clicks a link to page i.

1. Find the eigenvector $\vec{r}$ corresponding to the eigenvalue 1 of matrix M, representing the steady-state probabilities (PageRank) of each page.
2. Interpret the result: Which page has the highest rank, and what does this imply about user behavior on the website?

Solution:

We solve $(M - I)\mathbf{r} = 0$ for the transition matrix

$$M = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{2}{3} & 0 \end{pmatrix}.$$

Subtract I:

$$M - I = \begin{pmatrix} -1 & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \\ \frac{1}{2} & \frac{2}{3} & -1 \end{pmatrix}.$$

The system $(M - I)\mathbf{r} = \mathbf{0}$ gives

$$\begin{aligned} -r_A + \frac{1}{3}r_B + \frac{1}{2}r_C &= 0, \\ \frac{1}{2}r_A - r_B + \frac{1}{2}r_C &= 0, \\ \frac{1}{2}r_A + \frac{2}{3}r_B - r_C &= 0. \end{aligned}$$

From the second equation:

$$r_B = \frac{1}{2}r_A + \frac{1}{2}r_C \quad \Rightarrow \quad r_A = 2r_B - r_C.$$

Equating with the first equation $r_A = \frac{1}{3}r_B + \frac{1}{2}r_C$ gives

$$\frac{1}{3}r_B + \frac{1}{2}r_C = 2r_B - r_C \quad \Rightarrow \quad 10r_B = 9r_C \quad \Rightarrow \quad r_B = \frac{9}{10}r_C.$$

Substitute into $r_A = \frac{1}{3}r_B + \frac{1}{2}r_C$:

$$r_A = \frac{1}{3} \cdot \frac{9}{10}r_C + \frac{1}{2}r_C = \frac{3}{10}r_C + \frac{5}{10}r_C = \frac{4}{5}r_C.$$

Thus (unnormalized)

$$\mathbf{r} \propto \left(\frac{4}{5}, \frac{9}{10}, 1\right) = (8, 9, 10).$$

Normalizing so $r_A + r_B + r_C = 1$:

$$\boxed{\mathbf{r} = \left(\frac{8}{27}, \frac{9}{27}, \frac{10}{27}\right) \approx (0.2963, 0.3333, 0.3704)}.$$

Page C has the highest steady-state probability, meaning users spend the most time there in the long run.

1.3 Proof

Let $X \in \mathbb{R}^{m \times n}$ be mean-centered (each column has mean zero), and

$$C = \frac{1}{m-1}X^T X$$

its sample covariance matrix. Let $V \in \mathbb{R}^{n\times n}$ be orthogonal and whose columns are eigenvectors of C. Define the transformed data

$$Y = XV.$$

Solution:

$$Y = XV.$$

Compute the covariance of Y:

$$\begin{aligned}\mathrm{Cov}(Y) &= \frac{1}{m-1}Y^TY = \frac{1}{m-1}(XV)^T(XV)\\ &= \frac{1}{m-1}V^TX^TXV = V^T\left(\frac{1}{m-1}X^TX\right)V = V^TCV.\end{aligned}$$

Since the columns of V are eigenvectors of C, we have the eigendecomposition $C = V\Lambda V^T$ where Λ is diagonal with the eigenvalues of C. Therefore

$$\mathrm{Cov}(Y) = V^TCV = V^T(V\Lambda V^T)V = \Lambda,$$

which is diagonal.

1.4 Positive Semi-Definiteness

Let

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

(a) Show that A is symmetric.

(b) Compute the eigenvalues of A.

(c) Is A positive semi-definite (PSD)? Justify.

Solution:
xA matrix A is symmetric if $A = A^\top$.
The transpose of A is

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

Since $A = A^\top$, the matrix A is symmetric.

(b) Eigenvalues of A:

The eigenvalues λ satisfy

$$\det(A - \lambda I) = 0$$

where I is the identity matrix.

Calculate the determinant:

$$\det \begin{bmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{bmatrix} = (2-\lambda)(2-\lambda) - (-1)(-1) = (2-\lambda)^2 - 1 = 0$$

$$(2-\lambda)^2 - 1 = 0 \implies (2-\lambda)^2 = 1$$

$$2 - \lambda = \pm 1$$

$$2 - \lambda = 1 \implies \lambda = 1$$

$$2 - \lambda = -1 \implies \lambda = 3$$

Thus, the eigenvalues are:

$$\lambda_1 = 1, \quad \lambda_2 = 3$$

(c) Positive Semi-Definiteness of A:

A symmetric matrix A is positive semi-definite (PSD) if and only if all its eigenvalues are non-negative, i.e., $\lambda_i \geq 0$ for all i.

Since $\lambda_1 = 1 \geq 0$ and $\lambda_2 = 3 \geq 0$, all eigenvalues are positive.

Therefore, the matrix A is **positive definite** (which implies it is also positive semi-definite).

2 Matrix Calculus [25 pts]

2.1 Gradients and Jacobians

Let f be a differentiable function.

(a) If $f : \mathbb{R}^m \to \mathbb{R}$ (i.e. f returns a scalar), define the gradient $\nabla f(x)$ and state its dimensions.

(b) If $f : \mathbb{R}^m \to \mathbb{R}^n$ (i.e. f returns a vector), define the derivative (Jacobian matrix) $Df(x) = \dfrac{\partial f(x)}{\partial x}$ and state its dimensions.

Solution:

(a) For $f : \mathbb{R}^m \to \mathbb{R}$

$$\nabla f(x) = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_m \end{bmatrix} \in \mathbb{R}^m.$$

It is a column vector (size $m \times 1$).

(b) For $f : \mathbb{R}^m \to \mathbb{R}^n$ with component functions $f_1, \dots, f_n$,

$$Df(x) = \frac{\partial f(x)}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_m} \end{bmatrix} \in \mathbb{R}^{n\times m}.$$

Entry-wise: $(Df)_{ij} = \partial f_i/\partial x_j$.

.

2.2 Matrix Calculus of Quadratic Forms

Let $x \in \mathbb{R}^n$ be a column vector and A be an $n \times n$ constant matrix. Consider the scalar-valued function $f(x) = x^\top A, x$.
(a) Derive the gradient $\nabla_x f(x)$ (give the result in terms of x and A, and specify the gradient's size).
(b) If A is symmetric (i.e. $A = A^\top$), simplify your expression for the gradient from part (a) to a more familiar form. Hint: You can expand $f(x)$ as $\sum_{i,j} A_{ij}, x_i x_j$ and differentiate with respect to x_i to find each component of the gradient. (Alternatively, use known results for gradients of quadratic forms.)

Solution:
(a) Let $f(x) = x^\top A\, x,\ A \in \mathbb{R}^{n\times n}$.
General derivation

$$\big(\nabla_x f(x)\big)_k = \sum_{i=1}^{n} A_{ki} x_i + \sum_{j=1}^{n} A_{jk} x_j = \big(A^\top + A\big)_{k:} x.$$

Hence

$$\boxed{\nabla_x f(x) = \big(A + A^\top\big)\, x} \quad \text{(size } n \times 1).$$

(b) **Symmetric case ($A = A^\top$)**

$$\nabla_x f(x) = 2A\, x.$$

2.3 Hessian, Positive Definiteness, and Convex Optimization

Consider the quadratic function

$$f(x) \;=\; \frac{1}{2}\,x^T Q x \;-\; p^T x,$$

where $Q \in \mathbb{R}^{n\times n}$ is symmetric and $p \in \mathbb{R}^n$.

(a) Compute the Hessian $\nabla_x^2 f(x)$.
(b) Show that if Q is positive definite, then $f(x)$ has a unique global minimum.

Solution:
(a) First compute the gradient:

$$\nabla_x f(x) = \nabla_x\Big(\tfrac{1}{2}\,x^T Q x - p^T x\Big) = Qx - p,$$

using $\nabla_x(x^T Q x) = 2Qx$ when Q is symmetric. Hence the Hessian is

$$\nabla_x^2 f(x) = \nabla_x\,(Qx - p) = Q.$$

(b) If $Q \succ 0$ then $\nabla_x^2 f(x) = Q$ is positive definite for all x. A function with everywhere positive-definite Hessian is strictly convex, so it can have at most one global minimum. Moreover, since Q is invertible, setting $\nabla_x f(x) = 0$ yields a unique critical point, which must therefore be the unique global minimizer.

3 Probability & Statistics [25 pts]

3.1 Bayes' Theorem [4 pts]

Suppose the probability of snow is 20%, and the probability of a traffic accident is 10%. Suppose further that the conditional probability of an accident, given that it snows, is 40%. What is the conditional probability that it snows, given that there is an accident?

Solution:
By Bayes' Theorem,

$$P(\text{snow}|\text{accident}) = \frac{P(\text{snow})}{P(\text{accident})} \times P(\text{accident}|\text{snow}) = \frac{0.2}{0.1} \times 0.4 = 0.8$$

3.2 Random Variables and Distributions [12 pts]

Suppose X and Y have joint density function

$$f_{X,Y}(x,y) = \begin{cases} \frac{3}{13}\left(2 + x + 2xy + 4y^2\right), & 0 \le x \le 1,\ 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Compute $f_X(x)$ for all $x \in \mathbb{R}$.

Solution:
According to definition,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_0^1 f_{X,Y}(x,y)dy.$$

For $0 \le x \le 1$,

$$\int_0^1 f_{X,Y}(x,y)dy = \int_0^1 \frac{3}{13}\left(2 + x + 2xy + 4y^2\right)dy = \frac{10}{13} + \frac{6}{13}x.$$

Therefore,

$$f_X(x) = \begin{cases} \frac{10}{13} + \frac{6}{13}x, & 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

(b) Compute $f_Y(y)$ for all $y \in \mathbb{R}$.

Solution:
According to definition,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_0^1 f_{X,Y}(x,y)dx.$$

For $0 \le y \le 1$,

$$\int_0^1 f_{X,Y}(x,y)dx = \int_0^1 \frac{3}{13}\left(2 + x + 2xy + 4y^2\right)dx = \frac{15}{26} + \frac{3}{13}y + \frac{12}{13}y^2.$$

Therefore,

$$f_Y(y) = \begin{cases} \frac{15}{26} + \frac{3}{13}y + \frac{12}{13}y^2, & 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

(c) Determine whether or not X and Y are independent.

Solution:
X and Y are not independent.

Proof. For $0 \le x \le 1$,

$$\begin{aligned} f_X(x) f_Y(y) &= \left(\frac{6}{13}x + \frac{10}{13}\right)\left(\frac{15}{26} + \frac{3}{13}y + \frac{12}{13}y^2\right) \\ &= \frac{75}{169} + \frac{45}{169}x + \frac{30}{169}y + \frac{18}{169}xy + \frac{120}{169}y^2 + \frac{72}{169}xy^2 \\ &\neq \frac{3}{13}\left(2 + x + 2xy + 4y^2\right). \end{aligned}$$

Since $f_X(x) f_Y(y) = f_{X,Y}(x, y)$ is not satisfied for $0 \le x \le 1$ and $0 \le y \le 1$, X and Y are not independent. $\square$

(d) Prove that if two random variables X and Y are independent, then their covariance is zero.

Solution:
From the definition of covariance,

$$\mathrm{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

When X and Y are independent,

$$\begin{aligned} \mathbb{E}[XY] &= \int\int xy f_{X,Y}(x, y) dx dy \\ &= \int\int xy f_X(x) f_Y(y) dx dy \\ &= \left(\int x f_X(x) dx\right)\left(\int y f_Y(y) dy\right) \\ &= \mathbb{E}[X]\mathbb{E}[Y] \end{aligned}$$

Therefore, $\mathrm{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$.

3.3 Likelihood Inference [9 pts]

Consider the 1-dimensional Gaussian distribution

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\},$$

where μ denotes the mean and σ^2 denotes the variance. Suppose we have a data set of observations $\mathbf{x} = (x_1, x_2, ..., x_N)^\top$, where each x_i is independently sampled from $\mathcal{N}(x|\mu, \sigma^2)$.

(a) Write down the log-likelihood function for the observed data.

Solution:

$$\ln p\left(\mathbf{x}|\mu, \sigma^2\right) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi).$$

(b) Derive the maximum likelihood estimators μ_{ML} and σ^2_{ML} for the unknown parameters μ and σ^2, given the observations $\mathbf{x}$.

Solution:
Taking the partial derivatives of $\ln p\left(\mathbf{x}|\mu, \sigma^2\right)$ with respect to μ and σ^2, we obtain

$$\frac{\partial \ln p\left(\mathbf{x}|\mu, \sigma^2\right)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n - \mu),$$

$$\frac{\partial \ln p\left(\mathbf{x}|\mu, \sigma^2\right)}{\partial \sigma^2} = \frac{1}{2(\sigma^2)^2}\sum_{n=1}^{N}(x_n - \mu)^2 - \frac{N}{2\sigma^2}.$$

Setting these derivatives to zero gives

$$\mu_{\mathrm{ML}} = \frac{\sum_{n=1}^{N} x_n}{N},$$

$$\sigma^2_{\mathrm{ML}} = \frac{\sum_{n=1}^{N}(x_n - \mu_{\mathrm{ML}})^2}{N}.$$

(c) Determine whether the estimators μ_{ML} and σ^2_{ML} are unbiased estimators.

Solution:
μ_{ML} is unbiased estimator and σ^2_{ML} is not unbiased estimator.

Proof.

$$\mathbb{E}\left[\frac{\sum_{n=1}^{N} x_n}{N}\right] = \frac{\sum_{n=1}^{N}\mathbb{E}[x_n]}{N} = \frac{N\mu}{N} = \mu.$$

$$
\begin{aligned}
\frac{\sum_{n=1}^{N}(x_n-\mu_{\mathrm{ML}})^2}{N} &= \frac{\sum_{n=1}^{N}\left((x_n-\mu)-(\mu_{\mathrm{ML}}-\mu)\right)^2}{N} \\
&= \frac{\sum_{n=1}^{N}(x_n-\mu)^2 + N(\mu_{\mathrm{ML}}-\mu)^2 - 2\sum_{n=1}^{N}(x_n-\mu)(\mu_{\mathrm{ML}}-\mu)}{N} \\
&= \frac{\sum_{n=1}^{N}(x_n-\mu)^2 + N(\mu_{\mathrm{ML}}-\mu)^2 - 2N(\mu_{\mathrm{ML}}-\mu)^2}{N} \\
&= \frac{\sum_{n=1}^{N}(x_n-\mu)^2 - N(\mu_{\mathrm{ML}}-\mu)^2}{N}.
\end{aligned}
$$

Taking the expectation of $\frac{\sum_{n=1}^{N}(x_n-\mu_{\mathrm{ML}})^2}{N}$ gives

$$
\begin{aligned}
\mathbb{E}\left[\frac{\sum_{n=1}^{N}(x_n-\mu_{\mathrm{ML}})^2}{N}\right] &= \frac{\mathbb{E}\left[\sum_{n=1}^{N}(x_n-\mu)^2\right]}{N} - \mathbb{E}\left[(\mu_{\mathrm{ML}}-\mu)^2\right] \\
&= \sigma^2 - \frac{\sigma^2}{N} \\
&= \frac{N-1}{N}\sigma^2 \\
&\neq \sigma^2.
\end{aligned}
$$

□

References [2 pts]

Solution:
Please mention any AI tools, people, post or blog etc. you used.

4 Programming

Please use this link to download all the required files. This homework contains only a ipynb, which you can make a copy and run on Google Colab.

Deliverables

For the programming part, please submit your `.ipynb` file to the programming autograder. Then, use `File` (top-left corner) → `Print` to generate and submit a PDF. The PDF is for future manual grading despite it is not used for this homework. For HW0, as long as you submit, you will receive 100%.

Expected files

- `HW0.pdf`
- `HW0.ipynb`
- `hw0.ipynb - Colab.pdf`