CIS5200: Machine Learning Spring 2023

Homework 0

Release Date: January 12, 2023 *Due Date: January 20, 2023*

Name: Lucas Wu

PennKey: lucaswu

PennID: 17125403

1 Written Questions

A1

1. False. Consider the matrices
$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$$
and
$$B = \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$$
We find that
$$AB = \begin{bmatrix} 5 & 6 \\ 8 & 10 \end{bmatrix}$$
And that
$$BA = \begin{bmatrix} 5 & 8 \\ 6 & 10 \end{bmatrix}$$
Clearly, we see that
$$AB \neq BA$$

2. True. Suppose that we have $AB = C$ and $BA = D$. Define $A_{i,j}, B_{i,j}, C_{i,j}, D_{i,j}$ to be the entry in the ith row and jth column of the desired matrix. Note that by the definition of matrix multiplication, we have that $C_{i,i}$ equals the ith row of A multiplied by the ith column of B. Similarly, $D_{i,i}$ is the ith row of B multiplied by the ith column of A.

   Thus, we find that
$$tr(C) = \sum_{i=1}^{\dim(A)} C_{i,i} = \sum_{i=1}^{\dim(A)} \sum_{j=1}^{\dim(A)} A_{i,j}B_{j,i} = \sum_{j=1}^{\dim(A)} \sum_{i=1}^{\dim(A)} A_{i,j}B_{j,i} = \sum_{j=1}^{\dim(A)} D_{j,j} = tr(D)$$
   with the middle step applied using a simple swap of summation, and we are done.

3. False. Consider the matrices
$$A = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$$
and
$$B = \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$$
We find that
$$(AB)^T = \begin{bmatrix} 9 & 8 \\ 10 & 10 \end{bmatrix}$$
And that
$$A^T B^T = \begin{bmatrix} 5 & 6 \\ 10 & 14 \end{bmatrix}$$
Clearly, we see that
$$(AB)^T \neq A^T B^T$$

4. True. We assume here that both A and B are invertible; otherwise if non-invertible the determinant of one of them will be 0, and their product will also be non-invertible with determinant 0, satisfying the desired equality.

   It's relatively straightforward to show that if A and B represent elementary row operations, the property $det(AB) = det(A) \cdot det(B)$ holds. Note additionally that we can break down A into the product of a series of elementary row operations $a_1 a_2 a_3 \cdots a_m$ and B into the product of a series of elementary row operations $b_1 b_2 b_3 \cdots b_n$. Thus, we find that
$$det(AB) = det(a_1) \cdot det(a_2) \cdots det(a_m) \cdot det(b_1) \cdot det(b_2) \cdots det(b_n) = det(A) \cdot det(B)$$
   as desired.

A2

1. The rank is 2.

2.
$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$
$$\begin{bmatrix} 2 & 1 & 4 \end{bmatrix}$$

A3

1.
$$6, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$
$$\sqrt{2}, \begin{bmatrix} -\frac{13}{7} + \frac{9}{7}\sqrt{2} \\ \frac{5}{7} - \frac{11}{7}\sqrt{2} \\ 1 \end{bmatrix}$$
$$-\sqrt{2}, \begin{bmatrix} -\frac{13}{7} - \frac{9}{7}\sqrt{2} \\ \frac{5}{7} + \frac{11}{7}\sqrt{2} \\ 1 \end{bmatrix}$$

2. No, $-\sqrt{2}$ is negative.

3. Scaling matrices scales eigenvalues as well, and adding a scalar adds that to eigenvalue. With that knowledge, we know that our eigenvalues will be

$$6\lambda + \gamma, \lambda\sqrt{2} + \gamma, \text{ and } -\lambda\sqrt{2} + \gamma$$

4. We know that $Ax = \lambda x$ for eigenvalue λ, and that

$$A^2 x = A(Ax) = A(\lambda x) = \lambda \cdot \lambda x = \lambda^2 x$$

So the eigenvalues of A^2 are λ^2.

A4

1. w

2. $-2w + 2 \cdot ww^\top x$

3.
$$\frac{\exp(-yw^\top x)}{1 + \exp(-yw^\top x)} \cdot (-yw)$$

4. Note that the output of $x^\top Ax$ is (defining the ith entry of x as x_i)

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

So our desired gradient is (noting the abuse of notation and the absence of the directional vectors)

$$\sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j} x_j + A_{j,i} x_j = Ax + A^\top x = \boxed{(A + A^\top)x}$$

When the matrix is symmetric, we know that $A_{i,j} = A_{j,i}$ (alternatively $A = A^\top$), so the gradient is now

$$2 \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j} x_j = \boxed{2Ax}$$

A5

1. We need $b = \mathbf{0}$.

2. We first scale w and b such that w is normal. This gives us the new equation $\frac{w^\top}{||w||}x + \frac{b}{||w||} = 0$ Note that the magnitude of the projection of x_0 onto $\frac{w}{||w||}$ is given by $\frac{w^T x_0}{||w||}$. Adding the scalar value to find the direct distance, we find that our desired answer is

$$\boxed{\left| \frac{w^\top x_0}{||w||} + \frac{b}{||w||} \right|}$$

A6

1. True by triangle inequality.
2. True (worst case scenario $x = \mathbf{1}$).
3. True (also by triangle inequality).
4. False (counterexamples exist, consider all 0s with one 0.5 entry).

A7

1. $$\frac{\partial f}{\partial x_1} = 2x_1$$
$$\frac{\partial f}{\partial x_2} = 4x_2^3 - 2x_2$$
Hessian:
$$\begin{bmatrix} 2 & 0 \\ 0 & 12x_2^2 - 2 \end{bmatrix}$$
No, it is not always convex because the Hessian is not positive semi-definite for all x_1 and x_2.
2. The gradient is 0 when the following is true
$$2x_1 = 0 \implies x_1 = 0$$
$$4x_2^3 - 2x^2 = 0 \implies x_2 = 0 \text{ or } \pm\sqrt{2}$$
Thus, the critical points are
$$(0, 0)$$
$$(0, \sqrt{2})$$
$$(0, -\sqrt{2})$$
3. Note that the function obtains minimums when $x_2 = \pm\sqrt{2}$ and $x_1 = 0$, thus giving the answer $\boxed{2}$

A8 From conditional probability (we can alternatively use Bayes' rule) we have

$$\frac{0.1 \cdot 0.85}{0.1 \cdot 0.85 + 0.9 \cdot 0.05} = \boxed{\frac{17}{26}}$$

A9

1. $a = \frac{1}{\sigma}$ $b = -\frac{\mu}{\sigma}$
2. We have the following:
$$\text{Var}[z] = \mathbb{E}[z^2] - \mathbb{E}[z]^2$$
$$\mathbb{E}[z^2] = \text{Var}[z] + \mathbb{E}[z]^2 = \boxed{\sigma^2 + \mu^2}$$
3. The distribution will be $z + \bar{z} \sim N(\mu + \bar{\mu}, \sigma^2 + \bar{\sigma}^2)$

A10

1. With parameter $p = \frac{1}{2}$, this is $\frac{1}{p} = \boxed{2}$.
2. There are three possible scenarios, two of which only one of them is heads, and the last of which both are heads. This gives a final answer of $\boxed{\frac{1}{3}}$.