Assignment 1 Solutions

1 Convexity and Linear Algebra

1.1 Taylor Expansions

a) **Solution:**
Gradient of f can be computed by:
$\nabla f =< 4x_1, 2x_2 + 2x_3, 2x_3 + 2x_2 >$
Hessian of f:

$$H = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix}$$

b) **solution:**
The second order Taylor expansion of $f(x)$ at point x_0 is given by:

$$f(x) \approx f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T H_f(x_0)(x - x_0)$$

Given the result from part(a), we will evaluate this expression at point $x_0 = (0, 0, 0)$.
Substituting these values into the Taylor expansion formula, with $x - x_0 = x$:

$$f(x) \approx 0 + \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$
$$f(x) \approx \frac{1}{2}\left(4x_1^2 + 2x_2^2 + 4x_2x_3 + 2x_3^2\right)$$
$$f(x) \approx 2x_1^2 + x_2^2 + 2x_2x_3 + x_3^2$$

c) **solution:**
We know that:
A function is convex if its Hessian is positive semi-definite everywhere
It is strictly convex if its Hessian is positive definite everywhere

To determine this, we find the eigenvalues of the Hessian matrix H_f.

$$H_f = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix}$$

The eigenvalues λ are found by solving $\det(H_f - \lambda I) = 0$:

$$
\begin{aligned}
\det \begin{bmatrix} 4-\lambda & 0 & 0 \\ 0 & 2-\lambda & 2 \\ 0 & 2 & 2-\lambda \end{bmatrix} &= 0 \\
(4-\lambda)[(2-\lambda)(2-\lambda) - (2)(2)] &= 0 \\
(4-\lambda)\left[\left(4 - 4\lambda + \lambda^2\right) - 4\right] &= 0 \\
(4-\lambda)\left(\lambda^2 - 4\lambda\right) &= 0 \\
(4-\lambda)\lambda(\lambda - 4) &= 0
\end{aligned}
$$

The eigenvalues are $\lambda_1 = 4$, $\lambda_2 = 4$, and $\lambda_3 = 0$.

Therefore, since: A matrix is positive semi-definite if all its eigenvalues are non-negative, it is positive definite if all its eigvenvalues are strictly positive
The Hessian Matrix H_f is positive semi-definite but not positive definite.
Therefore, the function is **convex** but **not strictly convex**

1.2 Matrix Rank and Inverse

a) **Solution:**
WTS: $Ax = 0 \iff x = 0$
$\Rightarrow$:
can write $Ax = 0$ as the linear combination of columns of A:

$$x_1 a_1 + x_2 a_2 + \cdots + x_n a_n = 0$$

Since A is full rank, its n columns are all linearly independent
Therefore, by the definition of linearly independent, the only way for the above equation to be true is if all coefficients are zero:

$$x_1 = x_2 = \cdots = x_n = 0$$

which means $x = 0$.
$\Leftarrow$:
If $x = 0$, then it is trivial that $Ax = A0 = 0$.

b) **Solution:**
Now let's prove that $A^T A$ is positive definite To prove this, let's evaluate the expression:

$$x^T (A^T A) x$$

$\forall x \in \mathbb{R}^n, x \neq \mathbf{0}$

$$
\begin{aligned}
x^T (A^T A) x &= (x^T A^T)(Ax) \\
&= (Ax)^T (Ax) \\
&= ||Ax||^2
\end{aligned}
$$

$||Ax||^2$ is 0 $\iff Ax = 0$
From part a, we have shown that $\forall x \neq 0$, $Ax \neq 0$
Therefore, for any non-zero vector x, the vector Ax will also be non-zero
This means that its squared form is strictly positive:

$$||Ax||^2 > 0$$

Therefore, it means that $x^T(A^TA)x > 0$ for all $x \neq 0$.
Therefore, A^TA is **positive definite**

WTS: **Symmetric Positive Definite Matrix is Always Invertible**
From the previous part b, we have shown that A^TA is **positive definite** Note that A^TA is always symmetric by Matrix Multiplication and definition of Transpose

Let M be a symmetric positive definite matrix. By the spectral theorem, it has an eigendecomposition $M = Q\Lambda Q^\top$, where:

- Q is an orthogonal matrix ($Q^\top Q = I$) whose columns are the eigenvectors of M.
- Λ is a diagonal matrix containing the positive eigenvalues λ_i of M.

We can construct the inverse M^{-1} as follows:

$$M^{-1} = Q\Lambda^{-1}Q^\top$$

The inverse of the diagonal matrix, Λ^{-1}, exists because all its diagonal entries λ_i are non-zero. Λ^{-1} is simply a diagonal matrix with entries $1/\lambda_i$. Let's verify this is the correct inverse:

$$\begin{aligned}
MM^{-1} &= \left(Q\Lambda Q^\top\right)\left(Q\Lambda^{-1}Q^\top\right) \\
&= Q\Lambda\left(Q^\top Q\right)\Lambda^{-1}Q^\top \\
&= Q\Lambda(I)\Lambda^{-1}Q^\top \\
&= Q\left(\Lambda\Lambda^{-1}\right)Q^\top \\
&= Q(I)Q^\top \\
&= QQ^\top = I
\end{aligned}$$

Same logic applied to showing $M^{-1}M = I$.

Therefore, since we can always compose an inverse of M by eigendecomposition, which means that M is **invertible**.

Therefore, we have shown that any symmetric positive definite matrix is always invertible by using eigendecomposition to construct the inverse.

Since we have previously shown that A^TA is positive definite and A^TA is symmetric by definition, A^TA is invertible.

2 SVD and Eigendecomposition

a) **Solution:**

$$B = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 5 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

b) **Solution:** To prove that the given form is the eigendecomposition of $B = U\Lambda U^T$ We must first compute the eigenvalues of B:
By solving the linear equation:

$$B - \lambda I = 0$$

the eigenvalues of B are:
$\lambda_1 = 6, \lambda 2 = 1, \lambda_3 = 0$

Now, the quick way to verify the above matrix decomposition is an egigendecomposition of B, is that we check each column of alleged U, denoted by u_i, wether for each u_i, we have the following equation holds:

$$\Lambda u_i = \lambda_i p_i$$

Now, let's check if this entity holds one by one:

- $\lambda_1 = 6, u_1 = \frac{1}{\sqrt{30}} \begin{bmatrix} -2 \\ -5 \\ -1 \end{bmatrix}$:

$$\Lambda u_1 = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 5 & 1 \\ 0 & 1 & 1 \end{bmatrix} \frac{1}{\sqrt{30}} \begin{bmatrix} -2 \\ -5 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{30}} \begin{bmatrix} -12 \\ -30 \\ 6 \end{bmatrix}$$

$$\lambda_1 u_1 = \frac{1}{\sqrt{30}} \begin{bmatrix} -12 \\ -30 \\ 6 \end{bmatrix}$$

Therefore,$\Lambda u_1 = \lambda_1 u_1$

- $\lambda_2 = 1, u_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}$:

$$\Lambda u_2 = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 5 & 1 \\ 0 & 1 & 1 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix}$$

$$\lambda_2 u_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix}$$

Therefore,$\Lambda u_2 = \lambda_2 u_2$

- $\lambda_3 = 0, u_3 = \frac{1}{\sqrt{6}}\begin{bmatrix}2\\-1\\1\end{bmatrix}$:

$$\Lambda u_3 = \begin{bmatrix}1 & 2 & 0\\2 & 5 & 1\\0 & 1 & 1\end{bmatrix}\frac{1}{\sqrt{6}}\begin{bmatrix}2\\-1\\1\end{bmatrix} = \frac{1}{\sqrt{6}}\begin{bmatrix}0\\0\\0\end{bmatrix}$$

$$\lambda_3 u_3 = \frac{1}{\sqrt{6}}\begin{bmatrix}0\\0\\0\end{bmatrix}$$

Therefore,$\Lambda u_3 = \lambda_3 u_3$ Therefore, now we can verify that the above matrix is the eigen-decomposition of matrix B

c) **Solution:** Since the **singular values** of matrix A denoted as σ_i are the square roots of the non-zero eigenvalues of matrix AA^T
Since we have found the eigenvalues of $B = AA^T$:
$\lambda_1 = 6, \lambda 2 = 1, \lambda_3 = 0$
We now can calculate the singular values of A by taking the square root of non-zero eigenvalues of B:

$$\sigma_1 = \sqrt{6}, \sigma_2 = \sqrt{1}$$

Therefore, we can construct

$$\Sigma = \begin{bmatrix}\sigma_1 & 0\\0 & \sigma_2\\0 & 0\end{bmatrix} = \begin{bmatrix}\sqrt{6} & 0\\0 & 1\\0 & 0\end{bmatrix}$$

d) **Solution:**

- For U: first, we prove that U is orthogonal:
For columns of U, we check whether $u_1 \cdot u_2 = 0$

$$(1/2, \sqrt{3}/2)(-\sqrt{(}3)/2, 1/2) = 0$$

Then we check the determinant, whether $\det(U) = 1$

$$\det(U) = 1/4 + 3/4 = 1$$

Therefore, basis vectors of transformation is orthonormal, the matrix U is a rotation matrix

- For V: first, we prove that V is orthogonal:
For columns of V, we check whether $u_1 \cdot u_2 = 0$

$$(\sqrt{2}/2, \sqrt{2}/2)(-\sqrt{(}2)/2, \sqrt{2}/2) = 0$$

Then we check the determinant, whether $\det(V) = 1$

$$\det(V) = 1/2 + 1/2 = 1$$

Therefore, basis vectors of transformation is orthonormal, the matrix V is a rotation matrix, V^T is also a rotation matrix by rotation matrix's property

Now we compute the rotation angles for U and V:

- for U: $\theta_u = \frac{\pi}{3}$

$$U = \begin{bmatrix} \cos(\frac{\pi}{3}) & -\sin(\frac{\pi}{3}) \\ \sin(\frac{\pi}{3}) & -\cos(\frac{\pi}{3}) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$$

- for V: $\theta_v = -\frac{\pi}{4}$

$$V = \begin{bmatrix} \cos(\frac{\pi}{3}) & -\sin(\frac{\pi}{3}) \\ \sin(\frac{\pi}{3}) & -\cos(\frac{\pi}{3}) \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

e) **Solution:**
Finally, let's explain the gemetric intuition behind the singular value decomposition matrix:
First, V^T rotates the input space by $-\pi/4$, then it scales the input space by 4 in r rotated x direction and $1/2$ in rotated y direction
Then it rotates again by $\pi/3$ clockwise to final position

3 Convexity

a) **Solution:**

We need to prove that the Huber loss function with parameter $\delta > 0$:

$$\text{Huber}_\delta(x) := \begin{cases} \frac{1}{2}x^2, & |x| \le \delta, \\ \delta|x| - \frac{1}{2}\delta^2, & |x| > \delta, \end{cases}$$

is convex using only the definition of convex functions.

To prove convexity, we need to show that for any $x, y \in \mathbb{R}$ and any $t \in [0,1]$:

$$\text{Huber}_\delta(tx + (1-t)y) \le t \cdot \text{Huber}_\delta(x) + (1-t) \cdot \text{Huber}_\delta(y)$$

We consider four cases based on the values of x and y:

Case 1: $|x| \le \delta$ and $|y| \le \delta$

In this case, we need to check if $|tx + (1-t)y| \le \delta$.

$$\begin{aligned} |tx + (1-t)y| &\le t|x| + (1-t)|y| \quad \text{(triangle inequality)} \\ &\le t\delta + (1-t)\delta = \delta \end{aligned}$$

Therefore, all three points fall in the quadratic region:

$$\begin{aligned} \text{Huber}_\delta(tx + (1-t)y) &= \frac{1}{2}(tx + (1-t)y)^2 \\ &\le t \cdot \frac{1}{2}x^2 + (1-t) \cdot \frac{1}{2}y^2 \quad \text{(convexity of } \frac{1}{2}x^2) \\ &= t \cdot \text{Huber}_\delta(x) + (1-t) \cdot \text{Huber}_\delta(y) \end{aligned}$$

Case 2: $|x| > \delta$ and $|y| > \delta$, with x and y having the same sign

Without loss of generality, assume $x, y > \delta$. Then $tx + (1-t)y > \delta$ for $t \in [0, 1]$.

$$\begin{aligned}\text{Huber}_\delta(tx + (1-t)y) &= \delta(tx + (1-t)y) - \frac{1}{2}\delta^2 \\ &= t(\delta x - \frac{1}{2}\delta^2) + (1-t)(\delta y - \frac{1}{2}\delta^2) \\ &= t \cdot \text{Huber}_\delta(x) + (1-t) \cdot \text{Huber}_\delta(y)\end{aligned}$$

The function is linear in this region, so the inequality holds with equality.

Case 3: $|x| > \delta$ and $|y| > \delta$, with x and y having opposite signs

Let $x > \delta$ and $y < -\delta$. Then $tx + (1-t)y$ may fall in any region. We have:

$$\begin{aligned}t \cdot \text{Huber}_\delta(x) + (1-t) \cdot \text{Huber}_\delta(y) &= t(\delta x - \frac{1}{2}\delta^2) + (1-t)(\delta(-y) - \frac{1}{2}\delta^2) \\ &= t\delta x - (1-t)\delta y - \frac{1}{2}\delta^2\end{aligned}$$

Since the Huber function is convex on each piece and continuous at the boundary, the convexity property holds.

Case 4: One point inside $[-\delta, \delta]$ and one outside

By the continuity of the Huber function at $\pm\delta$ and the convexity of each piece, the overall function remains convex.

Therefore, $\text{Huber}_\delta(x)$ is convex for all $\delta > 0$.

b) **Solution:**

We need to prove that

$$f(\boldsymbol{x}) = \|A\boldsymbol{x} + \boldsymbol{b}\|_2 + \lambda\|\boldsymbol{x}\|_\infty$$

is convex, where $A \in \mathbb{R}^{n\times n}$, $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^n$, and $\lambda \geq 0$.

Step 1: Show that any norm is convex

Let $\|\cdot\|$ be any norm. For any $\boldsymbol{u}, \boldsymbol{v}$ and $t \in [0, 1]$:

$$\begin{aligned}\|t\boldsymbol{u} + (1-t)\boldsymbol{v}\| &\leq \|t\boldsymbol{u}\| + \|(1-t)\boldsymbol{v}\| \quad \text{(triangle inequality)} \\ &= t\|\boldsymbol{u}\| + (1-t)\|\boldsymbol{v}\| \quad \text{(positive homogeneity)}\end{aligned}$$

Therefore, any norm is a convex function. In particular, both $\|\cdot\|_2$ and $\|\cdot\|_\infty$ are convex.

Step 2: Show $\|A\boldsymbol{x} + \boldsymbol{b}\|_2$ is convex

Let $g(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}$. This is an affine transformation. Let $h(\boldsymbol{z}) = \|\boldsymbol{z}\|_2$, which is convex by Step 1.

Then $\|A\boldsymbol{x} + \boldsymbol{b}\|_2 = h(g(\boldsymbol{x}))$.

Using the property that $f(A\boldsymbol{x} + \boldsymbol{b})$ is convex if f is convex (composition of convex function with affine transformation), we have that $\|A\boldsymbol{x} + \boldsymbol{b}\|_2$ is convex.

Step 3: Show $\|\boldsymbol{x}\|_\infty$ is convex

From Step 1, $\|\boldsymbol{x}\|_\infty$ is convex.

Step 4: Combine using the property of weighted sums

We have:

- $\|A\boldsymbol{x} + \boldsymbol{b}\|_2$ is convex (Step 2)
- $\|\boldsymbol{x}\|_\infty$ is convex (Step 3)
- $\lambda \geq 0$ (given)

Using the property that $\sum_i w_i f_i(\boldsymbol{x})$ is convex if f_i are convex and $w_i \geq 0$, we have:

$$f(\boldsymbol{x}) = \underbrace{\|A\boldsymbol{x} + \boldsymbol{b}\|_2}_{\text{convex}} + \underbrace{\lambda}_{\geq 0} \cdot \underbrace{\|\boldsymbol{x}\|_\infty}_{\text{convex}}$$

is convex.

Therefore, $f(\boldsymbol{x}) = \|A\boldsymbol{x} + \boldsymbol{b}\|_2 + \lambda\|\boldsymbol{x}\|_\infty$ is convex.

c) **Solution:**

We need to prove that the Swish activation function

$$f(x) = x\sigma(x) = \frac{x}{1 + e^{-x}}$$

is neither convex nor concave on $\mathbb{R}$.

To show that f is neither convex nor concave, we will compute $f''(x)$ and show that it changes sign.

Step 1: Compute $f'(x)$

Given that $\sigma'(x) = \sigma(x)(1-\sigma(x))$, we use the product rule and since $\sigma^{'}(x) = \sigma(x)(1-\sigma(x))$

$$\begin{aligned} f'(x) &= \sigma(x) + x\sigma'(x) \\ &= \sigma(x) + \sigma(x)(1 - \sigma(x))x \end{aligned}$$

Step 2: Compute $f''(x)$

$$\begin{aligned} f''(x) &= \sigma'(x) + \frac{d}{dx}(x\sigma'(x)) = \sigma'(x) + (\sigma'(x) + x\sigma''(x)) \\ f''(x) &= 2\sigma'(x) + x\sigma''(x) \end{aligned}$$

Now we find $\sigma''(x)$ by differentiating $\sigma'(x) = \sigma(x) - \sigma(x)^2$:

$$\sigma''(x) = \sigma'(x) - 2\sigma(x)\sigma'(x) = \sigma'(x)(1 - 2\sigma(x))$$

Substitute this back into the expression for $f''(x)$:

The term $\sigma'(x) = \sigma(x)(1 - \sigma(x))$ is always positive because $\sigma(x) \in (0, 1)$ for all $x \in \mathbb{R}$. Therefore, the sign of $f''(x)$ is determined by the sign of the term $S(x) = 2 + x(1 - 2\sigma(x))$. Let's test the sign of $S(x)$ at two different points: - At $x = 0$:

$$\sigma(0) = \frac{1}{1 + e^0} = \frac{1}{2}$$
$$S(0) = 2 + 0\left(1 - 2\left(\frac{1}{2}\right)\right) = 2$$

Since $S(0) = 2 > 0$, we have $f''(0) > 0$. - As $x \to \infty$:

As x becomes very large, $e^{-x} \to 0$, so $\sigma(x) \to 1$. The term $(1 - 2\sigma(x))$ approaches $1 - 2(1) = -1$. So, for large positive $x, S(x) \approx 2 + x(-1) = 2 - x$. This value is negative for $x > 2$. For example, at $x = 5, S(5) \approx -3 < 0$, which means $f''(5) < 0$.

Since we have found a point where $f''(x) > 0$ (e.g., $x = 0$) and a point where $f''(x) < 0$ (e.g., $x = 5$), the second derivative changes sign.

Since $f''(x)$ changes sign on $\mathbb{R}$, the function f is neither convex (which would require $f''(x) \geq 0$) nor concave (which would require $f''(x) \leq 0$) on $\mathbb{R}$.

Therefore, the Swish activation function is neither convex nor concave on $\mathbb{R}$.