Assignment 1

1. Given a system of linear equatioins below for a vector $\mathrm{x} = [x_1, x_2, x_3]^\top$

$x_1 + 2x_2 = 1$

$3x_1 + x_2 + 4x_3 = 7$

$-2x_1 + x_2 - 4x_3 = -6$

1.1 construct the A matrix and b vector

Matrix A:

$$\begin{bmatrix} 1 & 2 & 0 \\ 3 & 1 & 4 \\ -2 & 1 & -4 \end{bmatrix}$$

Vector b:

$$\begin{bmatrix} 1 \\ 7 \\ -6 \end{bmatrix}$$

1.2 Compute eigenvalues and eigenvector fo A

If we denote the eigenvector as v and eigenvalues as λ. For all possible eigenvalues and eigenvectors, we have the following equation:

$$\mathrm{Av} = \lambda \mathrm{v}$$

This can be equivalent to

$$\begin{aligned} \mathrm{Av} &= \lambda \mathrm{Iv} \\ (\mathrm{A} - \lambda \mathrm{I})\mathrm{v} &= 0 \\ \Rightarrow \det((\mathrm{A} - \lambda \mathrm{I})\mathrm{v}) &= 0 \end{aligned}$$

Where I is an 3-by-3 identity matrix. Vector v cannot be o vector, and matrix $\mathrm{A} - \lambda \mathrm{I}$ should be a singular matrix with determine = 0 .

To expand the matrix:

$$\begin{aligned} &\det\left(\begin{bmatrix} 1-\lambda & 2 & 0 \\ 3 & 1-\lambda & 4 \\ -2 & 1 & -4-\lambda \end{bmatrix}\right) = 0 \\ &\Rightarrow (\ 1-\lambda)[(1-\lambda\ ) * (-4-\lambda) - 1 * 4] + \\ &\quad 3 * [1 * 0 - 2 * (-4-\lambda)] + -2 * [2 * 4 - 0 * (1-\lambda)] = 0 \\ &= -\lambda^3 - 2\lambda^2 + 11\lambda - 8 + 24 + 6\lambda - 16 \\ &= -\lambda(\lambda^2 + 2\lambda - 17) \\ &= \lambda((\lambda+1)^2 - 18) \\ &\Rightarrow \lambda_1 = 0, \lambda_2 = -1 - \sqrt{18}, \lambda_3 = -1 + \sqrt{18} \end{aligned}$$

Bring the eigenvalue back to the equation

$$\begin{aligned} &\mathrm{Av} = \lambda \mathrm{v} \\ &\Rightarrow (1+3-2-\lambda)v_1 + (2+1+1-\lambda)v_2 + (-\lambda)v_3 = 0 \\ &\Rightarrow \begin{bmatrix} (1-\lambda)v_1 + 2v_2 + 0v_3 \\ 3v_1 + (1-\lambda)v_2 + 4v_3 \\ -2v_1 + v_2 + (-4-\lambda)v_3 \end{bmatrix} = 0 \end{aligned}$$

Another hidden condition is the each eigen vector should have the norm as 1:

$$\sqrt{v_1^2 + v_2^2 + v_3^2} = 1$$

Solve this function set, we have the eigen vector as:

The corresponding eignvector for $\lambda_1 = 0$: $[-\frac{8}{\sqrt{105}}, \frac{4}{\sqrt{105}}, \frac{5}{\sqrt{105}}]^\top$

The corresponding eignvector for $\lambda_2 = -1 - \sqrt{18}$: $[-\frac{2}{2\sqrt{15+3\sqrt{18}}}, \frac{2+\sqrt{18}}{2\sqrt{15+3\sqrt{18}}}, -\frac{4+\sqrt{18}}{2\sqrt{15+3\sqrt{18}}}]^\top$

The corresponding eignvector for $\lambda_2 = -1 + \sqrt{18}$: $[\frac{2}{2\sqrt{15-3\sqrt{18}}}, -\frac{2-\sqrt{18}}{2\sqrt{15-3\sqrt{18}}}, \frac{4-\sqrt{18}}{2\sqrt{15-3\sqrt{18}}}]^\top$

1.3 Compute the determinant of A and the rank of A.

By eigendecomposition:

$$A = U^{-1}\Sigma U$$

Via 1.2, we know Σ is only a 2-by-2 diagnal matrix. So the determinant of A is 0, and the rank of A is 2.

1.4 solve x interms A and b. Is there an unique solution for x?

$$\begin{aligned} \mathrm{Ax} &= \mathrm{b} \\ \mathrm{A}^{-1}\mathrm{Ax} &= \mathrm{A}^{-1}\mathrm{b} \\ \mathrm{x} &= \mathrm{A}^{-1}\mathrm{b} \end{aligned}$$

However, there is no unique solution, because our unknown variable is 3 dim, but the rank of the matrix A is only $2 < 3$. So there is multiple solution.

2. Given $f(\mathrm{x}) = \log\{(A\mathrm{x}+b)^\top(C\mathrm{x}+d)+\lambda\mathrm{x}^\top\mathrm{x}\}$, where x is Nx1 vector, A and C are MxN matrix, b and d are Mx1 vector, λ is a scalr. Derive the equation ofr computing $\frac{\partial f(\mathrm{x})}{\partial \mathrm{x}}$

We can decompose $f(x)$ into several steps:

- $f = logS$
- $S = M_1 + \lambda M_2$
- $M_1 = P_1^\top P_2$
- $M_2 = \mathrm{x}^\top\mathrm{x}$
- $P_1 = A\mathrm{x} + b$
- $P_2 = C\mathrm{x} + d$

According to the chain rule:

$$\begin{aligned} \frac{\partial f}{\partial \mathrm{x}} &= \frac{\partial f}{\partial S}\left[\frac{\partial S}{\partial M_1}\left(\frac{\partial M_1}{\partial P_1}\frac{\partial P_1}{\partial \mathrm{x}} + \frac{\partial M_1}{\partial P_2}\frac{\partial P_2}{\partial \mathrm{x}}\right) + \frac{\partial S}{\partial M_2}\frac{\partial S}{\partial \mathrm{x}}\right] \\ \frac{\partial f}{\partial S} &= \frac{1}{(A\mathrm{x}+b)^\top(C\mathrm{x}+d)+\lambda\mathrm{x}^\top\mathrm{x}}\left[\left(\frac{\partial M_1}{\partial P_1}\frac{\partial P_1}{\partial \mathrm{x}} + \frac{\partial M_1}{\partial P_2}\frac{\partial P_2}{\partial \mathrm{x}}\right) + 2\lambda\mathrm{x}\right] \\ \frac{\partial f}{\partial S} &= \frac{1}{(A\mathrm{x}+b)^\top(C\mathrm{x}+d)+\lambda\mathrm{x}^\top\mathrm{x}}\left[\left(A^\top P_2 + C^\top P_1\right) + 2\lambda\mathrm{x}\right] \\ &= \frac{1}{(A\mathrm{x}+b)^\top(C\mathrm{x}+d)+\lambda\mathrm{x}^\top\mathrm{x}}\left[\left(A^\top(C\mathrm{x}+d) + C^\top(A\mathrm{x}+b)\right) + 2\lambda\mathrm{x}\right] \end{aligned}$$

3. Let $\mathrm{U}^{N\times 1}$,$\mathrm{V}^{K\times 1}$, $\mathrm{X}^{M\times 1}$ be vectors and $\mathrm{A}^{N\times K}$ a matrix

- Compute $\frac{\partial\sigma(U^\top AV)}{\partial X}$, where U and V are a function of X but A is not.
- Compute $\frac{\partial\sigma(U^\top AV)}{\partial X}$, where U and V are Not a function of X but A is.

where $\sigma(z)$ is the function and it:

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

3.1 Compute $\frac{\partial\sigma(U^\top AV)}{\partial X}$

Old trick, we can decompose it into 2 steps:

- $f = \frac{1}{s}, s \in \mathbb{R}$
- $s = 1 + e^{-z}, z \in \mathbb{R}$
- $z = U^\top AV$

$$\begin{aligned}
\frac{\partial\sigma(U^\top AV)}{\partial X} &= \frac{\partial f}{\partial s}\frac{\partial s}{\partial z}\left(\frac{\partial z}{\partial U}\frac{\partial U}{\partial X} + \frac{\partial z}{\partial V}\frac{\partial V}{\partial X}\right)\\
&= -\frac{1}{s^2}\frac{\partial s}{\partial z}\left(\frac{\partial z}{\partial U}\frac{\partial U}{\partial X} + \frac{\partial z}{\partial V}\frac{\partial V}{\partial X}\right)\\
&= -\frac{1}{s^2}(-e^{-z})\left(\frac{\partial z}{\partial U}\frac{\partial U}{\partial X} + \frac{\partial z}{\partial V}\frac{\partial V}{\partial X}\right)\\
&= \frac{1}{s^2}e^{-z}\left(\frac{\partial z}{\partial U}\frac{\partial U}{\partial X} + \frac{\partial z}{\partial V}\frac{\partial V}{\partial X}\right)\\
&= \frac{1}{s^2}(s-1)\left(\frac{\partial z}{\partial U}\frac{\partial U}{\partial X} + \frac{\partial z}{\partial V}\frac{\partial V}{\partial X}\right)\\
&= \frac{1}{s}(1-\frac{1}{s})\left(\frac{\partial z}{\partial U}\frac{\partial U}{\partial X} + \frac{\partial z}{\partial V}\frac{\partial V}{\partial X}\right)\\
&= \frac{1}{s}(1-\frac{1}{s})\left(\frac{\partial U^\top AV}{\partial U}\frac{\partial U}{\partial X} + \frac{\partial (U^\top AV)^\top}{\partial V}\frac{\partial V}{\partial X}\right)\\
&= f(1-f)(\frac{\partial U}{\partial X}\frac{\partial U^\top}{\partial U}AV + \frac{\partial V}{\partial X}\frac{\partial V^\top}{\partial V}A^\top U)\\
&= -\frac{1}{1+e^{-U^\top AV}}\frac{e^{-U^\top AV}}{1+e^{-U^\top AV}}(\frac{\partial U}{\partial X}AV + \frac{\partial V}{\partial X}A^\top U)
\end{aligned}$$

Sicne the first dimension is 1 not m, so this is a scalar over vector problem. What we need to pay attention is only to match the dimension to sure dim of the $\frac{\partial\sigma(U^\top AV)}{\partial X}$: $M \times 1$

3.2 Compute $\frac{\partial \sigma(U^\top AV)}{\partial X}$ when A is the function of X

Similar to 3.1

$$
\begin{aligned}
\frac{\partial \sigma(U^\top AV)}{\partial X} &= \frac{1}{s}(1-\frac{1}{s})\frac{\partial z}{\partial A}\frac{\partial A}{\partial X} \\
&= \frac{1}{s}(1-\frac{1}{s})\frac{\partial A}{\partial X}UV \\
&= -\frac{1}{1+e^{-U^\top AV}}\frac{e^{-U^\top AV}}{1+e^{-U^\top AV}}(\frac{\partial A}{\partial X})UV
\end{aligned}
$$

The matrix Dim of the $\frac{\partial A}{\partial X}$: $M \times K \times N$, and the dim of $\frac{\partial \sigma(U^\top AV)}{\partial X}$ is $M \times 1$

4 Given the softmax function $\sigma_M(k) = \frac{e^{W[k]X+b[k]}}{\sum_{k'=1}^{K} e^{W[k']X+b[k']}}$, compute $\frac{\partial \sigma_M(k)}{\partial X}$, where W is KxN matrix, X is a Nx1 vector, b is a kx1 vector, W[k] the kth of W and b[k] is the k's element of b.

let's F the matrix formation of $\sigma(k)$ for all k.

- $\sum_{k=1}^{K} \sigma(k) = \sum_{k=k'} \left[\frac{s[k]}{\sum_{k'=1}^{K} s[k']}\right] + \sum_{k \neq k'} \left[\frac{s[k']}{\sum_{k'=1}^{K} s[k']}\right]$
- $s[k] = e^{z[k]}$
- $Z = WX + b$

Here we can use a indicator matirx, which is a K X K diagonal matrix:

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

And another I', a K X K matrix with only the diagonal is not 1:

$$I' = \begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ \dots \\ 1 & 1 & \dots & 0 \end{bmatrix}$$

$$\begin{aligned}
\frac{\partial F}{\partial X} &= \frac{\partial F}{\partial S} \frac{\partial S}{\partial Z} \frac{\partial Z}{\partial X} \\
&= \Bigg[I \left[\frac{\partial s[k]}{\partial s[k]} \frac{\partial s[k]}{\partial z[k]} (\sum_{k'=1}^{K} s[k'])^{-1} + s[k] \frac{\partial (\sum_{k'=1}^{K} s[k'])^{-1}}{\partial s[k]} \frac{\partial s[k]}{\partial z[k]} \right] \\
&\quad + I' \left[\frac{\partial s[k']}{\partial s[k]} \frac{\partial s[k]}{\partial z[k]} (\sum_{k'=1}^{K} s[k'])^{-1} + s[k'] \frac{\partial (\sum_{k'=1}^{K} s[k'])^{-1}}{\partial s[k]} \frac{\partial s[k]}{\partial z[k]} \right] \Bigg] \frac{\partial Z}{\partial X} \\
&= \left[I \left[e^{z[k]} (\sum_{k'=1}^{K} s[k'])^{-1} + s[k](-(\sum_{k'=1}^{K} e^{z[k']})^{-2}) e^{z[k]} \right] + I' \left[0 + s[k'](-(\sum_{k'=1}^{K} e^{z[k']})^{-2}) e^{z[k]} \right] \right] \frac{\partial Z}{\partial X} \\
&= \left[I \left[\frac{e^{z[k]}}{\sum_{k'=1}^{K} e^{z[k']}} - \frac{e^{z[k]}}{\sum_{k'=1}^{K} e^{z[k']}} \frac{e^{z[k]}}{\sum_{k'=1}^{K} e^{z[k']}} \right] + I' \left[- \frac{e^{z[k']}}{\sum_{k'=1}^{K} e^{z[k']}} \frac{e^{z[k]}}{\sum_{k'=1}^{K} e^{z[k']}} \right] \right] \frac{\partial Z}{\partial X} \\
&= \left[I(\sigma[k](1 - \sigma[k])) + I'(-\sigma[k]\sigma[k']) \right] \frac{\partial Z}{\partial X} \\
&= \left(\begin{bmatrix} \sigma[1](1-\sigma[1]) & -\sigma[1]\sigma[2] & \dots & -\sigma[1]\sigma[k] \\ -\sigma[2]\sigma[1] & \sigma[2](1-\sigma[2]) & \dots & -\sigma[2]\sigma[k] \\ \dots \\ -\sigma[k]\sigma[1] & -\sigma[k]\sigma[2] & \dots & \sigma[k](1-\sigma[k]) \end{bmatrix} W \right)^{\top}
\end{aligned}$$

This is the transpose of a $[K \times K] \times [K \times N] = K \times N$ matrix, so it is a $N \times K$ matrix (from $X : N \times 1$ convention).