Ex 1

a) $f(x) = x^T \cdot Ax$ where $f: \mathbb{R}^N \to \mathbb{R}$, $x \in \mathbb{R}^N$ and $A \in \mathbb{R}^{N \times N}$ is possibly asymmetric.

$\Rightarrow$ ~~[illegible]~~ if symmetric: $A^T \cdot x = Ax$

if asymetric: $A \neq A^T$ $\Rightarrow$ we need to keep both terms in gradient

$$f(x) = x^T \cdot Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i A_{ij} x_j$$

$$\frac{df}{dx_k} = \frac{d}{dx_k} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} x_i A_{ij} x_j \right)$$

$$\left.\begin{array}{ll} \text{if } i=k; & x_k A_{kj} x_j \\ \text{if } j=k; & x_i A_{ik} x_k \end{array}\right\} \frac{df}{dx_k} = \sum_{j=1}^{n} A_{kj} x_j + \sum_{i=1}^{n} A_{ik} x_i$$

$\to (Ax)_k = \sum_{j=1}^{n} \cdot A_{kj} x_j$.

$\to (A^T x)_k = \sum_i A_{ik} x_i$ So $\nabla f(x) = Ax + A^T x$ //

$\nabla f(x) = Ax + A^T x$ $\to$ $\nabla^2 f(x) = ?$ $= \frac{d}{dx}(Ax + A^T \cdot x)$

$= \frac{d}{dx} Ax + \frac{d}{dx} A^T \cdot x$

$= A + A^T$

$\nabla^2 f(x) = A + A^T$ //

b) If A symmetric;

$\nabla f(x) = Ax + A^T x$ we found; if $A = A^T$

$\nabla f(x) = Ax + Ax = 2Ax$ //

$$\nabla^2 f(x) = A + \underbrace{A^T}_{A} = 2A$$ //

c) $f(x) = \log\left(\sum_{i=1}^{k} e^{a_i^T x + b_i}\right)$

$z_i := a_i^T x + b_i \rightarrow f(x) = \log\left(\sum_{i=1}^{k} e^{z_i}\right)$ //

$S := \sum_{i=1}^{k} e^{z_i} \rightarrow f(x) = \log(S)$ //

$\rightarrow \nabla f(x) = \frac{1}{S} \nabla S$, we need to find ∇S

$\nabla e^{z_i} = e^{z_i} \cdot \nabla z_i \Rightarrow \nabla (a_i^T x + b_i) = a_i$ so;

$= e^{z_i} \cdot a_i$ // plug in;

$\underset{\sim}{\nabla S} = \sum_{i=1}^{k} \nabla(e^{z_i}) = \underbrace{\sum_{i=1}^{k} e^{z_i} a_i}$; we found ∇S so plug in to $\nabla f(x) = \frac{1}{S} \nabla S$

$\nabla f(x) = \frac{1}{S} \sum_{i=1}^{k} e^{z_i} \cdot a_i$

$= \frac{1}{\sum_{i=1}^{k} e^{z_i}} \cdot \sum_{i=1}^{k} e^{z_i} \cdot a_i$

$= \sum_i \Big(\underbrace{\frac{e^{z_i}}{\sum_j e^{z_j}}}_{P} \Big) \cdot a_i$? looks like softmax

$p_i := \frac{e^{z_i}}{\underbrace{\sum_{j=1}^{k} e^{z_j}}} = \frac{e^{z_i}}{S}$ so; $\nabla f(x) = \sum_{i=1}^{k} p_i \cdot a_i$ //

we defined this as S

* $\nabla p_i = p_i (a_i - \nabla f(x))$ softmax derivative ?

$\nabla^2 f(x) = \cancel{\sum_{j=1}^{k} p_i (a_i - \nabla f(x))}$

$\rightarrow \mu := \nabla f(x) = \sum_{i=1}^{k} p_i a_i$ from above

$\nabla^2 f(x) = \sum_{i=1}^{k} \nabla p_i \cdot a_i^T$ since a_i is constant

$\rightarrow \mu := \nabla f(x) = \sum_{i=1}^{k} p_i a_i$ $\quad\rightarrow p_i = \frac{e^{z_i}}{S}$

$\rightarrow \nabla^2 f(x) = \sum_{i=1}^{k} \nabla p_i \cdot a_i^T$

$\nabla p_i = \nabla\left(\frac{e^{z_i}}{S}\right)$

$$\nabla p_i = \frac{e^{z_i} \cdot a_i \cdot S - e^{z_i} \cdot \sum_{j=1}^{k} e^{z_j} a_j}{S^2}$$

$$= \underbrace{\frac{e^{z_i}}{S}}_{\text{plug in } \rightarrow p_i} \left(a_i - \frac{1}{S} \sum_{j=1}^{k} e^{z_j} a_j\right)$$

$$= p_i \left(a_i - \underbrace{\sum_{j=1}^{k} p_j \cdot a_j}_{\nabla f(x) = \mu \text{ plug in}}\right)$$

$$\nabla p_i = \underline{p_i (a_i - \mu)}$$

$$\nabla^2 f(x) = \sum_{i=1}^{k} \nabla p_i \cdot a_i^T$$

plug in $p_i (a_i - \mu)$

$$\nabla^2 f(x) = \sum_{i=1}^{k} p_i (a_i - \mu) \cdot a_i^T \quad //$$

EX2 $\rightarrow$ a) $tr(A^T X) \leq \|A\|_F \|X\|_F$

trace $tr(A^T X)$: sum of all diag. elements

$\|X\|_F := \sqrt{tr(X^T \cdot X)}$ Forbenius form.

Cauchy-Schwarz inequality from lecture notes page 180; (A3: Norms) states for all $x, y \in E$

$\rightarrow |\langle x, y \rangle| \leq \|x\| \cdot \|y\|$

$tr(A^T \cdot X) = \langle A, X \rangle_F = \sum_{i,j} A_{ij} X_{ij}$ therefore.

$\langle A, X \rangle_F \leq \|A\|_F \cdot \|X\|_F$

b) $\|A^T.x + x^T A\|_F^2 \leq 4\|A\|_F^2 \, \|x\|_F^2$

→ by def; $\|x\|_F^2 = tr(x^T.x)$ so;

$\|A^T.x + x^T.A\|_F^2 := tr((A^T.x + x^T.A)^T.(A^T.x + x^T.A))$

rule: $(AB)^T = B^T.A^T$ so; $(A^T.x + x^T.A)^T = (A^T.x)^T + (x^T.A)^T$

$= x^T.A + A^T.x$ plug in

$tr(\underline{(x^T.A + A^T.x)}.\underline{(A^T.x + x^T.A)})$

identical so;

$= tr((x^T.A + A^T.x)^2)$

→ Forbenius norm inequality: $\|B+C\|_F^2 \leq 2\|B\|_F^2 + 2\|C\|_F^2$

let's say $B = x^T.A$ and $C = A^T.x$

$\|A^T.x + x^T.A\|_F^2 \leq 2\|A^T.x\|_F^2 + 2\|x^T.A\|_F^2$

→ Forbenius norm satisfies $\|AB\|_F \leq \|A\|_F . \|B\|_F$

So; $\|A^T.x\|_F \leq \|A^T\|_F . \|x\|_F$ and

$\|x^T.A\|_F \leq \|x^T\|_F . \|A\|_F$

That's also true if we square both sides.

$\|A^T.x\|_F^2 \leq \|A^T\|_F^2 . \|x\|_F^2$

→ def. $\|A^T\|_F = \|A\|_F$ so if we plug in. Therefore;

$\|A^T.x\|_F^2 \leq \|A\|_F^2 . \|x\|_F^2$ //

$\|A^T.x + x^T.A\|_F^2 \leq 2.\|A^T.x\|_F^2 + 2\|x^T.A\|_F^2$

$\leq 2\|A\|_F^2\|x\|_F^2 + 2\|x\|_F^2\|A\|_F^2$

$= 4\|A\|_F^2 . \|x\|_F^2$ //

Ex3 $\quad C = \{ X \in S_+(N) : tr(X) = 1 \}$ where;

positive $\quad S_+(N) \subset \mathbb{R}^{N \times N}$ is the set of symmetric semi-definite mat.

$tr(X) := \sum_{i=1}^{N} X_{i,i}$ is the trace of X

show C is non-empty and convex set

→ $S_+(N)$ example $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ → symmetric
→ all eigen vals are pos.

→ $tr(X)$: sum of diagonal $tr\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} = 1+3 = 4$ //

we must show $tr(X_0) = 1$?

Let $X \in \mathbb{R}^{N \times N}$ be $X = diag(1, 0 \dots)$

then $tr(X) = 1 + 0 + \dots + 0 = 1$ ✓

✓ X is symmetric ⟹ $X^T = X$ $\quad X = \begin{bmatrix} x_1 & 0 & \dots & 0 \\ 0 & x_2 & \dots & 0 \\ 0 & 0 & x_3 \dots & 0 \\ 0 & \dots & & x_n \end{bmatrix}$

$X_{ij} = 0$ for all $i \neq j$

$(X^T)_{ij} = X_{ij} \rightarrow X^T = X$ $\quad$ diagn. matrix is symmetric.

$X = diag(x_1, x_2 \dots x_n)$, $x_{ij} = \begin{cases} x_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$

✓ $X \in C$ so C is non-empty

Convexity: Let $X_1, X_2 \in C$ and $\theta \in [0,1]$

$X_\theta := \theta X_1 + (1-\theta) X_2$

① $X_\theta \in S_+(N)$ $\quad$ ② $tr(X_\theta) = 1$:

$tr(X_\theta) = \theta \cdot tr(X_1) + (1-\theta) tr(X_2)$

$= \theta \cdot 1 + (1-\theta) \cdot 1$

$= 1$ //

so $X_\theta \in C$