Reworked solutions STAT5810 HW 1

Nicole Fleming
A02079019

① a) $A = (B^T)^{-1} \longrightarrow AB^T = I$

so $(A \cdot B^T)^T = I^T \longrightarrow B \cdot A^T = I$

In the case above B is the inverse of A^T and A is the inverse of B^T.

So $A^T = B^{-1}$ transpose both sides;

$$(A^T)^T = (B^{-1})^T$$

$$A = (B^{-1})^T$$

$$\underline{\underline{A = (B^T)^{-1} = (B^{-1})^T}}$$

---

① b) $A^{-1}B^{-1} = (BA)^{-1}$

Where $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{m \times n}$ OR $A, B \in \mathbb{R}^{n \times n}$

Rectangle

$$A^{-1}_{n \times m} B^{-1}_{m \times n} = (B_{m \times n} A_{n \times m})^{-1}$$

$$(B_{m \times n} A_{n \times m}) \cdot [A^{-1}_{n \times m} B^{-1}_{m \times n}] = (B_{m \times n} A_{n \times m})^{-1} (B_{m \times n} A_{n \times m})$$

$$AA^{-1} \cdot BB^{-1} = I$$

$$I_{n \times m} \cdot I_{m \times n}$$

$$\underline{\underline{I_{n \times n} = I_{n \times n}}}$$

SQUARE

$$A^{-1}B^{-1} = (BA)^{-1}$$

$$(BA)A^{-1}B^{-1} = (BA)^{-1}BA$$

$$AA^{-1}BB^{-1} = I$$

$$I \cdot I = I$$

$$\underline{I = I}$$

---

③ Take the definition of the trace:

$$(AB) = [AB]_{ij} = \sum_k^n a_{ik} b_{kj}$$

where AB is a square, $j = i$

so $$(AB) = [AB]_{ii} = \sum_k^n a_{ik} b_{ki}$$

But here

$$\underline{\underline{tr(AB^T) = \sum_i^n \left(\sum_k^n a_{ik} b_{ik} \right)}}$$

---

④ $\boxed{5.5}$

for each function individually:

1) $f_1(x) = \sin(x_1)\cos(x_2)$

$f_1 : \mathbb{R}^2 \longrightarrow \mathbb{R}$

$\underline{\underline{J_{f_1} \in \mathbb{R}^{1\times 2}}}$

$$\underline{\underline{J_f = [\cos(x_1)\cos(x_2), -\sin(x_1)\sin(x_2)]}}$$

2) $f_2(x) = x^T y$

$f_2 : \mathbb{R}^n \longrightarrow \mathbb{R}$

$\underline{\underline{J_{f_2} : \mathbb{R}^{1\times n}}}$ (row vector)

$f_2(x) = x^T y = \sum_i x_i y_i$

$\frac{df}{dx_1} = y_1 \quad \frac{df}{dx_2} = y_2$

$$\underline{\underline{J_{f_2} = [y_1, y_2, y_3, \cdots, y_n] = y^T}}$$

3) $f_3(x) = xx^T$

$f_3 : \mathbb{R}^n \to \mathbb{R}^{n\times n}$

$\underline{\underline{J_{f_3} : \mathbb{R}^{(n\times n)\times n}}}$

$$\frac{d}{dx}(xx^T) = \frac{d}{dx}\left(\begin{bmatrix} x \end{bmatrix} \begin{bmatrix} & x^T & \end{bmatrix} \right) = \left(\begin{bmatrix} x_1x_1 & x_1x_2 & \cdots & x_1x_n \\ \vdots & x_2x_2 & & \vdots \\ & & \ddots & \\ x_nx_1 & \cdots & & x_nx_n \end{bmatrix} \right) \frac{d}{dx}$$

$$\frac{d}{dx_1} \rightarrow \begin{bmatrix} 2x_1 & x_2 & \cdots & x_n \\ x_2 & & & \\ \vdots & & 0 & \\ x_n & & & \end{bmatrix} \quad \frac{d}{dx_2} = \begin{bmatrix} 0 & x_1 & 0 & \cdots & 0 \\ x_1 & 2x_2 & x_3 & \cdots & x_n \\ 0 & x_3 & & & \\ \vdots & \vdots & & 0 & \\ 0 & x_n & & & \end{bmatrix} \quad \frac{d}{dx_3} = \begin{bmatrix} 0 & 0 & x_1 & 0 & \cdots & 0 \\ 0 & 0 & x_2 & 0 & \cdots & 0 \\ x_1 & x_2 & 2x_3 & x_4 & \cdots & x_n \\ 0 & 0 & x_4 & & & \\ \vdots & \vdots & \vdots & & 0 & \\ 0 & 0 & x_n & & & \end{bmatrix}$$

$$\underline{\underline{\frac{d[xx^T]_{ij}}{dx_k} = \begin{cases} 2x_i, & i=j=k \quad (x_{ii})_k \\ 0, & i \neq j \neq k \\ [x_{i-k}, x_{i-(k-1)} \cdots x_{i-1}], & i=k \neq j \quad (x_{ij} \neq x_{ii})_k \\ [x_{j-k}, x_{j-(k-1)} \cdots x_{j-1}]^T, & j=k \neq i \quad (x_{ij} \neq x_{ii})_k \end{cases}}}$$

Example: for $\frac{d[xx^T]_{ij}}{dx_5} = \begin{cases} 2x_5 & ; i=j=k \\ 0 & ; i \neq j \neq k \\ x_1, x_2, x_3, x_4 & ; i=k \neq j \\ x_1, x_2, x_3, x_4 & ; j=k \neq i \end{cases}$

$\boxed{5.6}$

1) $f(t) = \sin(\log(t^T t)) \qquad t \in \mathbb{R}^D$

$f_t : \mathbb{R}^D \longrightarrow \mathbb{R}^1$

$J_f: \mathbb{R}^{1\times D}$

$$\frac{d}{dt}(t^{\top}t) = 2t^{\top}$$

$$\underbrace{[t_1 t_2 \cdots t_D]}_{t^{\top}} \underbrace{\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_D \end{bmatrix}}_{t} = [t_1 \cdot t_1, t_2 \cdot t_2, \ldots, t \cdot t_D]$$

$$\frac{d}{dt}\left(\sin\left(\log\left(t^{\top}t\right)\right)\right)$$

$$\sin(\ln(t^{\top}t))$$

$$\sin(\ln(t^{\top}t)) \qquad u = \ln(t^{\top}t) \quad du = \frac{1}{t^{\top}t} \qquad w = t^{\top}t \quad dw = 2t^{\top}$$

$$\frac{\cos(\ln(t^{\top}t) \cdot \frac{d}{dt}(\ln(t^{\top}t)))}{\ln(t^{\top}t)}$$

$$\Longrightarrow \frac{\cos\left(\ln(t^{\top}t) \cdot \frac{2t^{\top}}{t^{\top}t}\right)}{\ln(t^{\top}t)} \quad (\in \mathbb{R}^{1\times D})$$

2) $f(X): tr(AXB)$

$$f_X: \mathbb{R}^{\underbrace{D\times E\times F\times D}} \longrightarrow \mathbb{R}^{D\times E}$$

$$\text{so} \quad g'(X) = tr(AXB)' = \underline{tr(AB)} \in \mathbb{R}^{D\times E}$$

$$\mathbb{R}^{D\times E} \ \mathbb{R}^{F\times D}$$

$$\hookrightarrow \left(\sum_E \left(\sum_D a_{iD} b_{Dj}\right)\right)$$

$\boxed{5.7}$

a) $f(z) = \log(1+\vec{z})$, $\vec{z} = g(\vec{x}) = \vec{x}^{\top}\vec{x}$, $x \in \mathbb{R}^D$

$f(g(\vec{x}))$ $\quad f: \mathbb{R}^D \longrightarrow \mathbb{R}$

$\quad g: \mathbb{R}^D \longrightarrow \mathbb{R}$

$$\vec{x}^{\top}\vec{x} = [x_1\ x_2\ x_3 \cdots x_D] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_D \end{bmatrix} = [x_1^2, x_2^2 \cdots x_n^2]$$

$J_f \in \mathbb{R}^{1\times D}$

$J_g \in \mathbb{R}^{1\times D}$ $\qquad g' = 2\vec{x}^{\top}$

$\underline{J_{TOT} \in \mathbb{R}^{D\times 1}}$

$$f' = \frac{1}{1+g(\vec{x})} \cdot g'(\vec{x})$$

$$f' = \frac{2x^{\top}}{1+x^{\top}x}$$

b) $f(z) = \sin(\vec{z})$, $\vec{z} = A\vec{x} + \vec{b}$, $A \in \mathbb{R}^{E\times D}$, $\vec{b} \in \mathbb{R}^E$, $\vec{x} \in \mathbb{R}^D$

$g(x) = \vec{z}$.

$J_g \in \mathbb{R}^{E\times D}$

$g : \mathbb{R}^D \longrightarrow \mathbb{R}^E$ $\qquad J_g$

$f : \mathbb{R}^E \longrightarrow \mathbb{R}^E$ $\qquad J_f \in \mathbb{R}^{E \times E}$

$J_{TOT} \in \mathbb{R}^{E \times D}$

$J_g = A$

$$J_f = \begin{bmatrix} \cos(\vec{z}_1) & & & 0 \\ & \cos(\vec{z}_2) & & \\ 0 & & \ddots & \\ & & & \cos(\vec{z}_n) \end{bmatrix} = \operatorname{diag}(\cos(\vec{z}))$$

$$\underline{\underline{J_{TOT} = A\left[\operatorname{diag}\left(\cos(A\vec{x} + \vec{b})\right)\right] \in \mathbb{R}^{E \times D}}}$$

$\boxed{5.8}$

a) $f(z) = e^{-1/2 z}$, $S \in \mathbb{R}^{D \times D}$, $x \in \mathbb{R}^D$, $m \in \mathbb{R}^D$

$z = g(y) = y^T S^{-1} y$

$y = h(x) = x - m$

$h : \mathbb{R}^D \longrightarrow \mathbb{R}^D$ $\qquad J_h : \mathbb{R}^{D \times D}$

$g : \mathbb{R}^D \longrightarrow \mathbb{R}^D$ $\qquad J_g : \mathbb{R}^{1 \times D}$

$f : \mathbb{R} \longrightarrow \mathbb{R}$ $\qquad J_f : \mathbb{R}^{1 \times 1}$

$J_h = I_{(D \times D)}$

$J_g = y^T\left(S^{-1} + (S^{-1})^T\right)_{(D \times D)}$

$J_f = \frac{1}{2} e^{-1/2 z}$

$$\underline{\underline{J_{TOT} = -\frac{1}{2} e^{-\frac{1}{2}\left((x-m)^T S^{-1} (x-m)\right)} (x-m)^T \left(S^{-1} + (S^{-1})^T\right)^T}}$$

b) $f(x) = tr(x x^T + \sigma^2 I)$, $\qquad x \in \mathbb{R}^D$

$xx^T \quad \mathbb{R}^D \longrightarrow \mathbb{R}^{D \times D}$

$f : \mathbb{R}^D \longrightarrow \mathbb{R}$

$f: \mathbb{R}^D \rightarrow \mathbb{R}$

$\sigma^2 I$ will be $\mathbb{R}^{D\times D}$ since adding.

$$\frac{d}{dx}\left(xx^T\right) = \frac{d}{dx}\left(\begin{bmatrix} x \end{bmatrix}\begin{bmatrix} & x^T & \end{bmatrix}\right)$$

$$= \frac{d}{dx}\left(\begin{bmatrix} x_1x_1 & x_1x_2 & \cdots & x_1x_n \\ x_2x_1 & x_2x_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ x_nx_1 & \cdots & \cdots & x_nx_n \end{bmatrix}\right)$$

(Refer to problem 4 (5.5 $f_3(x)$))

$$\frac{d[xx^T]_{ij}}{dx_k} = \begin{cases} 2x_i & i=j=k \\ 0 & i\neq j\neq k \\ x_{i-k,j}, x_{i-k-1\,j} \cdots & i=k\neq j \\ x_{i\,k-1}, x_{i\,k-2} \cdots & k=j\neq i \end{cases}$$

$$\frac{d}{dx}\left[tr\left(xx^T + \sigma^2 \cdot I\right)\right]$$

$$= tr(\underset{C}{xx^T} + \underset{0}{\sigma \cdot I}) \cdot \frac{d[xx^T]_{ij}}{dx_k}$$

$$\Rightarrow \sum_i^D 2\left(\frac{d[xx^T]_{ii}}{dx_k}\right)$$

$$\Rightarrow \sum_i^D 2(2x_{ii})$$

$$\Rightarrow \underline{\underline{\sum_i^D 4x_{ii}}}$$

---

5 $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{x} \in \mathbb{R}^2$ $\quad \vec{y}(\vec{x}) : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ $\quad \vec{y} = \begin{bmatrix} x_1x_2 \\ x_1+x_2 \\ 3x_2 \end{bmatrix}$

and $\vec{w}(\vec{y}) = \mathbb{R}^3 \rightarrow \mathbb{R}^2$, $\vec{w} = \begin{bmatrix} y_1+y_2 \\ y_1y_3+y_2 \end{bmatrix}$ find $\frac{\partial w}{\partial x}$.

$y : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ $\quad J_y : \mathbb{R}^{2\times 3}$

$w : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ $\quad J_w : \mathbb{R}^{3\times 2}$

$J_{Tot} : \mathbb{R}^{2\times 2}$

$$\frac{\partial y}{\partial x} = \begin{bmatrix} x_2 & x_1 \\ 1 & 1 \\ 0 & 3 \end{bmatrix}$$

$$\frac{\partial w}{\partial y} = \begin{bmatrix} 1 & 1 & 0 \\ y_3 & y_2 & y_1 \end{bmatrix}$$

$$\partial w = \frac{\partial w}{\partial y} \cdot \frac{\partial y}{\partial x} = \begin{bmatrix} 1 & 1 & 0 \\ y_3 & y_2 & y_1 \end{bmatrix}\begin{bmatrix} x_2 & x_1 \\ 1 & 1 \end{bmatrix}$$

$$\frac{\partial x}{} \quad \frac{\partial y}{} \quad \begin{bmatrix} & \\ 0 & 3 \end{bmatrix}$$ (part of prior page)

$$\frac{\partial w}{\partial x} = \begin{bmatrix} x_2+1 & y_3x_2 + y_2 \\ x_1+1 & y_3x_1 + y_2 + 3y_1 \end{bmatrix}$$

$$\frac{\partial w}{\partial x} = \begin{bmatrix} x_2+1 & 3x_2^2 + (x_1+x_2) \\ x_1+1 & 3x_2x_1 + (x_1+x_2) + 3x_1x_2 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

---

(6)

a) $\frac{\partial(\vec{x}^T A)}{\partial x}$ where $\vec{x} \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times m}$ where n is not necessarily equal to m.

$$\frac{\partial}{\partial x}\left(\begin{bmatrix} & x^T & \end{bmatrix} \begin{bmatrix} A \end{bmatrix} \right)$$

will result in a matrix $\in \mathbb{R}^{n\times m}$, and A is not a function of x.

$f(x) = \vec{x}^T A$

$$\left.\begin{matrix} x \in \mathbb{R}^n \\ A \in \mathbb{R}^{n\times m} \\ f \in \mathbb{R}^{n\times m} \end{matrix}\right\}$$

The derivative rule with vectors is:

$$\frac{dx^T a}{dx} = \frac{da^T x}{\partial x} = a^T$$

where $a^T x = x^T a$ since this goes to a scalar.

So for a matrix,

$$\frac{d(x^T A)}{dx} = \frac{d}{dx}\left(\begin{bmatrix} x_1 a_{11}, & x_2 a_{12} & \cdots \\ x_1 a_{21} & x_2 a_{22} & \\ x_1 a_{31} & & \ddots \end{bmatrix} \right)$$

$$\Rightarrow \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots \\ a_{21} & a_{22} & & \\ a_{31} & & \ddots & \\ \vdots & & & \end{bmatrix} = \underline{\underline{A}}$$

Since one component of x is applied to each column of A, when the derivative is taken of x the result is A, since $x^T A$ is a matrix, the derivative will be a matrix. Since the derivative is taken with respect to x, the result will be a matrix.

---

b) $\dfrac{\partial tr(X)}{\partial X}$ where X is some matrix that is square. (the trace is only valid on square matrices).

$$\frac{\partial tr(X)}{\partial X} = \frac{\partial}{\partial x}\Big(tr(X)\Big) = tr(X + \partial X)$$

$$\Rightarrow \partial tr(X) = tr(X + \partial X) - tr(X)$$

$$= tr(X_{ij} + \partial X_{ij})$$

$$= \frac{\partial}{\partial X_{mn}} \sum_{kl} X_{kl} \rightarrow \frac{\partial}{\partial X_{mn}} \sum_{kl} \partial X_{km} \, \partial X_{nl} \Rightarrow \Sigma 1.$$

which tells us for the $\dfrac{\partial tr(X)}{\partial X}$ the s lution is the identity matrix of the same dimensions as X.

---

c) $\dfrac{d\, tr(XX^T)}{dX}$, $X \in \mathbb{R}^{n \times n}$ call $XX^T = a$

$$X : \mathbb{R}^{n \times n}$$
$$a : \mathbb{R}^{n \times n} \longrightarrow \mathbb{R}^{(n \times n)n}$$
$$a' : \mathbb{R}^{(n \times n)n} \longrightarrow \mathbb{R}^{n \times n}$$

The $tr(XX^T)$ results in a matrix, and for the trace to be taken, the matrix will be square for the trace to be taken. For the matrix $X \in \mathbb{R}^{n \times n}$ the diagonal elements will remain

if $X \in \mathbb{R}^{n \times m}$ $X^T \in \mathbb{R}^{m \times n}$

the same : $X = \begin{bmatrix} n_{11} & n_{12} & \cdots \\ n_{21} & n_{22} & \\ \vdots & & \ddots \end{bmatrix}$ $X^T = \begin{bmatrix} n_{11} & n_{21} & \cdots \\ n_{12} & n_{22} & \\ \vdots & & \ddots \end{bmatrix}$ if $X \in \mathbb{R}^{n \times m}$, $X^T \in \mathbb{R}^{m \times n}$

but the values on the diagonals will remain the same, and the matrix $Y = XX^T$ will be square s.t. $Y \in \mathbb{R}^{n \times n}$.

With this in mind, in the $X \in \mathbb{R}^{n \times n}$ case

XX^T and X^TX will result in the same operation inside the $tr(\cdot)$ operation.

The derivative is as follows:

$$\frac{d}{dX}\left(tr(XX^T)\right) = \frac{d}{dX}\left(\sum_{i=1}^{n} X_{ii} \cdot X^T_{ii}\right) \quad \text{where } X^T_{ii} = X_{ii}$$

$$= \underline{\underline{2X}}$$

OR by using Taylor Series for my brain with Rectangle matrix X:

$$\frac{df}{dx} \Rightarrow \frac{d}{dx}\left(tr(X \cdot X^T)\right) = f(x) + tr\left(XX^T\right) + tr(X^TX) + f'(\xi) \cdots$$

and $X \cdot X^T$ and $X^T \cdot X$ are equivalant even with Rectangular matrices, since the diagonal is not changed when the transpose is taken. Thus

$$f(x) + tr(2XX^T) + \text{error terms}$$

so $df(x) = \underline{\underline{2X}}$

---

d) $\frac{\partial}{\partial x}\left(\sqrt{(\vec{x}-\vec{a})^T(\vec{x}-\vec{a})}\right)$ $\vec{x}, \vec{a} \in \mathbb{R}^D$. I will denote $\vec{x} = x$ and $\vec{a} = a$.

$\frac{\partial}{\partial x}\left(\left((x-a)^T(x-a)\right)^{\frac{1}{2}}\right)$

Note: $\vec{a}$ is not a vector of $\vec{x}$, thus with the derivitive $\frac{d}{dx}(\vec{a}) = \vec{0}$, as a act as constants.

$u = \vec{x} - \vec{a}$ $\quad du = x - a$

$v = \vec{u}^T\vec{u}$ $\quad dv = 2 \cdot \frac{\partial u}{\partial x} = 2\vec{x} - 2\vec{a}$

$$\frac{\partial}{\partial x} = \left((u^Tu)^{1/2}\right)$$

$$= \frac{1}{2}(u^Tu)^{-1/2} \cdot du$$

$$\Rightarrow \frac{1}{2}\left((x-a)^T(x-a)\right)^{-1/2} \cdot (2\vec{x} - 2\vec{a})$$

$$= \boxed{\frac{2\vec{x} - 2\vec{a}}{2\sqrt{(\vec{x}-\vec{a})^T(\vec{x}-\vec{a})}}}$$
