Manifold Learning and Sparse Representation: Assignment 2

ZHANG Yuan, 1601111332

Exercise 24.

1. $\|\cdot\|_p^* = \|\cdot\|_q$, where $p^{-1}+q^{-1}=1$.

Proof. By definition, $\|\mathbf{x}\|_p^* = \sup\{\langle \mathbf{x}, \mathbf{y}\rangle | \|\mathbf{y}\|_p \leq 1\}$ for any vector $\mathbf{x}$. Thanks to Holder's inequality,

$$\langle \mathbf{x}, \mathbf{y}\rangle \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q \leq \|\mathbf{x}\|_p,$$

where $p^{-1}+q^{-1}=1$ and it becomes an equality if and only if $\forall i : (\sum_j x_j^q) y_i^p = x_i^q$ such that the vector $\mathbf{y}$ always exists for any given vector $\mathbf{x}$. Thus, $\|\mathbf{x}\|_p^* = \sup\{\langle \mathbf{x}, \mathbf{y}\rangle | \|\mathbf{y}\|_p \leq 1\} = \|\mathbf{x}\|_q$. □

2. $\|\cdot\|_2^* = \|\cdot\|_*$.

Proof. When $\|\mathbf{Y}\|_2 \leq 1$ which implies $\max_i \sigma_i(Y) \leq 1$, we use the von Neumann theorem

$$\langle \mathbf{X}, \mathbf{Y}\rangle \leq \sum_i^{\min(m,n)} \sigma_i(\mathbf{X})\sigma_i(\mathbf{Y}) \leq \sum_i^{\min(m,n)} \sigma_i(\mathbf{X}) = \|\mathbf{X}\|_*.$$

On the other hand, when $Y = I$,

$$\langle \mathbf{X}, \mathbf{Y}\rangle = tr(\mathbf{X}) = \sum_i^{\min(m,n)} \sigma_i(\mathbf{X}) = \|\mathbf{X}\|_*.$$

Thus, by definition, $\|X\|_2^* = \sup\{\langle \mathbf{X}, \mathbf{Y}\rangle | \|\mathbf{Y}\|_2 \leq 1\} = \|\mathbf{X}\|_*$. □

3. $\|\cdot\|_F^* = \|\cdot\|_F$.

Proof. When $\|\mathbf{Y}\|_F \leq 1$, we have the following thanks to the Cauchy-Schwarz inequality,

$$\langle \mathbf{X}, \mathbf{Y}\rangle = \sum_{i,j} x_{ij}y_{ij} \leq \sqrt{\sum_{ij} x_{ij}^2 \sum_{ij} y_{ij}^2} = \|X\|_F \|Y\|_F = \|X\|_F$$

where the equality holds when $\forall i,j : y_{ij} = x_{ij}/\sum_j x_{ij}$, (we define $y_{ij}=0$ when $\sum_j x_{ij} = 0$). Therefore, $\|X\|_F^* = \sup\{\langle \mathbf{X}, \mathbf{Y}\rangle | \|\mathbf{Y}\|_F \leq 1\} = \|X\|_F$. □

Exercise 28

1. $\frac{\partial \mathbf{XY}}{\partial t} = \frac{\partial \mathbf{X}}{\partial t}\mathbf{Y} + \frac{\partial \mathbf{Y}}{\partial t}\mathbf{X}$.

Proof.

$$
\begin{aligned}
(\frac{\partial \mathbf{XY}}{\partial t})_{ij} = \frac{\partial (\mathbf{XY})_{ij}}{\partial t} = \frac{\partial \sum_k x_{ik} y_{kj}}{\partial t} &= \sum_k \frac{\partial x_{ik} y_{kj}}{\partial t} \\
&= \sum_k \frac{\partial x_{ik}}{\partial t} y_{kj} + \frac{\partial x_{ik} y_{kj}}{\partial t} x_{ik} \\
&= \sum_k (\frac{\partial \mathbf{X}}{\partial t})_{ik} \cdot y_{kj} + \sum_k x_{ik} \cdot (\frac{\partial \mathbf{Y}}{\partial t})_{kj} \\
&= (\frac{\partial \mathbf{X}}{\partial t}\mathbf{Y})_{ij} + (\mathbf{X}\frac{\partial \mathbf{Y}}{\partial t})_{ij} \\
&= (\frac{\partial \mathbf{X}}{\partial t}\mathbf{Y} + \mathbf{X}\frac{\partial \mathbf{Y}}{\partial t})_{ij}
\end{aligned}
$$

□

2. $\frac{\partial \mathbf{a^T x}}{\partial \mathbf{x}} = \mathbf{a}$.

Proof.

$$
(\frac{\partial \mathbf{a^T x}}{\partial \mathbf{x}})_i = \frac{\partial \sum_j a_j x_j}{\partial x_i} = a_i
$$

□

3. $\frac{\partial \mathbf{x^T A x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A^T})\mathbf{x}$

Proof.

$$
\begin{aligned}
(\frac{\partial \mathbf{x^T A x}}{\partial \mathbf{x}})_i &= \frac{\partial \sum_j (a_{ij} x_i x_j + a_{ji} x_j x_i)}{\partial x_i} = \sum_j a_{ij} x_j + \sum_j a_{ji} x_j \\
&= (\mathbf{Ax})_i + (\mathbf{A^T x})_i = ((\mathbf{A} + \mathbf{A^T})\mathbf{x})_i
\end{aligned}
$$

□

4. calculate $\frac{\partial \mathbf{X}}{\partial \mathbf{X}}$, assuming $\mathbf{X} \in \mathbb{R}^{m \times n}$.

$$
\begin{aligned}
\frac{\partial \mathbf{X}}{\partial \mathbf{X}} &= \begin{pmatrix} \frac{\partial \mathbf{X}}{\partial x_{11}} & \cdots & \frac{\partial \mathbf{X}}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{X}}{\partial x_{m1}} & \cdots & \frac{\partial \mathbf{X}}{\partial x_{mn}} \end{pmatrix} = \begin{pmatrix} \mathbf{e_1 e_1'^T} & \cdots & \mathbf{e_1 e_n'^T} \\ \vdots & \ddots & \vdots \\ \mathbf{e_m e_1'^T} & \cdots & \mathbf{e_m e_n'^T} \end{pmatrix} = \begin{pmatrix} \mathbf{e_1}(\mathbf{e_1'^T}, \cdots, \mathbf{e_n'^T}) \\ \vdots \\ \mathbf{e_m}(\mathbf{e_1'^T}, \cdots, \mathbf{e_n'^T}) \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{e_1} \\ \vdots \\ \mathbf{e_m} \end{pmatrix} \begin{pmatrix} \mathbf{e_1'^T} & \cdots & \mathbf{e_n'^T} \end{pmatrix} = vec(\mathbf{I}_{m\times m}) \cdot vec(\mathbf{I}_{n \times n})^T
\end{aligned}
$$

where indicator vectors $\mathbf{e}_i \in \mathbb{R}^m$ and $\mathbf{e}_i \in \mathbb{R}^n$.

Exercise 29.

By definition, for any $x \in \mathbb{R}$ and $\mathbf{Y} \in \mathbb{R}^{3\times 3}$

$$\begin{aligned}\langle \mathcal{A}^*(x), \mathbf{Y}\rangle = \langle x, \mathcal{A}(\mathbf{Y})\rangle &= \langle x, Y_{11} + Y_{12} - Y_{3,1} + 2Y_{33}\rangle \\ &= x \cdot (Y_{11} + Y_{12} - Y_{3,1} + 2Y_{33}) \\ &= \langle \begin{pmatrix} x & x & 0 \\ 0 & 0 & 0 \\ -x & 0 & 2x \end{pmatrix}, Y\rangle.\end{aligned}$$

Therefore, $\mathcal{A}^*(x) = \begin{pmatrix} x & x & 0 \\ 0 & 0 & 0 \\ -x & 0 & 2x \end{pmatrix}$.

Exercise 33.

Proof. Suppose $K \in \mathbb{R}^{n\times n}$ and $Y \in \mathbb{R}^{r\times n}$. We denote $t = -\min_i(\lambda_i(\mathbf{K}))$ and then $\mathbf{K} + t\mathbf{I}$ is positive semi-definite with equivalent eigenvalues and singular values, i.e., $\lambda_i = \sigma_i$.

$$\begin{aligned} tr(\mathbf{YKY^T}) &= tr(\mathbf{Y}(\mathbf{K} + t\mathbf{I} - t\mathbf{I})\mathbf{Y^T}) = tr(\mathbf{Y}(\mathbf{K} + t\mathbf{I})Y^T) - tr(t\mathbf{I}) \\ &= tr(\mathbf{Y^T Y}(\mathbf{K} + t\mathbf{I})) - rt, \end{aligned}$$

According to the von Neumann trace inequality,

$$\sum_{i=1}^{n} \sigma_i(\mathbf{Y^T Y})\sigma_{n-i+1}(\mathbf{K} + t\mathbf{I}) \le tr(\mathbf{Y^T Y}(\mathbf{K} + t\mathbf{I})) \le \sum_{i=1}^{n} \sigma_i(\mathbf{Y^T Y})\sigma_i(\mathbf{K} + t\mathbf{I})$$

$$\sum_{i=1}^{r} \sigma_i(\mathbf{Y Y^T})\sigma_{n-i+1}(\mathbf{K} + t\mathbf{I}) \le tr(\mathbf{Y^T Y}(\mathbf{K} + t\mathbf{I})) \le \sum_{i=1}^{r} \sigma_i(\mathbf{Y Y^T})\sigma_i(\mathbf{K} + t\mathbf{I})$$

$$\sum_{i=1}^{r} \sigma_i(\mathbf{I})\lambda_{n-i+1}(\mathbf{K} + t\mathbf{I}) \le tr(\mathbf{Y^T Y}(\mathbf{K} + t\mathbf{I})) \le \sum_{i=1}^{r} \sigma_i(\mathbf{I})\lambda_i(\mathbf{K} + t\mathbf{I})$$

$$\sum_{i=1}^{r} \lambda_{n-i+1}(\mathbf{K}) + rt \le tr(\mathbf{Y^T Y}(\mathbf{K} + t\mathbf{I})) \le \sum_{i=1}^{r} \lambda_i(\mathbf{K}) + rt$$

$$\sum_{i=1}^{r} \lambda_{n-i+1}(\mathbf{K}) \le tr(\mathbf{Y^T Y}(\mathbf{K} + t\mathbf{I})) - rt \le \sum_{i=1}^{r} \lambda_i(\mathbf{K})$$

$$\sum_{i=1}^{r} \lambda_{n-i+1}(\mathbf{K}) \le tr(\mathbf{YKY^T}) \le \sum_{i=1}^{r} \lambda_i(\mathbf{K})$$

We note that the first and second equalities hold iff. Y are the tailing and leading eigenvectors of $\mathbf{K}$ associated with the largest and smallest r eigenvalues. Therefore, the solutions to $\min_{\mathbf{Y}} tr(\mathbf{YKY^T})$ and $\max_{\mathbf{Y}} tr(\mathbf{YKY^T})$ correspond to the tailing and leading eigenvectors of $\mathbf{K}$, respectively. □