

# Manifold Learning Homework 2

安捷 1601210097

2017-03-01

习题 (24). *Proof 1.*

$$\|\cdot\|_p^* = \|\cdot\|_q$$

由于 Holder inequality

**Theorem.** 对于任意的  $p > 1, q > 1$ , 如果满足

$$\frac{1}{p} + \frac{1}{q} = 1$$

那么有如下关系成立

$$\sum_{k=1}^n |a_k b_k| \leq \left( \sum_{k=1}^n |a_k|^p \right)^{\frac{1}{p}} \left( \sum_{k=1}^n |b_k|^q \right)^{\frac{1}{q}}$$

由于  $\|\mathbf{y}\|_p \leq 1$ , 因此有  $\langle \mathbf{x}, \mathbf{y} \rangle \leq \|\mathbf{x}\|_q$ , 因此有  $\|\mathbf{x}\|_p^* = \|\mathbf{x}\|_q$  □

*Proof 2.* 首先, 假设  $\mathbf{y}$  为对称矩阵, 有

$$\langle \mathbf{x}, \mathbf{y} \rangle = \text{tr}(\mathbf{x}^T \mathbf{y}) = \text{tr}(\mathbf{B}^{-1} \mathbf{x}^T \mathbf{y} \mathbf{B})$$

其中, 有

$$\mathbf{B}^{-1} \mathbf{y} \mathbf{B} = \mathbf{V}$$

$\mathbf{V}$  为对角矩阵且对角线元素为矩阵  $\mathbf{y}$  的特征值

$$\text{tr}(\mathbf{B}^{-1} \mathbf{x}^T \mathbf{y} \mathbf{B}) = \text{tr}(\mathbf{x}^T \mathbf{V})$$

又有

$$\|\mathbf{y}\|_2 \leq 1$$

所以  $V$  的元素的绝对值都小于等于 1, 因此有

$$\text{tr}(\mathbf{x}^T \mathbf{V}) \leq \sum_{k=1}^n |\lambda_k| = \|\mathbf{x}\|_*$$

而由 von Neumann 迹定理知, 此时, 取到的结果是  $\langle \mathbf{x}, \mathbf{y} \rangle$  可以取到的上确界, 而上述证明说明了这一上确界在满足对偶范数定义的条件中是可以取到的, 因此上述结论成立。 □

*Proof 3.* 由  $p = 2, q = 2$  时的 Holder inequality, 即 Cauchy inequality

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{ij} \mathbf{x}_{ij} \mathbf{y}_{ij} \leq \left( \sum_{ij} |\mathbf{x}_{ij}|^2 \right)^{\frac{1}{2}} + \left( \sum_{ij} |\mathbf{y}_{ij}|^2 \right)^{\frac{1}{2}} \leq \|\mathbf{x}\|_F$$

所以有

$$\|\mathbf{x}\|_F^* = \|\mathbf{x}\|_F$$

□

**习题 (28).** *Proof (2.20).*

$$\begin{aligned} \frac{\partial \mathbf{X} \mathbf{Y}}{\partial t} &= \frac{\partial \mathbf{X}(\mathbf{t}) \mathbf{Y}(\mathbf{t})}{\partial t} \\ &= \frac{\partial (\mathbf{X}(\mathbf{t}) \mathbf{Y}(\mathbf{t}))_{ij}}{\partial t} \end{aligned}$$

其中  $\partial (\mathbf{X}(\mathbf{t}) \mathbf{Y}(\mathbf{t}))_{ij}$  表示两矩阵乘积的第  $i, j$  个元素, 下面考虑  $\partial (\mathbf{X}(\mathbf{t}) \mathbf{Y}(\mathbf{t}))_{ij}$  关于  $t$  的偏导数

$$\begin{aligned} \frac{\partial \mathbf{X}(\mathbf{t}) \mathbf{Y}(\mathbf{t})_{ij}}{\partial t} &= \frac{\partial \sum_{k=1}^n \mathbf{X}_{ik}(t) \mathbf{Y}_{kj}(t)}{\partial t} \\ &= \sum_{k=1}^n \frac{\partial (\mathbf{X}_{ik}(t) \mathbf{Y}_{kj}(t))}{\partial t} \\ &= \sum_{k=1}^n \left( \frac{\partial \mathbf{X}_{ik}(t)}{\partial t} \mathbf{Y}_{kj}(t) + \frac{\partial \mathbf{Y}_{kj}(t)}{\partial t} \mathbf{X}_{ik}(t) \right) \\ &= \sum_{k=1}^n \left( \frac{\partial \mathbf{X}_{ik}(t)}{\partial t} \mathbf{Y}_{kj}(t) \right) + \sum_{k=1}^n \left( \frac{\partial \mathbf{Y}_{kj}(t)}{\partial t} \mathbf{X}_{ik}(t) \right) \\ &= \left( \frac{\mathbf{X}(t)}{(t)} \mathbf{Y}(t) \right)_{ij} + \left( \frac{\partial \mathbf{Y}(t)}{\partial t} \mathbf{X}(t) \right)_{ij} \\ &= \left( \frac{\mathbf{X}(t)}{\partial t} \mathbf{Y}(t) + \frac{\partial \mathbf{Y}(t)}{\partial t} \mathbf{X}(t) \right)_{ij} \end{aligned}$$

所以有

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

□

*Proof (2.25).*

$$\begin{aligned} \frac{\partial (\mathbf{a}^T \mathbf{x})}{\partial \mathbf{x}} &= \mathbf{a} \\ \frac{\partial (\mathbf{a}^T \mathbf{x})}{\partial \mathbf{x}_i} &= \frac{\partial (\sum_{i=1}^n \mathbf{a}_i \mathbf{x}_i)}{\partial \mathbf{x}_i} = \mathbf{a}_i \end{aligned}$$

所以有

$$\frac{\partial (\mathbf{a}^T \mathbf{x})}{\partial \mathbf{x}} = \mathbf{a}$$

□

*Proof (2.26).*

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

考虑导数的第  $k$  个元素，有

$$\begin{aligned} \frac{\partial (\mathbf{x}^T \mathbf{A} \mathbf{x})}{\partial \mathbf{x}_k} &= \frac{\partial \left( \sum_{j=1}^n (\sum_{k=1}^n \mathbf{x}_k \mathbf{A}_{kj}) \mathbf{x}_j \right)}{\partial \mathbf{x}_i} \\ &= \frac{\partial \left( \sum_{k \neq i} \mathbf{x}_k \mathbf{A}_{ki} \mathbf{x}_i + \sum_{j \neq i} \mathbf{A}_{ij} \mathbf{x}_j \mathbf{x}_i + \mathbf{A}_{ii} \mathbf{x}_i^2 \right)}{\partial \mathbf{x}_i} \\ &= \sum_{k \neq i} \mathbf{A}_{ki} \mathbf{x}_k + \sum_{j \neq i} \mathbf{A}_{ij} \mathbf{x}_j + 2\mathbf{A}_{ii} \mathbf{x}_i \\ &= \sum_{t=1}^n \mathbf{A}_{ti} \mathbf{x}_t + \sum_{t=1}^n \mathbf{A}_{it} \mathbf{x}_t \\ &= (\mathbf{A} + \mathbf{A}^T)_i \mathbf{x} \end{aligned}$$

所以有

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

□

*Proof  $\frac{\partial \mathbf{X}}{\partial \mathbf{X}}$ .*

$$\frac{\partial \mathbf{X}}{\partial \mathbf{X}}$$

首先看

$$\frac{\partial \mathbf{X}}{\partial \mathbf{X}_{ij}}$$

有

$$\frac{\partial \mathbf{X}}{\partial \mathbf{X}_{ij}} = \mathbf{e}_i \mathbf{e}_j^T$$

其中为只有第  $i$  行为 1，其余行皆为 0 的向量，则有

$$\frac{\partial \mathbf{X}}{\partial \mathbf{X}} = \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \\ \vdots \\ \mathbf{e}_n \end{pmatrix} (\mathbf{e}_1^T, \mathbf{e}_2^T, \mathbf{e}_3^T, \dots, \mathbf{e}_n^T)$$

□

习题 (29).

$$\langle \mathcal{A}^*(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, \mathcal{A}(\mathbf{y}) \rangle$$

所以有

$$\begin{aligned} \langle \mathcal{A}^*(x), \mathbf{Y} \rangle &= (Y_{11} + Y_{12} - Y_{31} + 2Y_{33})x \\ &= \left\langle \begin{pmatrix} x & x & 0 \\ 0 & 0 & 0 \\ -x & 0 & 2x \end{pmatrix}, \mathbf{Y} \right\rangle \end{aligned}$$

因此

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

习题 (33). 证明. 首先, 由于有

$$\text{tr}(\mathbf{Y}\mathbf{K}\mathbf{Y}^T) = \text{tr}(\mathbf{Y}(\mathbf{K} + a\mathbf{I})\mathbf{Y}^T) - a$$

因此, 可以通过选择一个足够大的  $a$  使得矩阵  $\mathbf{K}$  变为正定矩阵, 因此, 这里不妨设矩阵  $\mathbf{K}$  为正定矩阵, 这样做只会使得目标函数增加一个常数, 不会对优化的求解产生任何影响。

$$\text{tr}(\mathbf{Y}\mathbf{K}\mathbf{Y}^T) = \text{tr}(\mathbf{Y}^T\mathbf{Y}\mathbf{K})$$

由于  $\sigma(\mathbf{Y}^T\mathbf{Y}) = \sigma(\mathbf{Y}\mathbf{Y}^T) = 1$ , 由 von Neumann 迹定理, 有

$$\sum_{i=1}^m \sigma_{n-i+1}(\mathbf{K}) \leq \text{tr}(\mathbf{Y}^T\mathbf{Y}\mathbf{K}) \leq \sum_{i=1}^m \sigma_i(\mathbf{K})$$

其中,  $m$  为  $\mathbf{Y}$  的行数, 下面说明什么时候可以取到最值, 当  $\mathbf{Y}^T$  取为上述定理中最大的  $m$  个奇异值, 即最大特征值对应的特征向量时, 有

$$\text{tr}(\mathbf{Y}\mathbf{K}\mathbf{Y}^T) = \text{tr}(\mathbf{Y}\mathbf{Y}^T\mathbf{V})$$

其中,  $\mathbf{V}$  是矩阵  $\mathbf{K}$  的前  $m$  个特征值组成的用零扩展的对角矩阵, 即上半部分为对角矩阵, 矩阵维数为  $m$ , 下半部分为零矩阵, 此时, 可以取到最大值。对应的, 当取最小特征值对应的特征向量为  $\mathbf{Y}$  时, 可以取到最小值。我尝试用拉格朗日乘子法求解, 由于约束条件取范数之后难以求导, 因此我没有办法用拉格朗日乘子法证明上述结论。  $\square$