Problem Set

原生生物

一　二次型

§1.1　基本结果

1. 记 $Q(x)=\dfrac{1}{2}x^TAx+b^Tx$，其中 A 为 n 阶对称方阵，x 为 n 维向量。

   (a) 设 $A=(a_{ij})$，试说明 $\dfrac{\partial x^TAx}{\partial x_1}=2\sum_{i=1}^n a_{1i}x_i$。

   (b) 证明 $\nabla_x Q(x)=Ax+b$。

   (c) 当 A 正定时，证明 $Q(x)$ 存在唯一最小值点 $x=A^{-1}b$。[注意梯度为 0 只能得到驻点]

2. 记 $Q(x)=x^TAx$，其中 A 为 n 阶对称半正定方阵 [由半正定阵定义，$Q(x)$ 的最小值为 0]。

   (a) 证明 $Q(x)=0\Leftrightarrow Ax=0$。

   (b) 设 $A=P^TDP$，其中 P 正交，D 是对角阵，非零对角元为 $\lambda_1,\ldots,\lambda_r$，求 $Q(x)$ 全部最小值点。

   (c) 举例：A 半正定，$\frac{1}{2}x^TAx+b^Tx$ 不存在最小值。

   (d) 对一般对称阵 A，何时 $\frac{1}{2}x^TAx+b^Tx$ 存在最小值？

§1.2　最小二乘

1. 记 $Q(x)=\|Ax-b\|$，其中 A 为 $m\times n$ 阶矩阵，x,b 为 n 维向量。[这里 $\|\alpha\|$ 代表二范数，即 $\sqrt{\alpha^T\alpha}$]

   (a) 计算 $\nabla_x Q(x)$。[先计算 $\nabla_x Q(x)^2$]

   (b) 利用梯度结果证明，$Q(x)$ 取到最小值时必有 $A^TAx=A^Tb$。

   (c) 证明 $A^TAx=A^Tb$ 时 $Q(x)$ 取到最小值。[假设满足此条件时为 x_0，考虑 $Q(x)^2-Q(x_0)^2$]

2. 考虑方程 $A^TAx=A^Tb$[由上一题，求解此方程可直接解出最小二乘，此方程称最小二乘问题的正则化方程组]。

   此处广义逆的定义 [不同情境下广义逆定义未必相同]：满足 $AA^+A=A,(AA^+)^T=AA^+$ 的矩阵 A^+ 称为 A 的广义逆。值得注意的是，当 A 未必为方阵时，广义逆仍然可以存在，若 $A_{m\times n}$，则 A^+ 为 $n\times m$ 阶矩阵。

   (a) 若 A^TA 可逆，验证 $(A^TA)^{-1}A^T$ 是 A 的广义逆。

   (b) 证明：当 $x=A^+b$ 时，$A^TAx=A^Tb$。[注意这里可以是任何一个广义逆]

   (c) 若存在 A' 使得 $\forall b$，$A'b$ 是 $\|Ax-b\|$ 的最小值点，证明 $A^TAA'=A^T$。

   (d) 用上一问的式子说明，A' 一定满足 $(AA')^T=AA',AA'A=A$，从而 A' 是 A 的广义逆。

二 范数

1. 在线性空间 $\mathbb{R}^n$ 中，考虑函数 $p:\mathbb{R}^n\to\mathbb{R}$。若其满足：

$$\forall x, p(x)\geq 0; p(x)=0\Leftrightarrow x=\mathbf{0}$$

$$p(\mu x)=|\mu|p(x)$$

$$p(x)+p(y)\geq p(x+y)$$

则称为 n 维向量的一个范数。

   (a) 当 A 为对称正定阵时，验证 $p(x)=\sqrt{x^TAx}$ 是一个范数，此范数记作 $\|x\|_A$。

   (b) 当 A 半正定时，证明 $p(x)$ 不是一个范数。

2. 设 $p(x)$ 是一个向量范数，考虑函数 $g(A)=\max_{p(x)=1}p(Ax)$，它是 n 阶方阵映射到一个数的函数。

   (a) 证明 $g(A)\geq 0$，且其为 0 当且仅当 $A=O$，即各个分量全为 0。

   (b) 证明 $g(\mu A)=|\mu|g(A)$。

   (c) 证明 $g(A)+g(B)\geq g(A+B)$。[由这三问，我们已经说明了它是所有 n 阶方阵中的一个范数。]

   (d) 证明 $g(A)p(x)\geq p(Ax)$。[由此，矩阵范数与向量范数具有某种相容性，例如，若向量范数为 $\|x\|_P$，可以将矩阵范数也记作 $\|A\|_P$，其中 P 是对称正定阵]

   (e) 证明 $g(A)g(B)\geq g(AB)$。[这一问是矩阵范数的额外性质]

   (f) 证明 $g(I)=1$[而单位阵的 Frobenius 范数为 $\sqrt{n}$]，且 $g(A)g(A^{-1})\geq 1$。[这个 $g(A)g(A^{-1})$ 一般称为 A 在范数 g 下的条件数]

3. 这里假设如上题所述的 p,g 记为向量范数 $\|x\|_p$ 和矩阵范数 $\|A\|_p$，并将矩阵范数下的条件数记作 $\sigma_p(A)$。对线性方程组 $Ax=b$，我们试着用矩阵范数考察解的扰动。以下假设 A 可逆。

   (a) 证明 $\|A^{-1}b\|_p\geq\frac{\|b\|_p}{\|A\|_p}$。

   (b) 假设 b 变为 $b+e_b$ 时解从 x 变为 $x+e_x$，求证 $\dfrac{\|e_x\|_p}{\|x\|_p}\leq\sigma_p(A)\dfrac{\|e_b\|_p}{\|b\|_p}$。

   (c) 证明式中的逆都存在时 $(A+E_A)^{-1}-A^{-1}=-(A+E_A)^{-1}E_AA^{-1}$。

   (d) 证明式中的逆都存在时 $\|I-(A+E_A)^{-1}E_A\|_p\|A^{-1}\|_p\|E_A\|_p\geq\|(A+E_A)^{-1}E_A\|_p$

   (e) 假设 A 变为 $A+E_A$，保证 $A+E_A$ 仍可逆且 $\|A^{-1}\|_p\|E_A\|_p<1$。若解从 x 变为 $x+e_x$，求证 $\dfrac{\|e_x\|_p}{\|x\|_p}\leq\dfrac{\|A^{-1}\|_p\|E_A\|_p}{1-\|A^{-1}\|_p\|E_A\|_p}$。

三 矩阵求导

§3.1 基本定义

1. 若 A 的每个分量都是 x 的函数 [这里 x 为一维变量]，定义 $\frac{\partial A}{\partial x}$ 的第 i 行第 j 列为 $\frac{\partial a_{ij}}{\partial x}$。

   (a) 计算说明 $\frac{\partial AB}{\partial x}=\frac{\partial A}{\partial x}B+A\frac{\partial B}{\partial x}$。[注意矩阵乘法顺序不可交换]

   (b) 计算说明 $\sum_{i,j}a_{ij}b_{ij}=\mathrm{tr}(A^TB)$，其中 A,B 为同阶矩阵。

   (c) 证明 $\frac{\partial\,\mathrm{tr}(A^TB)}{\partial x}=\mathrm{tr}(\frac{\partial A^TB}{\partial x})$。[注意 $\mathrm{tr}(A^TB)$ 是 x 的一维函数]

2. 若 $f(A)$ 是矩阵映射到数的函数，定义 $\frac{\partial f(A)}{\partial A}$ 的第 i 行第 j 列为 $\frac{\partial f(A)}{\partial a_{ij}}$。[当 A 是列向量 a 的时候，$\frac{\partial f(a)}{\partial a}$ 就是梯度]

(a) 计算说明 $\frac{\partial f(A)g(A)}{\partial A}=f(A)\frac{\partial g(A)}{\partial A}+g(A)\frac{\partial f(A)}{\partial A}$。[注意 $f(A),g(A)$ 是数乘]

(b) 计算说明 $\frac{\partial f(A)}{\partial x}=\operatorname{tr}\left(\left(\frac{\partial f(A)}{\partial A}\right)^T\frac{\partial A}{\partial x}\right)$。[这里 A 的每个位置都是 x 的函数，而最后又综合成了一个 x 的一维函数]

(c) 计算 $\frac{\partial \operatorname{tr}(A)}{\partial A}$，并由此重新证明上一题的 c。

§3.2　更多计算

1. 有关行列式的导数。

   (a) 用 Laplace 展开证明 $\frac{\partial \det A}{\partial a_{ij}}$ 是 A 的第 ij 个代数余子式。

   (b) 证明 $\frac{\partial \det A}{\partial A}=(A^*)^T$，其中 A^* 为 A 的伴随方阵。

   (c) 利用上题的 b 计算 $\frac{\partial \det A}{\partial x}$，并进一步计算 $\frac{\partial \ln\det A}{\partial x}$。

2. 极值问题：以下 w 是向量，W 是 $d\times d'$ 矩阵，X 是 $d\times m$ 矩阵，且 $n>d>d'$。

   (a) 回顾 $\frac{\partial w^TAw}{\partial w}$ 的结果，并计算 $\frac{\partial \operatorname{tr}(W^TXX^TW)}{\partial W}$。

   (b) 从计算结果分析怎样的 W 可以使 $\operatorname{tr}(W^TXX^TW)$ 取到极值。

   下面限定 W 满足 $W^TW=I_{d'}$，求解怎样的 W 使得 $\operatorname{tr}(W^TXX^TW)$ 取到最大值。

   (c) 利用拉格朗日乘数法，假设 W^TW 的第 i 行第 j 列对应乘数 λ_{ij}，且其拼成矩阵 Λ，证明 Lagrange 函数 $L(W,\Lambda)=\operatorname{tr}(W^TXX^TW)-\operatorname{tr}(\Lambda^T(W^TW-I))$。

   (d) 计算乘子部分对 W 的导数 $\dfrac{\partial \operatorname{tr}(\Lambda^T(W^TW-I))}{\partial W}$。

   (e) 注意到 W^TW 对称，其 ij 位置与 ji 位置恒相同，因此乘子 $\lambda_{ij}=\lambda_{ji}$，由此可知 Λ 也为对称阵。证明 $\dfrac{\partial L(W,\Lambda)}{\partial W}=O\Leftrightarrow XX^TW=W\Lambda$。

   附加　对 XX^T 作正交相似对角化 P^TDP，使得 D 的对角元从大到小排列。这时，P 的前 d' 列构成的矩阵就是最优的 W，此时 Λ 为对角阵，对角元是 D 的前 d' 个对角元。

   [注：这就是主成分分析的数学表达]