最优化第四次作业

162050127 颜劭铭

2022 年 5 月 22 日

1 基础题

1.1 令 A 是一个 $n \times n$ 的实对称矩阵, 其特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n, \lambda_i$ 对应的特征向量为 $u_i.A$ 的特征分解为 $A = U^T \Lambda U$, 其中 U 是正交矩阵, U 的第 i 列为 u_i, $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ 为对角阵.

1.1.1 对于任意 $x \neq 0$,

$$R(x, A) = \frac{x^T A x}{x^T x}$$

称为 Rayleigh 商. 证明对于任意的 $c \neq 0$, 都有 $R(cx, A) = R(x, A)$.

已知对于任意 $x \neq 0$, 有瑞利商：

$$R(x, A) = \frac{x^T A x}{x^T x}$$

又因为 $c \neq 0$, 所以:

$$R(cx, A) = \frac{(cx)^T A (cx)}{(cx)^T (cx)} = \frac{cx^T A cx}{cx^T cx} = \frac{c^2 x^T A x}{c^2 x^T x} = \frac{x^T A x}{x^T x} = R(x, A)$$

所以得证.

1.1.2 证明

$$\max_x \frac{x^T A x}{x^T x} = \max_{\|x\|=1} x^T A x$$

设 $R(x, A)$ 在 x^* 处取到最大值，即:

$$R(x^*, A) = \max_x \frac{x^T A x}{x^T x}$$

又因为要证

$$\max_x \frac{x^T A x}{x^T x} = \max_{\|x\|=1} x^T A x$$

因此可将 x^* 单位化：

$$\exists c \neq 0, \text{使} \|cx^*\| = 1$$

由 1.1.1 可得：

$$R(x^*, A) = R(cx^*, A) = \frac{(cx^*)^T A (cx^*)}{(cx^*)^T (cx^*)} = \frac{c^2 x^{*T} A x^*}{\|cx^*\|} = c^2 x^{*T} A x^* = \max_{\|x\|=1} x^T A x$$

得证.

1.1.3 证明: 对于任一单位向量 x, 都可以表示成 $x = \sum_{i=1}^n w_i u_i$, 其中 u_i 为特征向量, w_i 满足 $\|w\|^2 = \sum_{i=1}^n w_i^2 = 1$.

转化为证明 $x = \sum\limits_{i=1}^n w_i \mu_i$ 和当 $\|w\| = 1$ 时，x 为单位向量.

由题意可得：μ_i 为 A 的特征向量，因此：

$$u_i^T u_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

且 n 个特征向量是线性无关的，为一组基，因此可得：$\forall x \in R^n$ 可由这 n 个特征向量线性表出，即 $x = \sum\limits_{i=1}^n w_i \mu_i$。

接下来证明当 $\|w\| = 1$ 时，x 为单位向量.

$$\begin{aligned} \|x\|^2 = x^T x &= (w_1\mu_1 + w_2\mu_2 + \cdots + w_n\mu_n)^T (w_1\mu_1 + w_2\mu_2 + \cdots + w_n\mu_n) \\ &= \sum_{i=1}^n w_i^2 \mu_i^T \mu_i \\ &= \|w\|^2 = 1 \end{aligned}$$

因此得证.

1.1.4 证明

$$\max_{\|x\|=1} x^T A x = \lambda_1$$

并分析取到极大值时, 最优的 x 是多少? (提示, 利用 c 的结论, 将 x 写成特征向量的线性组合).

1.1.5 证明

$$\min_{\|x\|=1} x^T A x = \lambda_n$$

并分析取到极小值时, 最优的 x 是多少? (提示, 利用 c 的结论, 将 x 写成特征向量的线性组合).

1.1.4 和 1.1.5 因为证明思路相同，所以合起来证明.

由于 A 为实对称矩阵，其特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, 可以对其进行特征分解为 $A = U^T \Lambda U$, 其中 U 是正交矩阵, U 的第 i 列为 u_i, $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ 为对角阵.

因此瑞利商可以进行如下转化：

$$R(x, A) = \frac{x^T A x}{x^T x} = \frac{x^T U \Lambda U^T x}{x^T U U^T x}$$

令 $y = u^T x$，继续进行变换：

$$R(x, A) = \frac{y^T \Lambda y}{y^T y} = \frac{\sum\limits_{i=1}^{n} \lambda_i y_i^2}{\sum\limits_{i=1}^{n} y_i^2}$$

由特征值大小的关系显然有：

$$\lambda_n \sum_{i=1}^{n} y_i^2 \leqslant \sum_{i=1}^{n} \lambda_i y_i^2 \leqslant \lambda_1 \sum_{i=1}^{n} y_i^2$$

于是有：

$$\begin{aligned} & \frac{\lambda_n \sum\limits_{i=1}^{n} y_i^2}{\sum\limits_{i=1}^{n} y_i^2} \leqslant \frac{\sum\limits_{i=1}^{n} \lambda_i y_i^2}{\sum\limits_{i=1}^{n} y_i^2} \leqslant \frac{\lambda_1 \sum\limits_{i=1}^{n} y_i^2}{\sum\limits_{i=1}^{n} y_i^2} \\ \Rightarrow & \lambda_n \leqslant R(x, A) \leqslant \lambda_1 \end{aligned}$$

又由 1.1.2 的结论可得：

$$\min_{\|x\|=1} x^T A x = \lambda_n \qquad \max_{\|x\|=1} x^T A x = \lambda_1$$

由 1.1.3 可得：

$$\begin{aligned}
x^T A x &= \left(\sum_{i=1}^{n} w_i u_i\right)^T A \left(\sum_{i=1}^{n} w_i u_i\right) \\
&= \left(\sum_{i=1}^{n} w_i u_i\right)^T \left(\sum_{i=1}^{n} w_i A u_i\right) \\
&\text{由于 } A\mu = \lambda\mu \\
&= \left(\sum_{i=1}^{n} w_i u_i\right)^T \left(\sum_{i=1}^{n} w_i \lambda_i u_i\right) \\
&= \sum_{i=1}^{n} \lambda_i w_i^2 u_i^T u_i \\
&= \sum_{i=1}^{n} \lambda_i w_i^2
\end{aligned}$$

因此当取得极大值时，$w = [1, 0, \cdots, 0]^T$.

取得极小值时，$w = [0, 0, \cdots, 1]^T$.

1.2 考虑约束优化问题

$$\max f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q \mathbf{x}, \ \textbf{s.t.} \ \|\mathbf{x}\| = 1.$$

使用固定步长的投影梯度法求解上述问题, 迭代公式为

$$\mathbf{x}_{k+1} = \Pi_{\mathbf{y}\in D}\left(\mathbf{x}_k + \alpha \nabla f\left(\mathbf{x}_k\right)\right)$$

其中, $\alpha > 0, \Pi(\mathrm{x}) = \arg\min_{y\in D}\|\mathrm{x}-\mathrm{y}\|, D$ **为约束集** $D = \{\mathrm{x} \mid \|\mathrm{x}\| = 1\}$**.**

1.2.1 对于 $\mathrm{x} \neq 0$, 求出上面投影函数的显示表达式.

由于 $\Pi(\mathrm{x}) = \arg\min_{y\in D}\|\mathrm{x}-\mathrm{y}\|$，因此投影函数 $\mathbf{x}_{k+1} = \Pi_{\mathbf{y}\in D}\left(\mathbf{x}_k + \alpha \nabla f\left(\mathbf{x}_k\right)\right)$ 的显示表达式为:

$$\Pi_{\mathbf{y}\in D}\left(\mathbf{x}_k + \alpha \nabla f\left(\mathbf{x}_k\right)\right) = \arg\min_{y\in D}\left\|\mathbf{x}_k + \alpha \nabla f\left(\mathbf{x}_k\right) - \mathrm{y}\right\|$$

由于 $f(\mathbf{x})=\frac{1}{2}\mathbf{x}^TQ\mathbf{x}$，因此 $\nabla f(x_k)=Qx_k$.
因此可以构造拉格朗日函数 $L(y,\lambda)=\|\mathbf{x}_k+\alpha\nabla f\left(\mathbf{x}_k\right)-\mathrm{y}\|^2+\lambda(1-y^Ty)$
对上述式子求偏导可得：

$$\begin{cases}\dfrac{\partial\mathcal{L}}{\partial y}=2(x_k+\alpha Qx_k-y)-2\lambda y\\ \dfrac{\partial\mathcal{L}}{\partial\lambda}=1-y^Ty\end{cases}$$

令偏导为 0，可以解得：

$$y=\frac{x_k(I_n+\alpha Q)}{1-\lambda}$$

求二阶偏导可得：

$$\frac{\partial^2L}{\partial y^2}=2(1-\lambda)$$

且由刚刚的计算结果可得：

$$(1-\lambda)^2=\frac{\|x_k(I_n+\alpha Q)\|^2}{\|y\|^2}=\|x_k(I_n+\alpha Q)\|^2$$

所以可得，$\frac{\partial^2L}{\partial y^2}=2\|x_k(I_n+\alpha Q)\|$
因为 $Q=diag(1,2)$ 且 $\|x_k\|=1$
所以 $\frac{\partial^2L}{\partial y^2}>0$，所以该极值点为极小值点，且

$$x_{k+1}=y=\frac{x_k(I_n+\alpha Q)}{\|x_k(I_n+\alpha Q)\|}$$

1.2.2 令 $Q=\mathrm{diag}(1,2)$. 写出上面问题的最优性条件.

已知优化问题：

$$\begin{cases}\max f(\mathbf{x})=\dfrac{1}{2}\mathbf{x}^TQ\mathbf{x}\\ \text{s.t. } \|\mathbf{x}\|=1\end{cases}$$

因此可以构造拉格朗日函数，$L(x,\lambda)=\frac{1}{2}\mathbf{x}^TQ\mathbf{x}+\lambda(1-\|\mathbf{x}\|)=\frac{1}{2}\mathbf{x}^TQ\mathbf{x}+\lambda(1-\sqrt{x^Tx})$.
所以 KKT 条件为：

$$\begin{cases}\nabla_x L(x,\lambda) = Qx - \lambda\dfrac{x}{\|x\|} = 0\\ \nabla_\lambda L(x,\lambda) = \|\mathbf{x}\| - 1 = 0\end{cases}$$

所以结合上述两个式子可得：x^* 是对应于特征值 λ^* 的特征向量.

由因为 $Q = diag(1,2)$，所以 $\lambda_1 = 1$ 对应的特征向量为 [0,1]，$\lambda_2 = 2$ 对应的特征向量为 [1,0].

1.2.3 令 $Q = \mathrm{diag}(1,2)$, 判断临界点类型, 确定最优点.

对于正定矩阵 $Q = U\Lambda U^T, UU^T = I.\Lambda$ 为特征值的对角矩阵，且 $\lambda_1 = 1, \lambda_2 = 2$，则 1.2.2 的拉格朗日函数的 Hessian 矩阵可写为：

$$\nabla_x^2 L(x,\lambda) = Q - \lambda I$$

对于 $\lambda_1 = 1$，$\nabla_x^2 L(x,\lambda) = Q - \lambda_1 I = 0$，代入原式 $f(x)$ 可算出结果为 2.

对于 $\lambda_2 = 2$，$\nabla_x^2 L(x,\lambda) = Q - \lambda_2 I = 0$，代入原式 $f(x)$ 可算出结果为 1.

因此特征值 λ_1 对应的特征向量 [0,1] 为最优点，特征值 λ_2 对应的特征向量 [1,0] 是鞍点.

1.3 考虑优化问题

$$\min_{(x,y)\in\mathbb{R}^2} f(x) = (x-1)^2 + y - 2$$

$$\textbf{s.t.} h(x) = y - x - 1 = 0$$

$$g(x) = x + y - 2 \leq 0$$

计算满足 KKT 条件的点, 并利用二阶条件验证上述点是否是局部极小值点.

已知优化问题：

$$\begin{cases}\min\limits_{(x,y)\in\mathbb{R}^2} f(x) = (x-1)^2 + y - 2\\ \text{s.t.} h(x) = y - x - 1 = 0\\ \qquad g(x) = x + y - 2 \leq 0\end{cases}$$

因此可以构造拉格朗日函数：$L(x,y,\lambda_1,\lambda_2) = f(x) + \lambda_1 h(x) + \lambda_2 g(x) = (x-1)^2 + y - 2 + \lambda_1(y - x - 1) + \lambda_2(x + y - 2)$.

解得 KKT 条件为：

$$
\begin{cases}
\dfrac{\partial \mathcal{L}}{\partial x} = 2\left(x-1\right) - \lambda_1 + \lambda_2 = 0 \\
\dfrac{\partial \mathcal{L}}{\partial y} = 1 + \lambda_1 + \lambda_2 = 0 \\
y - x - 1 = 0 \\
x + y - 2 = 0 \\
\lambda_2 (y + x - 2) = 0 \\
\lambda_2 \geqslant 0
\end{cases}
$$

解得：

$$
\begin{cases}
x^* = \dfrac{1}{2} \\
y^* = \dfrac{3}{2} \\
\lambda_1^* = -1 \\
\lambda_2^* = 0
\end{cases}
$$

对 f(x) 进行变换：

$$
\begin{aligned}
f(x) &= x^2 - 2x + 1 + y - 2 = x^2 - 2x + y - 1 \\
&= \frac{1}{2}\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} -2 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} - 1 \\
&\text{令 } t = \begin{bmatrix} x \\ y \end{bmatrix} \\
&= \frac{1}{2} t^T Q t + b^T t + c
\end{aligned}
$$

因此同理可以对约束函数进行变换得：

$$
h(x) = \begin{bmatrix} -1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} - 1 \quad g(x) = \begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} - 2
$$

因此再次构造拉格朗日函数：

$$
L(t, \lambda_1, \lambda_2) = \frac{1}{2} t^T Q t + b^T t + c + \lambda_1 h(X) + \lambda_2 g(x)
$$

求偏导得：

$$\frac{\partial L}{\partial t} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} t + \begin{bmatrix} -2 & 1 \end{bmatrix} + \lambda_1 \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
$$\frac{\partial^2 L}{\partial t^2} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

由于 $\frac{\partial^2 L}{\partial t^2} \succcurlyeq 0$，所以 $(\frac{1}{2}, \frac{3}{2})$ 为极小点.

1.4 考虑优化问题:

$$\begin{aligned} &\min x_1 \\ &\textbf{s.t.} 16 - (x_1 - 4)^2 - x_2^2 \geq 0 \\ &\quad x_1^2 + (x_2 - 2)^2 - 4 = 0 \end{aligned}$$

1.4.1 写出上述问题的最优性条件, 计算 KKT 点.

已知优化问题：

$$\begin{cases} \min x_1 \\ \text{s.t.} 16 - (x_1 - 4)^2 - x_2^2 \geq 0 \\ \qquad x_1^2 + (x_2 - 2)^2 - 4 = 0 \end{cases}$$

因此可以构造拉格朗日函数：$L(x_1, x_2, \lambda_1, \lambda_2) = x_1 + \lambda_1(x_2^2 + (x_1-4)^2 - 16) + \lambda_2(x_1^2 + (x_2-2)^2 - 4)$

解得 KKT 条件为：

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x_1} = 1 + 2\lambda_1(x_1 - 4) + 2\lambda_2 x_1 = 0 \\ \dfrac{\partial \mathcal{L}}{\partial x_2} = 2\lambda_1 x_2 + 2\lambda_2(x_2 - 2) = 0 \\ x_2^2 + (x_1 - 4)^2 - 16 \leqslant 0 \\ x_1^2 + (x_2 - 2)^2 - 4 = 0 \\ \lambda_1(x_2^2 + (x_1 - 4)^2 - 16) = 0 \\ \lambda_1 \geqslant 0 \end{cases}$$

解得：

$$\begin{cases} x_1^* = 0 \\ x_2^* = 0 \\ \lambda_1^* = -\dfrac{1}{8} \\ \lambda_2^* = 0 \end{cases} \text{或} \quad \begin{cases} x_1^* = \dfrac{8}{5} \\ x_2^* = \dfrac{12}{5} \\ \lambda_1^* = -\dfrac{1}{24} \\ \lambda_2^* = -\dfrac{1}{4} \end{cases}$$

因此 KKT 点为 (0,0) 和 $(\frac{8}{5}, \frac{12}{5})$.

1.4.2 判断上述 KKT 点是否是局部极小点, 鞍点, 或全局极小点.

对 1.4.1 中拉格朗日函数求二阶导:

$$\begin{cases} \dfrac{\partial^2 \mathcal{L}}{\partial x_1^2} = 2\lambda_1 + 2\lambda_2 \\ \dfrac{\partial^2 \mathcal{L}}{\partial x_2^2} = 2\lambda_1 + 2\lambda_2 \\ \dfrac{\partial^2 \mathcal{L}}{\partial x_1 \partial x_2} = \dfrac{\partial^2 \mathcal{L}}{\partial x_2 \partial x_1} = 0 \end{cases}$$

因此将 1.4.1 中的 KKT 点代入，可求得 Hessian 矩阵为：

$$\begin{bmatrix} -\frac{1}{4} & 0 \\ 0 & -\frac{1}{4} \end{bmatrix} (\text{点 } (0,0)) \quad \begin{bmatrix} -\frac{7}{12} & 0 \\ 0 & -\frac{7}{12} \end{bmatrix} (\text{点 } (\frac{8}{5}, \frac{12}{5}))$$

检验可发现两个 Hessian 矩阵都是 $\succcurlyeq 0$ 的，但是将两个 KKT 点代入原式 f(x) 中易得：点 (0,0) 是极小点，而点 $(\frac{8}{5}, \frac{12}{5})$ 为鞍点.

1.5 利用增广拉格朗日方法求解以下问题:

$$\min f(x) = 2x_1^2 + x_2^2 - 2x_1x_2$$
$$\textbf{s.t. } x_1 + x_2 = 1.$$

1.5.1 写出上述问题的最优性条件, 计算 KKT 点.

已知优化问题

$$\begin{cases} \min f(x) = 2x_1^2 + x_2^2 - 2x_1x_2 \\ \text{s.t. } x_1 + x_2 = 1. \end{cases}$$

因此可以构造拉格朗日函数：$L(x_1,x_2,\lambda)=2x_1^2+x_2^2-2x_1x_2+\lambda(x_1+x_2-1)$
解得 KKT 条件为：

$$\begin{cases}\dfrac{\partial\mathcal{L}}{\partial x_1}=4x_1-2x_2+\lambda=0\\ \dfrac{\partial\mathcal{L}}{\partial x_2}=2x_2-2x_1+\lambda=0\\ x_1+x_2-1=0\\ \lambda\neq 0\end{cases}$$

解得：

$$\begin{cases}x_1^*=\dfrac{2}{5}\\ x_2^*=\dfrac{3}{5}\\ \lambda^*=-\dfrac{2}{5}\end{cases}$$

所以 KKT 点为 $(\frac{2}{5},\frac{3}{5})$

1.5.2 写出增广的拉格朗日函数.

易得增广拉格朗日函数如下：

$$\mathcal{P}_E(x_1,x_2,\lambda,\sigma)=2x_1^2+x_2^2-2x_1x_2+\lambda(x_1+x_2-1)+\frac{\sigma}{2}(x_1+x_2-1)^2$$

1.5.3 令 $\sigma=2,\lambda=1$, 求解上述增广的拉格朗日函数的极小值.

由于 $\sigma=2,\lambda=1$，所以将其代入 1.5.2 中的 $\mathcal{P}_E(x_1,x_2,\lambda,\sigma)$ 可得：

$$\begin{aligned}\mathcal{P}_E(x_1,x_2,\lambda,\sigma)&=2x_1^2+x_2^2-2x_1x_2+\lambda(x_1+x_2-1)+\frac{\sigma}{2}(x_1+x_2-1)^2\\&=2x_1^2+x_2^2-2x_1x_2+x_1+x_2-1+(x_1+x_2-1)^2\\&=2x_1^2+x_2^2-2x_1x_2+x_1+x_2-1+x_1^2+x_1x_2-x_1+x_1x_2\\&+x_2^2-x_2-x_1-x_2+1\\&=3x_1^2+2x_2^2-x_1-x_2\end{aligned}$$

求上式偏导可得：

$$
\begin{cases}
\dfrac{\partial \mathcal{P}_E}{\partial x_1} = 6x_1 - 1 = 0 \\
\dfrac{\partial \mathcal{P}_E}{\partial x_2} = 4x_2 - 1 = 0
\end{cases}
$$

解得：

$$
\begin{cases}
x_1^* = \dfrac{1}{6} \\
x_2^* = \dfrac{1}{4}
\end{cases}
$$

将其代入 $\mathcal{P}_E$ 中，可得其极小值为 $-\frac{5}{24}$

1.5.4 计算修正的 λ.