瑞利商与极值计算

模式识别系列 2017年 11月27日

对于一个埃尔米特矩阵 M 及非零向量 x，定义瑞利商

$$R(M,x)=\frac{x^*Mx}{x^*x}$$

这里的 x^* 是 x 的共轭转置矩阵，如果 M,x 都由实数元素组成，那么瑞利商可以写成

$$R(M,x)=\frac{x^TMx}{x^Tx}$$

设 M 的特征值与特征向量分别为 $\lambda_1,\lambda_2,,,\lambda_n$，$v_1,v_2,,,v_n$，并且有

$$\lambda_{min}=\lambda_1\leq\lambda_2\leq\ldots\leq\lambda_n=\lambda_{max}$$

下面将证明，在 M 确定的情况下

$$\max_x R(M,x)=\lambda_n$$
$$\min_x R(M,x)=\lambda_1$$

由于 M 是一个埃尔米特矩阵，所以存在一个酉矩阵 U 满足

$$M=UAU^T$$

其中 $A=diag(\lambda_1,\lambda_2,,,\lambda_n)$，将上式代入瑞利商

$$\begin{aligned}R(M,x)&=\frac{x^TUAU^Tx}{x^Tx}\\&=\frac{(U^Tx)^TA(U^Tx)}{x^Tx}\end{aligned}$$

假设 $p=U^Tx$ 那么

$$\begin{aligned}R(M,x)&=\frac{p^TAp}{x^Tx}\\&=\frac{\sum_{i=1}^n\lambda_i|p_i|^2}{\sum_{i=1}^n|x_i|^2}\end{aligned}$$

根据特征值的大小关系，可得如下不等式

$$\lambda_1 \sum_{i=1}^{n} |p_i|^2 \leq \sum_{i=1}^{n} \lambda_i |p_i|^2 \leq \lambda_n \sum_{i=1}^{n} |p_i|^2$$

于是有

$$\frac{\lambda_1 \sum_{i=1}^{n} |p_i|^2}{\sum_{i=1}^{n} |x_i|^2} \leq R(M,x) \leq \frac{\lambda_n \sum_{i=1}^{n} |p_i|^2}{\sum_{i=1}^{n} |x_i|^2}$$

设 U 的第 i 行，第 j 列元素为 u_{ij}，U^T 的第 i 行，第 j 列元素为 u_{ji}，那么

$$p_i = \sum_{j=1}^{n} u_{ji} x_j$$

$$p_i^T = \sum_{j=1}^{n} x_j u_{ij}$$

$$|p_i|^2 = p_i^T p_i = \sum_{j=1}^{n} \sum_{k=1}^{n} x_j u_{ij} u_{ki} x_k$$

于是

$$\sum_{i=1}^{n} |p_i|^2 = \sum_{j=1}^{n} \sum_{k=1}^{n} \left(\sum_{i=1}^{n} u_{ki} u_{ij} \right) x_j x_k$$

由于 U 是酉矩阵，即

$$U^T U = I$$

写成展开形式为

$$I_{jk} = \sum_{i=1}^{n} u_{ji} u_{ik}$$

当 $j \neq k$ 时，$I_{jk} = 0$，当 $j = k$ 时，$I_{jk} = 1$。所以可以得到

$$\sum_{i=1}^{n} |p_i|^2 = \sum_{i=1}^{n} |x_i|^2$$

代入上述不等式，可得

$$\lambda_1 \leq R(M,x) \leq \lambda_n$$

并且当 $x = v_1$ 时 $R(M,x) = \lambda_1$，当 $x = v_n$ 时 $R(M,x) = \lambda_n$。这就证明了前面的结论。

另一方面，如果我们用 $x' = cx$ 来取代 x，其中 c 为非零的实数，发现

$$R(M, x') = \frac{x'^T M x'}{x'^T x} = \frac{c x^T M x c}{c x^T x c} = R(M, x)$$

也就是说，对 x 进行等比例缩放并不会影响瑞利商的值，即

$$R(M, cx) = R(M, x)$$

于是，我们可以令 $x^T x = 1$，这样就有 $R(M, x) = x^T M x$。此时对瑞利商求极值就是在约束 $x^T x = 1$ 条件下，对 $x^T M x$ 求极值。下面使用拉格朗日乘子法来解，定义拉格朗日函数

$$L(x, \lambda) = x^T M x - \lambda(x^T x - 1)$$

对 x 求梯度，并令值为0

$$\nabla_x L = Mx - \lambda x = 0$$

即 M 的特征值能使得瑞利商取得极值，并且 $R(M, x) = \lambda$。

瑞利商的另一种推广形式——广义瑞利商，在 Fisher 线性判别分析中有重要应用。定义

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

其中 Q 为对称正定矩阵，基于同样的理由，我们缩放 x 使得 $x^T Q x = 1$，然后利用拉格朗日乘子法求 $x^T M x$ 的极值，定义

$$L(x, \lambda) = x^T M x - \lambda(x^T Q x - 1)$$

然后求梯度取零

$$\begin{aligned} &\nabla_x L = Mx - \lambda Q x = 0 \\ &\Leftrightarrow Mx = \lambda Q x \\ &\Leftrightarrow Q^{-1} M x = \lambda x \end{aligned}$$

也就是说，$R(M, x, Q)$ 的极值在 $Q^{-1} M$ 的特征向量上取得，其驻值就为特征值。

clouswang@gmail.com