HW10

PB20111689 蓝俊玮

T1

因为有：

$$\sum_c P(c|\boldsymbol{x})P(c|\boldsymbol{z}) \le \max_{c\in\mathcal{Y}} P(c|\boldsymbol{x}) \sum_c P(c|\boldsymbol{z}) = \max_{c\in\mathcal{Y}} P(c|\boldsymbol{x})$$

所以可以得到：

$$err^*(\boldsymbol{x}) = 1 - \max_{c\in\mathcal{Y}} P(c|\boldsymbol{x}) \le 1 - \sum_c P(c|\boldsymbol{x})P(c|\boldsymbol{z}) = err(\boldsymbol{x})$$

记 $err^*(\boldsymbol{x}) = 1 - P(c^*|\boldsymbol{x})$ 同时也有：

$$\begin{aligned}
err(\boldsymbol{x}) =&1 - \sum_c P(c|\boldsymbol{x})P(c|\boldsymbol{z}) \\
\approx&1 - \sum_c P^2(c|\boldsymbol{x}) \\
=&1 - P^2(c^*|\boldsymbol{x}) - \sum_{c\ne c^*} P^2(c|\boldsymbol{x}) \\
=&(1 + P(c^*|\boldsymbol{x}))(1 - P(c^*|\boldsymbol{x})) - \sum_{c\ne c^*} P^2(c|\boldsymbol{x}) \\
=&(2 - err^*(\boldsymbol{x}))err^*(\boldsymbol{x}) - \sum_{c\ne c^*} P^2(c|\boldsymbol{x})
\end{aligned}$$

由柯西不等式可以得到：

$$\sum_{c\ne c^*} P^2(c|\boldsymbol{x}) \ge \frac{1}{|\mathcal{Y}|-1}(\sum_{c\ne c^*} P(c|\boldsymbol{x}))^2 = \frac{1}{|\mathcal{Y}|-1}(1 - P(c^*|\boldsymbol{x}))^2 = \frac{1}{|\mathcal{Y}|-1}(err^*(\boldsymbol{x}))^2$$

将该式代入前式，可以得到：

$$\begin{aligned}
err(\boldsymbol{x}) \le&(2 - err^*(\boldsymbol{x}))err^*(\boldsymbol{x}) - \frac{1}{|\mathcal{Y}|-1}(err^*(\boldsymbol{x}))^2 \\
=&err^*(\boldsymbol{x})(2 - \frac{|\mathcal{Y}|}{|\mathcal{Y}|-1} \times err^*(\boldsymbol{x}))
\end{aligned}$$

综上所述，可以知道：

$$err^*(\boldsymbol{x}) \le err(\boldsymbol{x}) \le err^*(\boldsymbol{x})(2 - \frac{|\mathcal{Y}|}{|\mathcal{Y}|-1} \times err^*(\boldsymbol{x}))$$

T2

> 在实践中，协方差矩阵 $\boldsymbol{X}\boldsymbol{X}^T$ 的特征值分解常由中心化后的样本矩阵 $\boldsymbol{X}$ 的奇异值分解代替，试述其原因

$\boldsymbol{X}$ 有奇异值分解为 $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$，则 $\boldsymbol{X}\boldsymbol{X}^T = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T$，而 $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$，因此有 $\boldsymbol{X}\boldsymbol{X}^T = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{U}^T = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$。

而对 $\boldsymbol{X}\boldsymbol{X}^T$ 特征值分解有 $\boldsymbol{X}\boldsymbol{X}^T=\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T$。只要让 $\boldsymbol{P}=\boldsymbol{U}$，所以可以用奇异值分解来代替特征值分解，可以节省计算和存储的成本，且计算精度较高。

T3

$$
L(\boldsymbol{W})=\operatorname{tr}\left(\boldsymbol{W}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}\right)-\operatorname{tr}\left(\Lambda^T(\boldsymbol{W}^T\boldsymbol{W}-\boldsymbol{I}_{d'})\right)
$$

$$
\frac{\partial L(\boldsymbol{W})}{\partial \boldsymbol{W}}=(\boldsymbol{X}\boldsymbol{X}^T+\boldsymbol{X}\boldsymbol{X}^T)\boldsymbol{W}-(\boldsymbol{W}\Lambda+\boldsymbol{W}\Lambda^T)=2\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}-2\boldsymbol{W}\Lambda
$$

令 $\frac{\partial L(\boldsymbol{W})}{\partial \boldsymbol{W}}=0$，可以求得：

$$
\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{W}=\boldsymbol{W}\Lambda
$$

将 $\boldsymbol{W}$ 和 Λ 展开（对协方差矩阵 $\boldsymbol{X}\boldsymbol{X}^T$ 进行特征值分解并将特征值进行排序），可以得到：

$$
\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{w}_i=\lambda_i\boldsymbol{w}_i \quad i=1,2,\ldots,d'
$$

将特征向量构成 $\boldsymbol{W}=(\boldsymbol{w}_1,\boldsymbol{w}_2,\ldots,\boldsymbol{w}_{d'})$ 就是主成分分析的解。

T4

不是凸优化问题。因为 P 不能满足是凸集。

同时

$$
f(P)=\sum_{(x_i,x_j)\in\mathcal{M}}(x_i-x_j)^TPP^T(x_i-x_j)
$$

$$
g(P)=1-\sum_{(x_i,x_j)\in\mathcal{M}}(x_i-x_j)^TPP^T(x_i-x_j)
$$

则对于原问题就有：

$$
\min_P f(P) \quad s.t.\ g(P)\leq 0
$$

可知 $f(P)$ 和 $g(P)$ 的二阶导数符号是相反的，因此 $f(P)$ 和 $g(P)$ 不能同时为凸函数，因此不是凸优化问题。