HW10

PB20111689 蓝俊玮

T1

因为有:

$$\sum_{c} P(c|oldsymbol{x}) P(c|oldsymbol{z}) \leq \max_{c \in \mathcal{Y}} P(c|oldsymbol{x}) \sum_{c} P(c|oldsymbol{z}) = \max_{c \in \mathcal{Y}} P(c|oldsymbol{x})$$

所以可以得到:

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

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

$$egin{aligned} err(oldsymbol{x}) =& 1 - \sum_{c} P(c|oldsymbol{x}) P(c|oldsymbol{z}) \ pprox & 1 - \sum_{c} P^2(c|oldsymbol{x}) \ & = 1 - P^2(c^*|oldsymbol{x}) - \sum_{c
eq c^*} P^2(c|oldsymbol{x}) \ & = (1 + P(c^*|oldsymbol{x}))(1 - P(c^*|oldsymbol{x})) - \sum_{c
eq c^*} P^2(c|oldsymbol{x}) \ & = (2 - err^*(oldsymbol{x}))err^*(oldsymbol{x}) - \sum_{c
eq c^*} P^2(c|oldsymbol{x}) \end{aligned}$$

由柯西不等式可以得到:

$$\sum_{c
eq c^*} P^2(c|m{x}) \geq rac{1}{|\mathcal{Y}|-1} (\sum_{c
eq c^*} P(c|m{x}))^2 = rac{1}{|\mathcal{Y}|-1} (1-P(c^*|m{x}))^2 = rac{1}{|\mathcal{Y}|-1} (err^*(m{x}))^2$$

将该式代入前式,可以得到:

$$egin{aligned} err(oldsymbol{x}) \leq &(2 - err^*(oldsymbol{x}))err^*(oldsymbol{x}) - rac{1}{|\mathcal{Y}| - 1}(err^*(oldsymbol{x}))^2 \ = &err^*(oldsymbol{x})(2 - rac{|\mathcal{Y}|}{|\mathcal{Y}| - 1} imes err^*(oldsymbol{x})) \end{aligned}$$

综上所述,可以知道:

$$err^*(oldsymbol{x}) \leq err(oldsymbol{x}) \leq err^*(oldsymbol{x})(2 - rac{|\mathcal{Y}|}{|\mathcal{Y}|-1} imes err^*(oldsymbol{x}))$$

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

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

而对 $m{X}m{X}^T$ 特征值分解有 $m{X}m{X}^T = m{P}m{\Lambda}m{P}^T$ 。只要让 $m{P} = m{U}$,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

$$L(oldsymbol{W}) = ext{tr}\left(oldsymbol{W}^Toldsymbol{X}oldsymbol{X}^Toldsymbol{W}
ight) - ext{tr}\left(oldsymbol{\Lambda}^T(oldsymbol{W}^Toldsymbol{W}-oldsymbol{I}_{d'}
ight))} \ rac{\partial L(oldsymbol{W})}{\partial oldsymbol{W}} = (oldsymbol{X}oldsymbol{X}^T + oldsymbol{X}oldsymbol{X}^T)oldsymbol{W} - (oldsymbol{W}oldsymbol{\Lambda} + oldsymbol{W}oldsymbol{\Lambda}^T) = 2oldsymbol{X}oldsymbol{X}^Toldsymbol{W} - 2oldsymbol{W}oldsymbol{\Lambda}$$

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

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

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

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

将特征向量构成 $\boldsymbol{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_{d'})$ 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

$$egin{aligned} f(P) &= \sum_{(x_i, x_j) \in \mathcal{M}} (x_i - x_j)^T P P^T (x_i - x_j) \ g(P) &= 1 - \sum_{(x_i, x_j) \in \mathcal{M}} (x_i - x_j)^T P P^T (x_i - x_j) \end{aligned}$$

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。