HW3

PB20111689 蓝俊玮

3.2

$$y=\frac{1}{1+e^{-(\boldsymbol{w}^T\boldsymbol{x}+b)}} \quad (3.18)$$

$$l(\boldsymbol{\beta})=\sum_{i=1}^{m}(-y_i\boldsymbol{\beta}^T\hat{\boldsymbol{x}_i}+\ln\left(1+e^{\boldsymbol{\beta}^T\hat{\boldsymbol{x}_i}}\right)) \quad (3.27)$$

那么可以计算

$$\frac{\partial y}{\partial \boldsymbol{w}}=\frac{\boldsymbol{x}e^{-(\boldsymbol{w}^T\boldsymbol{x}+b)}}{\left(1+e^{-(\boldsymbol{w}^T\boldsymbol{x}+b)}\right)^2}$$

$$\frac{\partial^2 y}{\partial \boldsymbol{w}\partial \boldsymbol{w}^T}=\boldsymbol{x}\boldsymbol{x}^Te^{-(\boldsymbol{w}^T\boldsymbol{x}+b)}\frac{-1+e^{-(\boldsymbol{w}^T\boldsymbol{x}+b)}}{\left(1+e^{-(\boldsymbol{w}^T\boldsymbol{x}+b)}\right)^3}$$

则可以看出 $\frac{\partial^2 y}{\partial \boldsymbol{w}^2}$ 在 $e^{-(\boldsymbol{w}^T\boldsymbol{x}+b)}=1$ 时取 0，则其有正有负，所以它的海森矩阵不是半正定的，因此它不是凸函数。

根据课本式子

$$\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}=\sum_{i=1}^{m}\hat{\boldsymbol{x}}_i{\hat{\boldsymbol{x}}_i}^Tp_1(\hat{\boldsymbol{x}}_i;\boldsymbol{\beta})(1-p_1(\hat{\boldsymbol{x}}_i;\boldsymbol{\beta})) \quad (3.31)$$

因为概率 $p_1(\hat{\boldsymbol{x}}_i;\boldsymbol{\beta})\in[0,1]$，$\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T}\geq 0$，所以式子 (3.31) 是半正定的，所以 $l(\boldsymbol{\beta})$ 是凸函数。

3.7

> 令码长为 9，类别数为 4，试给出海明距离意义下理论最优的 ECOC 二元码并证明之

这里的分歧我认为是这个定义不够准确，网上绝大多数的定义为正反码两个意义下的海明距离都要最优。

根据助教在 gitee 所言。如果只考虑正码之间两两海明距离最小的情况，而不考虑反码下的话，则有：

	f1	f2	f3	f4	f5	f6	f7	f8	f9
c1	1	1	1	1	1	1	1	1	1
c2	1	1	1	1	1	1	-1	-1	-1
c3	1	1	1	-1	-1	-1	1	1	1
c4	-1	-1	-1	1	1	1	1	1	1

则可以取到最小海明距离为 6 的 ECOC 二元码。

但如果考虑正反码意义下的两两之间海明距离最小的情况，则根据 Exhausted Code 可以给出

	f1	f2	f3	f4	f5	f6	f7	f8	f9
c1	1	1	1	1	1	1	1	1	1
c2	-1	-1	-1	-1	1	1	1	1	-1
c3	-1	-1	1	1	-1	-1	1	1	1
c4	-1	1	-1	1	-1	1	-1	1	1

则可以取到最小海明距离为 4 的 ECOC 二元码。

补充题

> 在 LDA 多分类情形下，试计算类间散度矩阵 S_b 的秩并证明

在 LDA 多分类情形下，类间散度矩阵 $S_b = \sum\limits_{i=1}^{N} m_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T$。因为有引理 $\operatorname{rank}(AB) \leq \min\left(\operatorname{rank} A, \operatorname{rank} B\right)$，所以可以知道 $(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T$ 的秩为 1。同时因为 $\boldsymbol{\mu} = \frac{1}{m}\sum\limits_{i=1}^{N} m_i\boldsymbol{\mu}_i$，所以对 S_b 中的所有矩阵，取 $\forall c_i = \frac{1}{m}$，将每一个矩阵求和乘上系数后有

$$\sum_{i=1}^{N} c_i m_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T = \frac{1}{m}\sum_{i=1}^{N} m_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T =$$

$$(\frac{m\boldsymbol{\mu}}{m} - \frac{1}{m}\sum_{i=1}^{N} m_i\boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T = (\boldsymbol{\mu} - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T = 0$$

使得 S_b 中只有 $N-1$ 个无关的矩阵，所以 S_b 的秩至多为 $N-1$，即有 $\operatorname{rank}\left(S_b\right) \leq N-1$。

> 给出公式 3.45 的推导证明

此处也固定分母为 1，则可得到该优化问题

$$\begin{aligned} \min_{w} \quad & -\operatorname{tr}\left(W^T S_b W\right) \\ s.t. \quad & \operatorname{tr}\left(W^T S_w W\right) = 1 \end{aligned}$$

由拉格朗日乘子法，则

$$L(W, \lambda) = -\operatorname{tr}\left(W^T S_b W\right) + \lambda \operatorname{tr}\left(W^T S_w W\right) - \lambda$$

则求导可得

$$\frac{\partial L(W,\lambda)}{\partial W} = -(S_b + S_b^T)W + \lambda(S_w + S_w^T)W = 0$$

$$\frac{\partial L(W,\lambda)}{\partial \lambda} = \operatorname{tr}\left(W^T S_w W\right) - 1 = 0$$

$$-2S_b W + 2\lambda S_w W = 0$$

$$S_b W = \lambda S_w W$$

> 证明 $X(X^TX)^{-1}X^T$ 是投影矩阵，并对线性回归模型从投影角度解释

$X(X^TX)^{-1}X^T \cdot X(X^TX)^{-1}X^T = X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T$

所以 $X(X^TX)^{-1}X^T$ 是幂等矩阵，所以它是投影矩阵。

对于线性回归模型有 $f(\hat{\boldsymbol{x}}_i) = \boldsymbol{x}_i\hat{\boldsymbol{w}}$，通过求其偏导使得 $\nabla_{\hat{\boldsymbol{w}}}E(\hat{\boldsymbol{w}}) = 0$ 得到 $\hat{\boldsymbol{w}} = (X^TX)^{-1}X^Ty$，则带入可以得到 $f(\hat{\boldsymbol{x}}_i) = \boldsymbol{x}_i(X^TX)^{-1}X^Ty$，即有 $\hat{Y} = X(X^TX)^{-1}X^TY$，所以可以这样解释 $X(X^TX)^{-1}X^T$ 这个矩阵将原来的数据集的标注矩阵 Y 映射到了测试集的预测矩阵 $\hat{Y}$ 上，即通过这个矩阵将其映射到了一个线性集上，因此可以认为是将空间中的标注投影到这个线性集上。所以从投影角度分析，这个投影矩阵就是将所有的数据通过投影的方式而映射到一个线性集上。同时根据这个是幂等矩阵，因此其无论投影多少次，其投影结果还是一致的。