IML 第二次作业

习题 3.2

令 $y=\dfrac{1}{1+e^{-(\boldsymbol{w}^\top\boldsymbol{x}+b)}}$，$l(\boldsymbol{\beta})=\sum_{i=1}^{m}\left(-y_i\boldsymbol{\beta}^\top\hat{\boldsymbol{x}}_i+\ln(1+e^{\boldsymbol{\beta}^\top\hat{\boldsymbol{x}}_i})\right)$，这两个函数关于 $\boldsymbol{w}$ 和 $\boldsymbol{\beta}=(\boldsymbol{w};b)$ 是二阶可微的，分别计算二者的 Hessian 矩阵：

$$\frac{\partial y}{\partial\boldsymbol{\omega}}=\frac{e^{-\left(\boldsymbol{\omega}^\top\boldsymbol{x}+b\right)}}{\left[1+e^{-\left(\boldsymbol{\omega}^\top\boldsymbol{x}+b\right)}\right]^2}\boldsymbol{x}$$

$$\begin{aligned}\frac{\partial^2 y}{\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}^\top}&=\frac{\partial}{\partial\boldsymbol{\omega}^\top}\frac{\partial y}{\partial\boldsymbol{\omega}}\\&=\frac{\partial}{\partial\boldsymbol{\omega}^\top}\frac{e^{-\left(\boldsymbol{\omega}^\top\boldsymbol{x}+b\right)}}{\left[1+e^{-\left(\boldsymbol{\omega}^\top\boldsymbol{x}+b\right)}\right]^2}\boldsymbol{x}\\&=\frac{e^{-\left(\boldsymbol{\omega}^\top\boldsymbol{x}+b\right)}\left[1-e^{-\left(\boldsymbol{\omega}^\top\boldsymbol{x}+b\right)}\right]}{\left[1+e^{-\left(\boldsymbol{\omega}^\top\boldsymbol{x}+b\right)}\right]^3}\boldsymbol{x}\boldsymbol{x}^\top\\&=y(1-y)(1-2y)\boldsymbol{x}\boldsymbol{x}^\top\end{aligned}$$

矩阵 $\boldsymbol{x}\boldsymbol{x}^\top$ 半正定，而 $y(1-y)(1-2y)<0(\text{as } y\in\left(\frac{1}{2},1\right))$ ，其 Hessian 矩阵不总非负，即 y 是非凸的。

$$\frac{\partial l}{\partial\boldsymbol{\beta}}=\sum_{i=1}^{m}\left(-y_i\hat{\boldsymbol{x}}_i+\frac{e^{\boldsymbol{\beta}^\top\hat{\boldsymbol{x}}_i}}{1+e^{\boldsymbol{\beta}^\top\hat{\boldsymbol{x}}_i}}\hat{\boldsymbol{x}}_i\right)$$

$$\begin{aligned}\frac{\partial^2 l}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}&=\frac{\partial}{\partial\boldsymbol{\beta}^\top}\frac{\partial l}{\partial\boldsymbol{\beta}}\\&=\frac{\partial}{\partial\boldsymbol{\beta}^\top}\sum_{i=1}^{m}\left(-y_i\hat{\boldsymbol{x}}_i+\frac{e^{\boldsymbol{\beta}^\top\hat{\boldsymbol{x}}_i}}{1+e^{\boldsymbol{\beta}^\top\hat{\boldsymbol{x}}_i}}\hat{\boldsymbol{x}}_i\right)\\&=\sum_{i=1}^{m}\frac{e^{\boldsymbol{\beta}^\top\hat{\boldsymbol{x}}_i}}{\left(1+e^{\boldsymbol{\beta}^\top\hat{\boldsymbol{x}}_i}\right)^2}\hat{\boldsymbol{x}}_i\hat{\boldsymbol{x}}_i^\top\end{aligned}$$

矩阵 $\hat{\boldsymbol{x}}_i\hat{\boldsymbol{x}}_i^\top$ 半正定，而 $\frac{e^{\boldsymbol{\beta}^\top\hat{\boldsymbol{x}}_i}}{\left(1+e^{\boldsymbol{\beta}^\top\hat{\boldsymbol{x}}_i}\right)^2}\hat{\boldsymbol{x}}_i\hat{\boldsymbol{x}}_i^\top>0$ ，所以其 Hessian 矩阵半正定，即 $l(\boldsymbol{\beta})$ 是凸的。

习题 3.7

设类别 i 的 ECOC 码为 r_i，定义 $d(r_i,r_j)$ 为其海明距离（编码不同的位数）。最大化的目标为

$$l=\prod_{1\le i<j\le 4}d(r_i,r_j)+\sum_{1\le i<j\le 4}d(r_i,r_j)$$

编写 C 代码程序，搜索得出解为

$$C_1=000000000 \quad C_2=000111111 \quad C_3=111000111 \quad C_4=111111000$$

其 $L=46692$。

作业 3.3

多分类情形下的 $\boldsymbol{S}_b=\sum_{i=1}^{N} m_i(\boldsymbol{\mu}_i-\boldsymbol{\mu})(\boldsymbol{\mu}_i-\boldsymbol{\mu})^{\top}$。

$$\boldsymbol{S}_b=[(\boldsymbol{\mu}_1-\boldsymbol{\mu}),(\boldsymbol{\mu}_2-\boldsymbol{\mu}),\ldots,(\boldsymbol{\mu}_N-\boldsymbol{\mu})]\begin{pmatrix}\boldsymbol{m}_1 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \ldots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{m}_N\end{pmatrix}\begin{pmatrix}(\boldsymbol{\mu}_1-\boldsymbol{\mu})^{\top}\\ \ldots\\ (\boldsymbol{\mu}_N-\boldsymbol{\mu})^{\top}\end{pmatrix}$$

记 $\boldsymbol{M}=\operatorname{diag}(m_1,m_2,\ldots,m_N)$, $\boldsymbol{A}=[(\boldsymbol{\mu}_1-\boldsymbol{\mu}),(\boldsymbol{\mu}_2-\boldsymbol{\mu}),\ldots,(\boldsymbol{\mu}_N-\boldsymbol{\mu})]^{\top}$, 则

$$\begin{aligned}\operatorname{rank}\boldsymbol{S}_b&=\operatorname{rank}\boldsymbol{A}^{\top}\boldsymbol{M}\boldsymbol{A}\\&=\operatorname{rank}\boldsymbol{A}^{\top}\boldsymbol{M}^{\frac{1}{2}}\boldsymbol{M}^{\frac{1}{2}}\boldsymbol{A}\\&=\operatorname{rank}\left(\boldsymbol{A}^{\top}\boldsymbol{M}^{\frac{1}{2}}\right)\left(\boldsymbol{A}^{\top}\boldsymbol{M}^{\frac{1}{2}}\right)^{\top}\\&=\operatorname{rank}\left(\boldsymbol{A}^{\top}\boldsymbol{M}^{\frac{1}{2}}\right)\\&=\operatorname{rank}\boldsymbol{A}^{\top}\end{aligned}$$

因为 $\sum_{i=1}^{N} m_i\boldsymbol{\mu}_i=\left(\sum_{i=1}^{N} m_i\right)\boldsymbol{\mu}$, 即 $\sum_{i=1}^{N} m_i(\boldsymbol{\mu}_i-\boldsymbol{\mu})=\mathbf{0}$, 所以 $\operatorname{rank}\boldsymbol{A}^{\top}\leq N-1$。

作业 3.4

式 3.44 是 $\max_{\boldsymbol{W}}\frac{\operatorname{tr}(\boldsymbol{W}^{\top}\boldsymbol{S}_b\boldsymbol{W})}{\operatorname{tr}(\boldsymbol{W}^{\top}\boldsymbol{S}_w\boldsymbol{W})}$，如果 $\boldsymbol{W}$ 是一个解，那么 $\alpha\boldsymbol{W}$,$\alpha\in\mathbb{R}$ 也是一个解，于是可固定 $\operatorname{tr}(\boldsymbol{W}^{\top}\boldsymbol{S}_w\boldsymbol{W})=1$，求解 $-\operatorname{tr}(\boldsymbol{W}^{\top}\boldsymbol{S}_b\boldsymbol{W})$ 的最小值。

由拉格朗日乘子法, 定义拉格朗日函数

$$L(\boldsymbol{W},\lambda)=-\operatorname{tr}\left(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{W}\right)+\lambda\left(\operatorname{tr}\left(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{W}\right)-1\right).$$

对上式关于 $\boldsymbol{W}$ 求偏导得

$$\begin{aligned}\frac{\partial L(\boldsymbol{W},\lambda)}{\partial\boldsymbol{W}}&=-\frac{\partial\left(\operatorname{tr}\left(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{W}\right)\right)}{\partial\boldsymbol{W}}+\lambda\frac{\partial\left(\operatorname{tr}\left(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{W}\right)-1\right)}{\partial\boldsymbol{W}}\\&=-\left(\boldsymbol{S}_b+\boldsymbol{S}_b^{\mathrm{T}}\right)\boldsymbol{W}+\lambda\left(\boldsymbol{S}_w+\boldsymbol{S}_w^{\mathrm{T}}\right)\boldsymbol{W}\\&=-2\boldsymbol{S}_b\boldsymbol{W}+2\lambda\boldsymbol{S}_w\boldsymbol{W}\end{aligned}$$

令 $L(\boldsymbol{W},\lambda)=0$ 可得 $\boldsymbol{S}_b\boldsymbol{W}=\lambda\boldsymbol{S}_w\boldsymbol{W}$。

作业 3.5

对称性:

$$\begin{aligned}(\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)^\top &= (\boldsymbol{X}^\top)^\top((\boldsymbol{X}^\top\boldsymbol{X})^{-1})^\top\boldsymbol{X}^\top \\ &= (\boldsymbol{X}^\top)^\top((\boldsymbol{X}^\top\boldsymbol{X})^\top)^{-1}\boldsymbol{X}^\top \\ &= \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\end{aligned}$$

幂等性:

$$\begin{aligned}(\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)^2 &= \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top \\ &= \boldsymbol{X}\boldsymbol{I}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top \\ &= \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\end{aligned}$$

所以矩阵 $\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ 是投影矩阵。

将特征矩阵 $\boldsymbol{X}$ 看作是一个由 d 个 n 维列向量组成的向量组。假设 $d < n$ 且所有列向量都线性无关,那 $\boldsymbol{X}$ 张成的空间是 d 维度空间。真实值 $\boldsymbol{y}$ 是一个 n 维空间中的 $n \times 1$ 向量。线性回归就是在 $\boldsymbol{X}$ 张成的 d 维空间中,寻找 n 维空间中 $\boldsymbol{y}$ 的投影,也就是一种降维的操作。