assignment3

Xiaoma

2022 年 10 月 23 日

题目 1.

解:

$$
\begin{aligned}
y(w) &= \frac{1}{1+e^{-(w^{\mathbf{T}}x+b)}} \\
\nabla_w y(w) &= \frac{e^{-(w^{\mathbf{T}}x+b)}}{(1+e^{-(w^{\mathbf{T}}x+b)})^2} \times x \\
&= y(w)(1-y(w))x \\
\nabla_w^2 y &= y(w)(1-y(w))^2 x x^{\mathbf{T}} - y(w)^2(1-y(w)) x x^{\mathbf{T}} \\
&= y(w)(1-y(w))(1-2y(w)) x x^{\mathbf{T}}
\end{aligned}
$$

当 $\frac{1}{2} < y(w) < 1$ 时，$\nabla_w^2 y(w) < 0$，则 $y(w)$ 非凸。

$$
\begin{aligned}
l(w) &= \sum_{i=1}^{m} (-y_i(w^{\mathbf{T}}x+b) + \ln(1+e^{w^{\mathbf{T}}x_i+b})) \\
\nabla_w l(w) &= \sum_{i=1}^{m} (-y_i x_i + \frac{e^{w^{\mathbf{T}}x_i+b}}{1+e^{w^{\mathbf{T}}x_i+b}} x_i) \\
&= \sum_{i=1}^{m} (-y_i x_i + \sigma(w^{\mathbf{T}}x_i+b) x_i) \\
\nabla_w^2 l(w) &= \sum_{i=1}^{m} (\sigma(w^{\mathbf{T}}x_i+b)(1-\sigma(w^{\mathbf{T}}x_i+b)) x_i x_i^{\mathbf{T}}) > 0
\end{aligned}
$$

则 $l(w)$ 是凸的。

题目 2.

解:

Class	f0	f1	f2	f3	f4	f5	f6	f7	f8
C1	1	1	1	1	1	1	1	1	1
C2	0	0	0	0	1	1	1	0	1
C3	0	0	1	1	0	0	1	1	0
C4	0	1	0	1	0	1	0	0	0

该 ECOC 二元码的海明距离为 5，满足行分离与列分离的条件。

题目 3.

解:

$$
\begin{aligned}
\mathbf{S}_b &= \mathbf{S}_t - \mathbf{S}_w \\
&= \sum_c m_c(\mu_c - \mu)(\mu_c - \mu)^{\mathbf{T}} \\
&= [(\mu_1 - \mu), \cdots, (\mu_N - \mu)] \begin{bmatrix} m_1 & & \\ & \ddots & \\ & & m_N \end{bmatrix} \begin{pmatrix} (\mu_1 - \mu)^{\mathbf{T}} \\ \vdots \\ (\mu_N - \mu)^{\mathbf{T}} \end{pmatrix}
\end{aligned}
$$

$$
rank(\mathbf{S}_b) = rank([(\mu_1 - \mu), \cdots, (\mu_N - \mu)] \begin{bmatrix} m_1 & & \\ & \ddots & \\ & & m_N \end{bmatrix} \begin{pmatrix} (\mu_1 - \mu)^{\mathbf{T}} \\ \vdots \\ (\mu_N - \mu)^{\mathbf{T}} \end{pmatrix})
$$

设 $\mathbf{M} = \begin{bmatrix} m_1 & & \\ & \ddots & \\ & & m_N \end{bmatrix}$，$\mathbf{U} = [(\mu_1 - \mu), \cdots, (\mu_N - \mu)]$

$$\begin{aligned} rank(\mathbf{S}_b) &= rank(\mathbf{U}\mathbf{M}\mathbf{U}^{\mathbf{T}}) \\ &= rank(\mathbf{U}\mathbf{M}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\mathbf{U}^{\mathbf{T}}) \\ &= rank(\mathbf{U}\mathbf{M}^{\frac{1}{2}})(\mathbf{U}\mathbf{M}^{\frac{1}{2}})^{\mathbf{T}} \\ &= rank(\mathbf{U}\mathbf{M}^{\frac{1}{2}}) \\ &= rank(\mathbf{U}) \end{aligned}$$

已知 $\sum_c m_c(\mu_c - \mu) = 0$，则 $rank(\mathbf{S}_b) \leq N - 1$。

题目 4.

解: 将问题转化为

$$J(\mathbf{W}) = Tr\{(\mathbf{W}^{\mathbf{T}}\mathbf{S}_{\mathbf{W}}\mathbf{W})^{-1}(\mathbf{W}^{\mathbf{T}}\mathbf{S}_{\mathbf{B}}\mathbf{W})\}$$

由 $\mathbf{S}_{\mathbf{B}}$ 的定义可知，$\mathbf{S}_{\mathbf{B}}$ 由 K 个矩阵的和组成，秩为 1，由 $m = \frac{1}{N}\sum_{n=1}^{N} x_n$，这些矩阵中只有 (K-1) 个是相互独立的。
所以 $rank(\mathbf{S}_{\mathbf{B}}) \leq (N-1)$，由 $\mathbf{S}_{\mathbf{B}}$ 张成的 (K-1) 维空间上的投影不会改变 $J(\mathbf{W})$ 的值，所以原问题需满足 $\mathbf{S}_{\mathbf{B}}\mathbf{W} = \lambda\mathbf{S}_{\mathbf{W}}\mathbf{W}$，$\mathbf{W}$ 的闭式解则是 $\mathbf{S}_{\mathbf{W}}^{-1}\mathbf{S}_{\mathbf{B}}$ 的 d' 个最大非零广义特征值所对应的特征向量组成的矩阵，$d' \leq (N-1)$。

题目 5.

解:

$$\begin{aligned} \mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}} &= \mathbf{X}(\mathbf{X}^{\mathbf{T}})^{-1}\mathbf{X}^{-1}\mathbf{X}^{\mathbf{T}} \\ &= \mathbf{X}(\mathbf{X}^{\mathbf{T}})^{-1}(\mathbf{X}(\mathbf{X}^{\mathbf{T}})^{-1})^{-1} \end{aligned}$$

设 $\mathbf{A} = \mathbf{X}(\mathbf{X}^{\mathbf{T}})^{-1}$，则证明 $\mathbf{A}\mathbf{A}^{-1}$ 是投影矩阵。

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

满足投影矩阵的性质，则 $\mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}$ 为投影矩阵。

我们可以将特征矩阵 X 看作是一个向量组，每一列（特征）都是一个 n 维向量，我们有 d 个这样的向量。我们假设 $d < n$ 且所有特征都线性无关，那 X 张成的空间是个 d 维度空间。真实值 y 是一个 $n \times 1$ 的向量，处于 n 维空间中。多元线性回归就是在 X 张成的 d 维空间中，寻找 n 维空间中 y 的投影。