

机器学习课程 第3次作业

黄昊 20204205

选择3.1 3.2 3.6 3.7 3.9 .

3.1 b 被确认是一个常数，与输入无关时。此时对线性模型进行变换： $y_{new} = y_{old} - b = \omega x$ 后即可不考虑偏置项 b

3.2

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

由半正定矩阵定义易知 $x x^T$ 是半正定的。又因为 $(1 - 2y)(y - y^2)$ 显然在 $y = 1/2$ 时为分界点，则Hessian矩阵不能在 y 的取值范围内一直是半正定的，从而目标函数是非凸的。

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

显然Hessian是正定的，故对数似然函数是凸的。

3.6 对所有样本进行一个非线性变换，映射到高维空间，使得映射后的样本满足线性可分的条件。

3.7 由于ECOC二代码在长度为9的情况下，最大海明距离为9，则满足最优性的编码中，类与类之间同时能够满足的最大的海明距离为6，编码策略为：

- 110110110
- 101101101
- 011011011
- 000000000

下面给出构造过程：考虑码长为3，类别为4的情况，其最优意义下（即每条编码的距离尽量远）的编码为：000，011，101，110。容易验证最小的海明距离最大为2。将此种编码重复三次，就得到了码长为9的情况。容易验证，上面所设计的编码方案，任意两个ECOC编码的海明距离为6，这意味着没有任何可能找到更优解。（若存在，则需要修改数字0或者1，但上面所提编码方案若再进行修改，只会引起最小的海明距离的最大值降低。）

3.9 因为对于OvR和MvM来说，不同类的处理方法相同，由于对称性，不平衡性产生的影响可以相互抵消。