

HW8

PB21111686_赵卓

1.

- 证明：假设不存在与训练集一致的决策树，那么训练集训练得到的决策树至少存在一个节点上存在无法划分的多个数据，这是因为如果节点上没有冲突数据，那么总是能够将数据分开的。这和不冲突数据的前提相矛盾，因此一定存在与训练集一致的决策树。

2.

- (1)
选择 $w^T D w$ 而非L2规范化项的原因是L2对所有特征权重进行相同调整，而前者对不同特征可以进行不同调整适应需求。其中 D_{ii} 表示特征的权重，值越大，权重越大。
- (2)
 $J(w) = (Xw - y)^T (Xw - y) + \lambda w^T D w$
则 $\frac{\partial J}{\partial w} = 2X^T(Xw - y) + 2\lambda D w$
令 $2X^T(Xw - y) + 2\lambda D w = 0$
有 $X^T X w + \lambda D w = X^T y$
解得 $w = (X^T X + \lambda D)^{-1} X^T y$

3.

- (1)
 $K_{i,j} = K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$
 $K_{j,i} = K(x_j, x_i) = \varphi(x_j) \cdot \varphi(x_i)$
则 $K_{i,j} = K_{j,i}$ ，即 K 是一个对称矩阵
- (2)
令 $z = (z_1, z_2 \dots z_n)$
则 $z^T K z = \sum_{i=1}^n \sum_{j=1}^n z_i z_j K_{i,j}$
 $= \sum_{i=1}^n \sum_{j=1}^n z_i z_j K_{j,i} \geq 0$
即 K 是一个正定矩阵

4.

- 原始形式 $\min_{w_1, b} \frac{1}{2} \|w\|^2 = \frac{1}{2} (w_1^2 + w_2^2)$

则有：

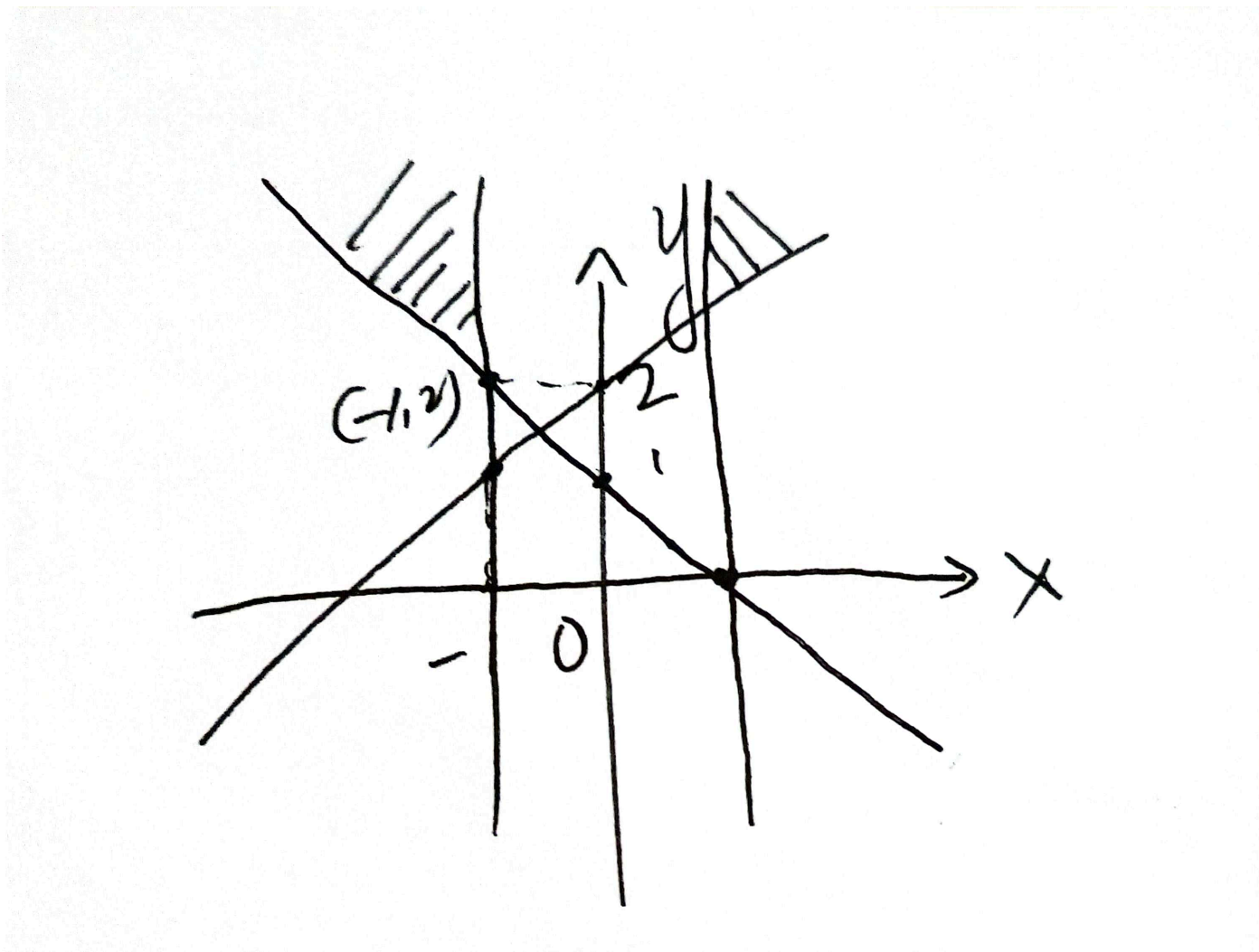
$$\begin{cases} w_1 + 2w_2 + b \geq 1 \\ 2w_1 + 3w_2 + b \geq 1 \\ 3w_1 + 3w_2 + b \geq 1 \\ -2w_1 - w_2 - b \geq 1 \\ -3w_2 - 2w_2 - b \geq 1 \end{cases}$$

解得

$$\begin{cases} -w_1 + w_2 \geq 2 \\ w_1 \leq -1 \\ w_2 \geq 1 \\ w_1 + 2w_2 \geq 2 \\ w_2 \geq 2 \end{cases}$$

由此可得 $w_1 = -1, w_2 = 2, b = -2$

如图所示：



5.

$$\begin{aligned} \bullet \frac{\partial}{\partial w_j} L_{CE}(w, b) &= -(y \frac{1}{\sigma(wx+b)} \sigma(wx+b)(1 - \sigma(wx+b))x + (1 - y) \frac{1}{1 - \sigma(wx+b)} \sigma(wx+b)(1 - \sigma(wx+b))x) \\ &= (\sigma(wx+b) - y)x \end{aligned}$$

6.

- $K - means$ 算法一定收敛。
- 证明：将目标函数记为 $f(T)$ ， T 是对给定数据集的一种划分方式，设 T 将原数据集划分为 $\{w_1, w_2, \dots, w_n\}$ ，则 $f(T) = \sum_{i=1}^k \rho(w_i) = \sum_{i=1}^k (x - x^*)'(x - x^*)$ ，其中 x^* 是 i 类数据的中心， $f(T) \geq 0$ 。从任意一个初始方式开始不断进行划分迭代，我们可以得到对数据集的一系列划分 T_1, T_2, \dots ，以及一系列目标函数 $f(T_1), f(T_2), \dots$ ，由于 $f(T_n) \geq 0$ 且递减，因此 $f(T_n)$ 收敛，即 $K - means$ 算法一定收敛。