

## 第1讲

(1) 查准率更重要, 因为将有病病人判断为无病病人的风险比较大, 而查全率关注的更多而是预测的准确性。

(2) 查全率. 对推荐系统来说, 更重要的应该是系统推荐的有多少是你真正需要的. 如果查全率比较低会降低用户体验, 造成用户流失。

2.

(1)

模型的泛化能力指的是在训练集上表现的很好的模型在测试集或者真实数据上的表现, 如果表现得较好, 说明模型的泛化能力好, 反之泛化能力差。

模型的过拟合指的是模型在训练集上表现得过于优异, 然而在测试集或者真实数据上表现得很一般。

一般来说, 过拟合的模型泛化能力就一般或者比较差。

(2) 模型的

2.

线性回归:

模型:  $h_0(x) = \theta^T x$ 目标函数:  $J_0(x) = \frac{1}{2n} \sum_{i=1}^n (h_0(x^{(i)}) - y^{(i)})^2$  $J_0(x) = \frac{1}{2n} \sum_{i=1}^n (h_0(x^{(i)}) - y^{(i)})^2$ 

求解算法:

梯度下降算法/正规方程

逻辑回归:

模型:  $h_0(x) = \frac{1}{1 + e^{-\theta^T x}}$  $J_0(x) = -\frac{1}{n} \sum_{i=1}^n [y^{(i)} \log(h_0(x^{(i)})) + (1 - y^{(i)})$  $\log(1 - h_0(x^{(i)}))]$ 

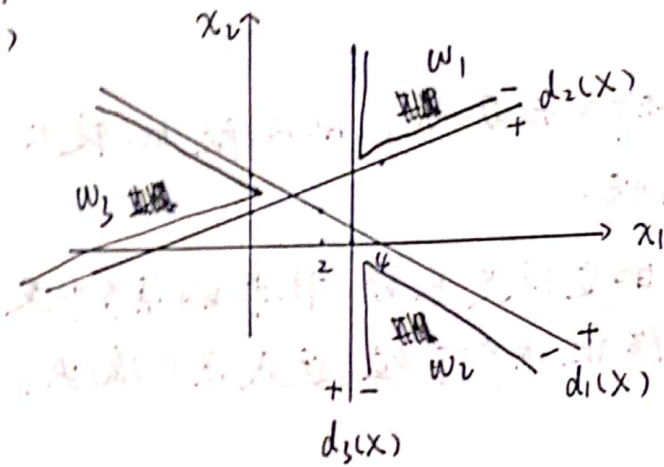
梯度下降算法。

在模型设定上, 线性回归的模型在逻辑回归中常要经过 sigmoid 函数, 因此线性回归模型常用来做预测而逻辑回归常用来做分类。

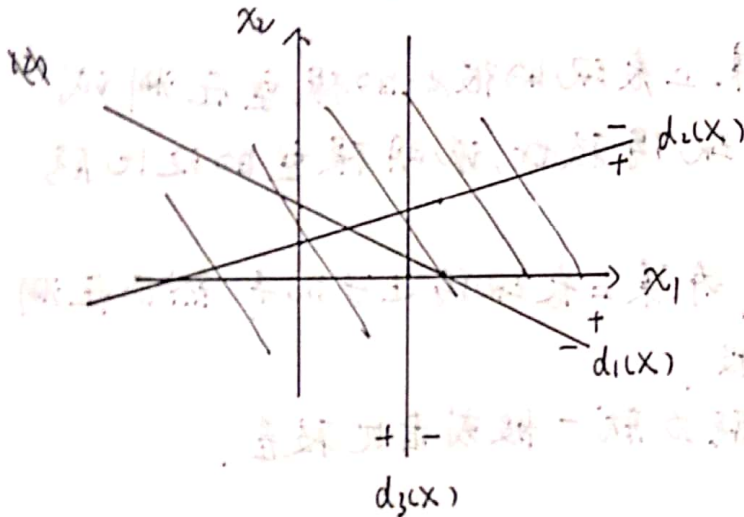
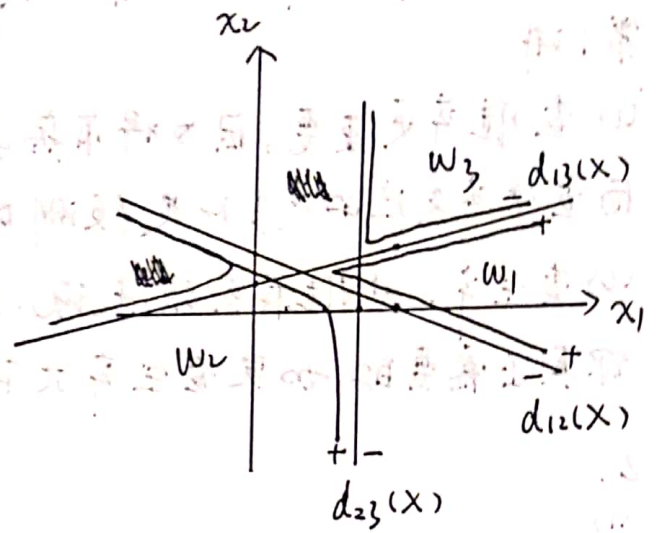
线性回归的损失函数为均方差函数, 而逻辑回归的是交叉熵函数, 但两者都是凸函数, 因此都可以使用梯度下降算法或一些更优的凸优化算法求解。但线性回归更常用正规方程求权重向量。



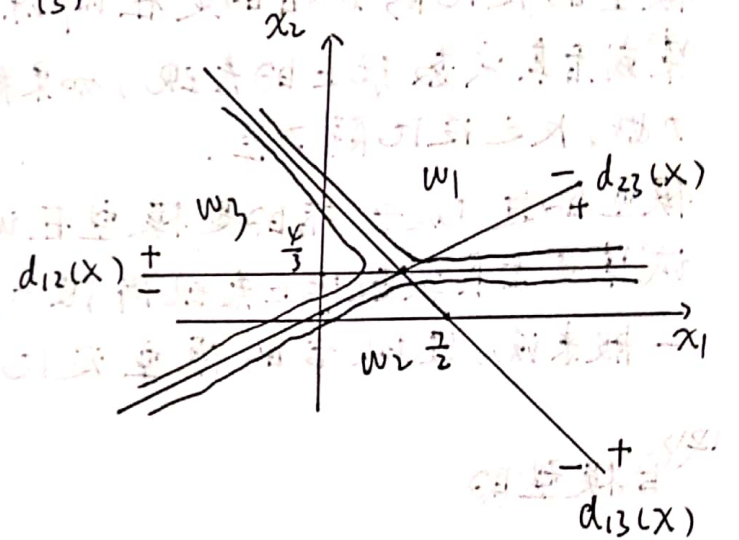
3.  
(1)



(2)



(3)



4.

$$P(w_1) \times P(x|w_1) = 0.04$$

$$P(w_2) \times P(x|w_2) = 0.18$$

该细胞为正常细胞

5.

$$\theta_{12} = \frac{(L_{12} - L_{22}) \cdot P(w_2)}{(L_{21} - L_{11}) \cdot P(w_1)} = \frac{0.9}{0.6} = 1.5$$

$$l_{12}(x) = \frac{P(x|w_1)}{P(x|w_2)} = 2$$

$$\theta_{12} < l_{12}(x), \quad x \in w_1$$

即为病变细胞



扫描全能王 创建