



中山大學
SUN YAT-SEN UNIVERSITY

第3章 线性模型

1. 基本形式
2. 回归问题
3. 分类问题

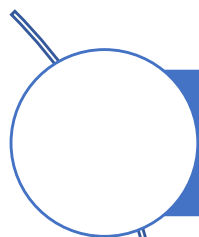


中山大學
SUN YAT-SEN UNIVERSITY

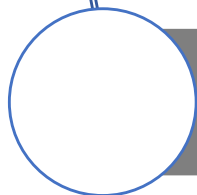
第3章 线性模型

1. 基本形式
2. 回归问题
3. 分类问题

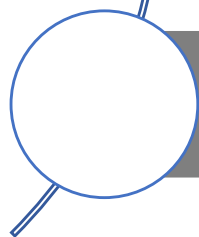
分类问题



(1) 什么是分类问题

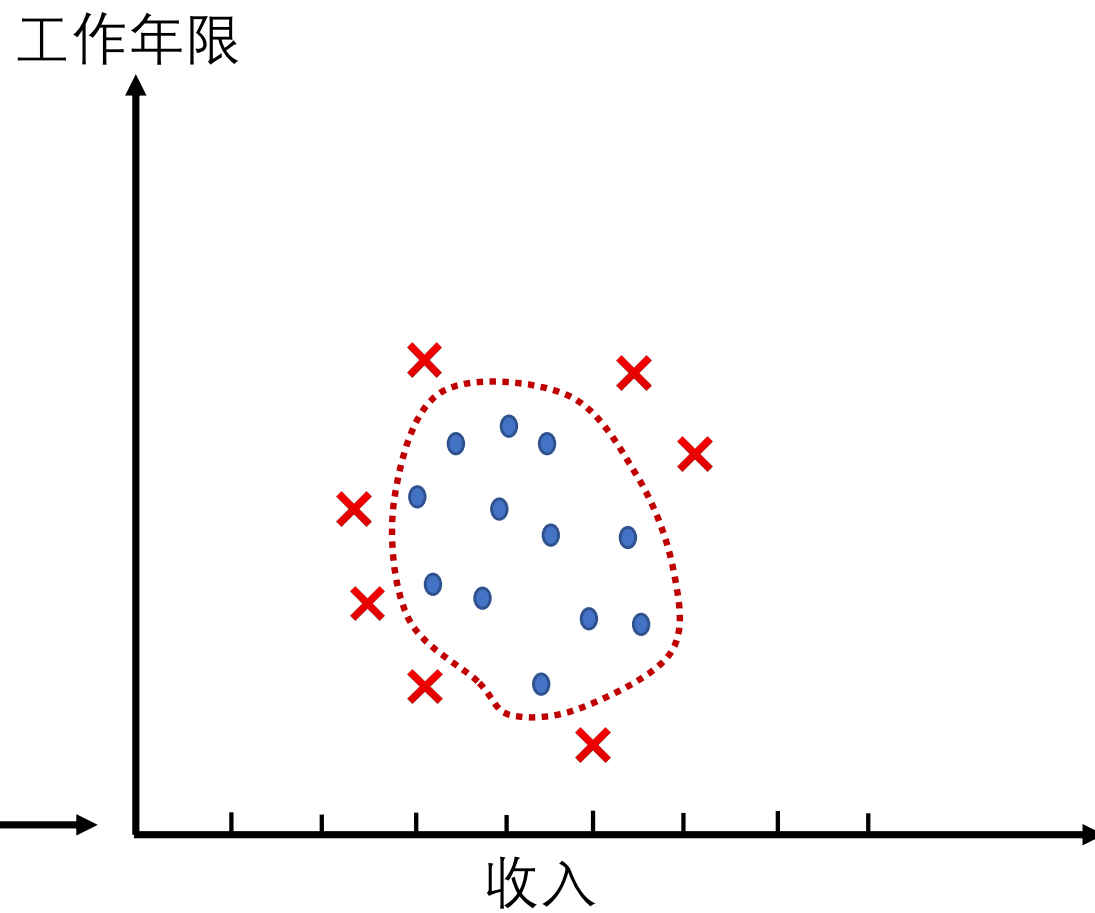
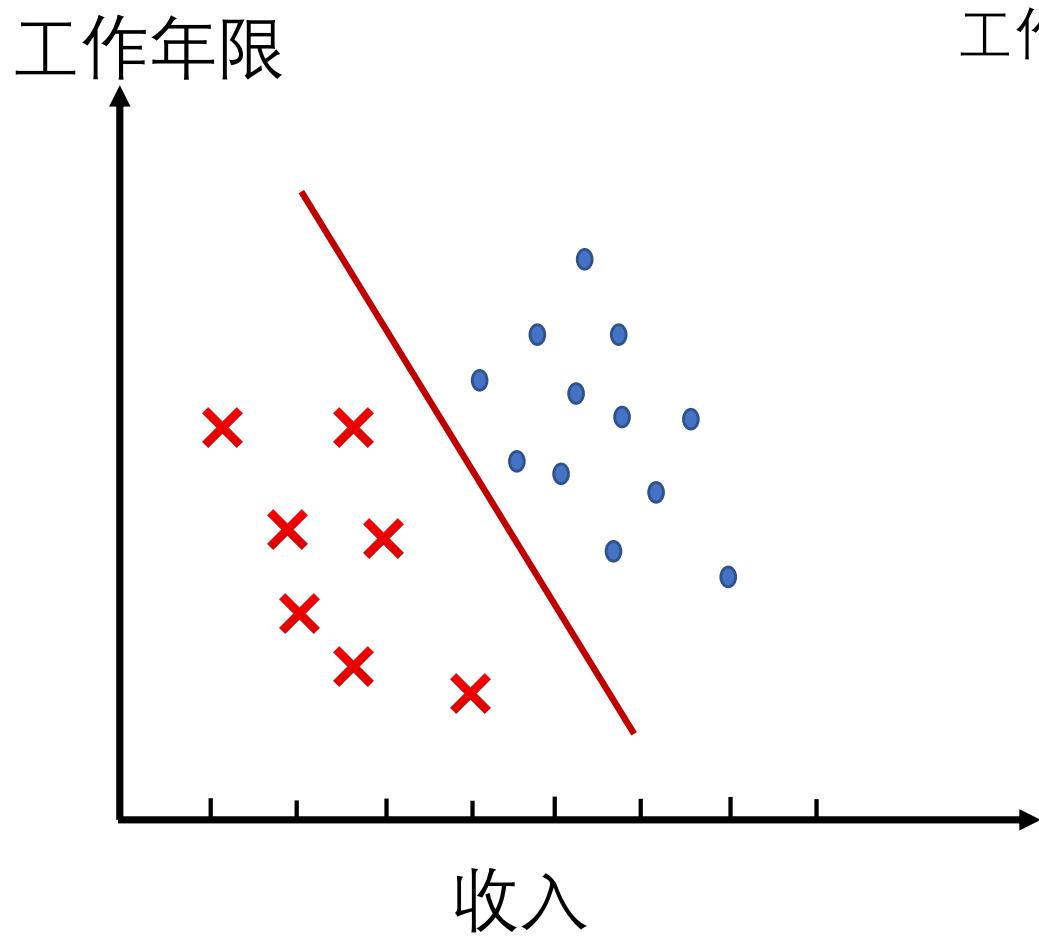


(2) 对数几率回归模型建立



(3) 多分类学习

分类问题



分类问题

- 已知:

- 训练集 $X, x \in X$
- 分类 $C = \{c_1, c_2, \dots, c_n\}$

- 决定:

- X 到 C 的一个映射关系
- 二分类: $C = \{0, 1\}$ 或 $C = \{-1, +1\}$

- 概率:

$$P_{\omega}(y|x) \quad P(y|x; \omega)$$

$$P(y = 1|x; \omega) \quad P(y = 0|x; \omega) = 1 - P(y = 1|x; \omega)$$

怎么转化成 $\{0,1\}$

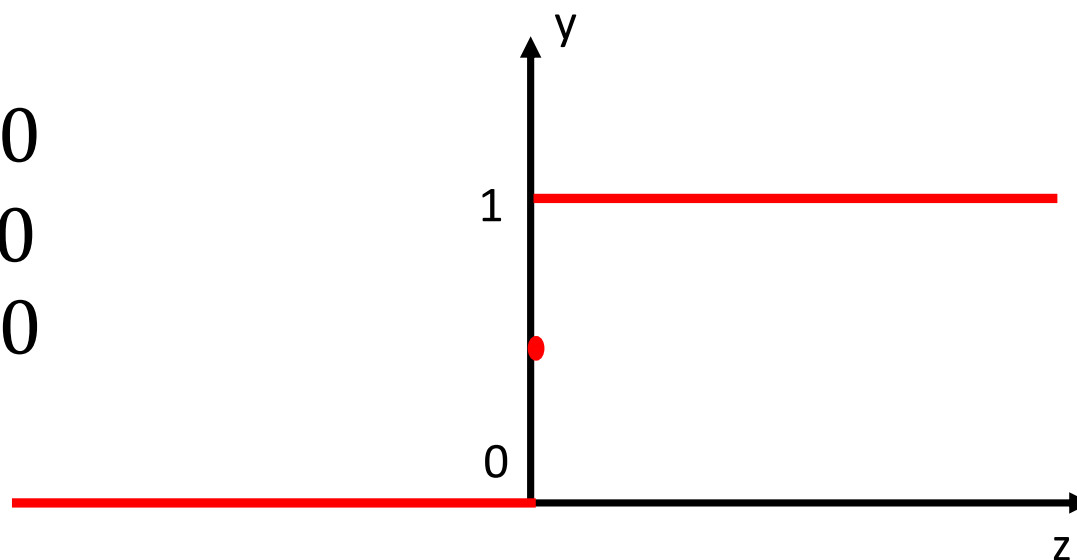
- 线性回归输出

$$f(\mathbf{x}) = \boldsymbol{\omega}^T \mathbf{x} + b = z$$

$$g(z) \in \{0,1\}$$

- 最理想的函数——单位阶跃函数

$$g(z) = \begin{cases} 0, & z < 0 \\ 0.5 & z = 0 \\ 1 & z > 0 \end{cases}$$



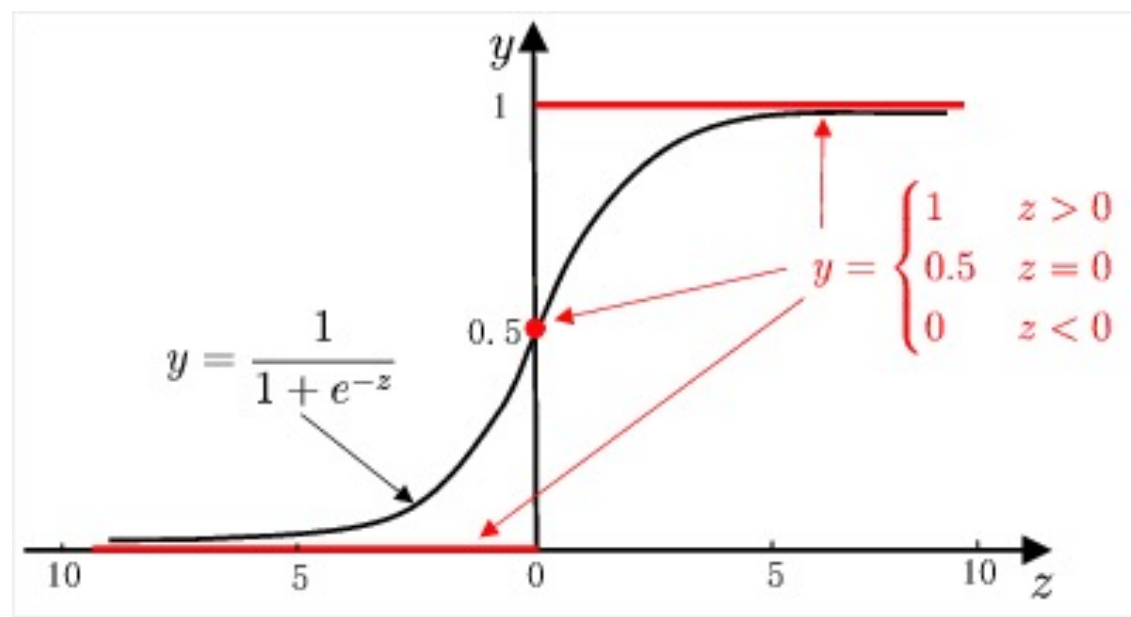
不连续

对数几率函数

- 替代函数——对数几率函数 (logistic function)
 - **优点：** 单调可微、任意阶可导

单位阶跃函数与对数几率函数的比较

$$g(z) = \frac{1}{1 + e^{-z}}$$



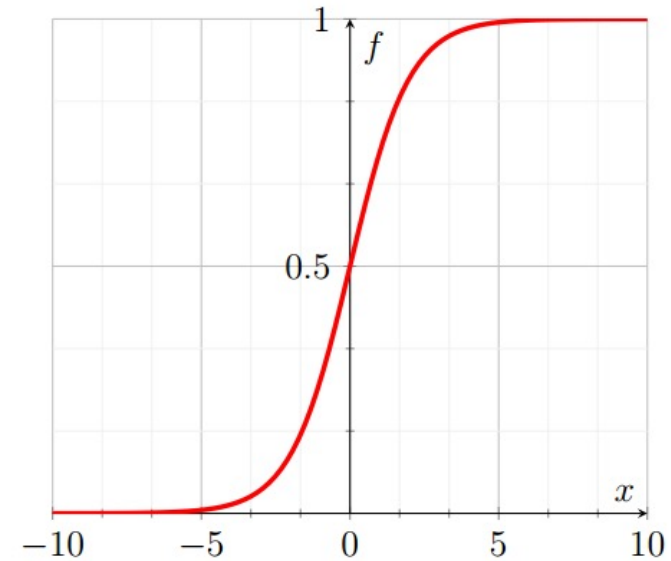
假设函数: $h_{\omega}(\mathbf{x}) = f(\mathbf{x}) = g(\boldsymbol{\omega}^T \mathbf{x} + b)$

对数几率函数性质

$$g(z) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}}$$

$$g(-\infty) = 0 \quad g(+\infty) = 1$$

$$g(0) = 0.5$$



➤ 连续，单调可微

➤ Sigmoid function of z

$$g'(z) = g(z)(1 - g(z))$$

$$\begin{aligned} g'(z) &= \frac{d}{dz} \frac{1}{1 + e^{-z}} \\ &= \frac{1}{(1 + e^{-z})^2} (e^{-z}) \\ &= \frac{1}{(1 + e^{-z})} \cdot \left(1 - \frac{1}{(1 + e^{-z})} \right) \\ &= g(z)(1 - g(z)). \end{aligned}$$

判别规则和判别边界

- 判别规则:

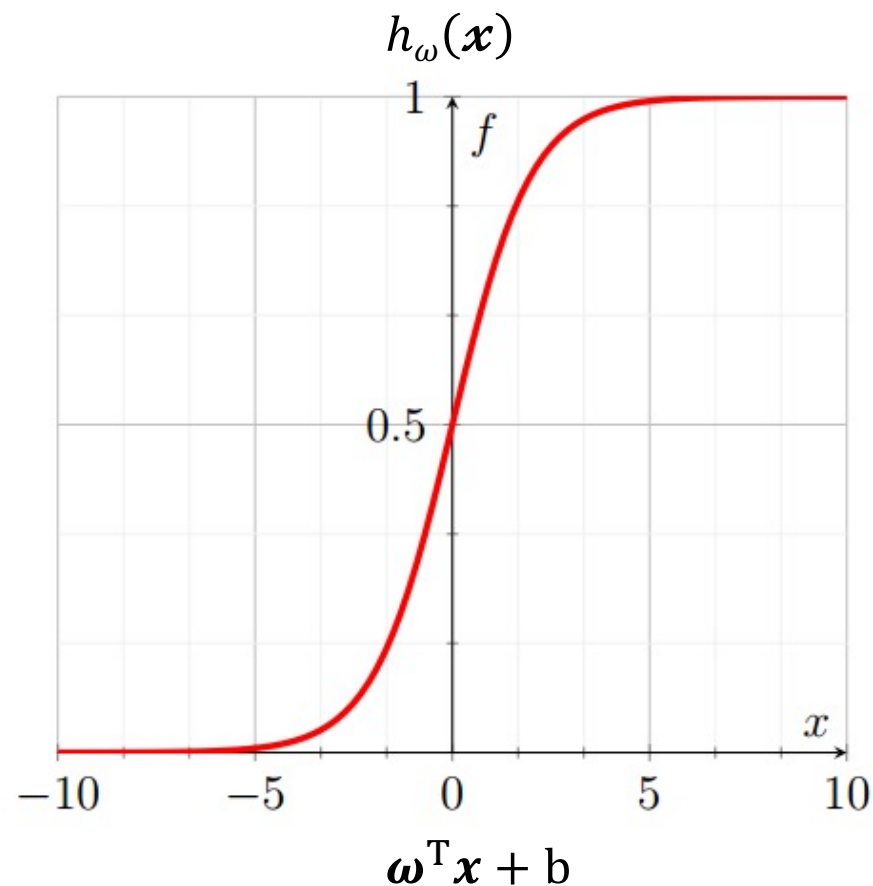
$$h_w(\mathbf{x}) = f(\mathbf{x}) = g(\boldsymbol{\omega}^T \mathbf{x} + b)$$

➤ 预测 $y=1$ $h_w(\mathbf{x}) \geq 0.5$ $\boldsymbol{\omega}^T \mathbf{x} + b \geq 0$

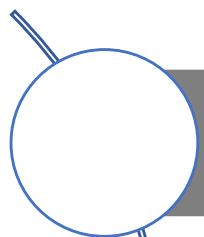
➤ 预测 $y=0$ $h_w(\mathbf{x}) < 0.5$ $\boldsymbol{\omega}^T \mathbf{x} + b < 0$

- 判别边界:

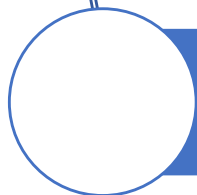
$$\boldsymbol{\omega}^T \mathbf{x} + b = 0$$



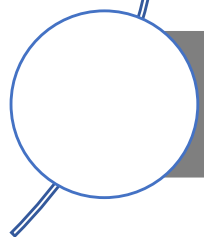
分类问题



(1) 什么是分类问题



(2) 对数几率回归模型建立



(3) 多分类学习

假设函数输出

收入：sal

发信用卡

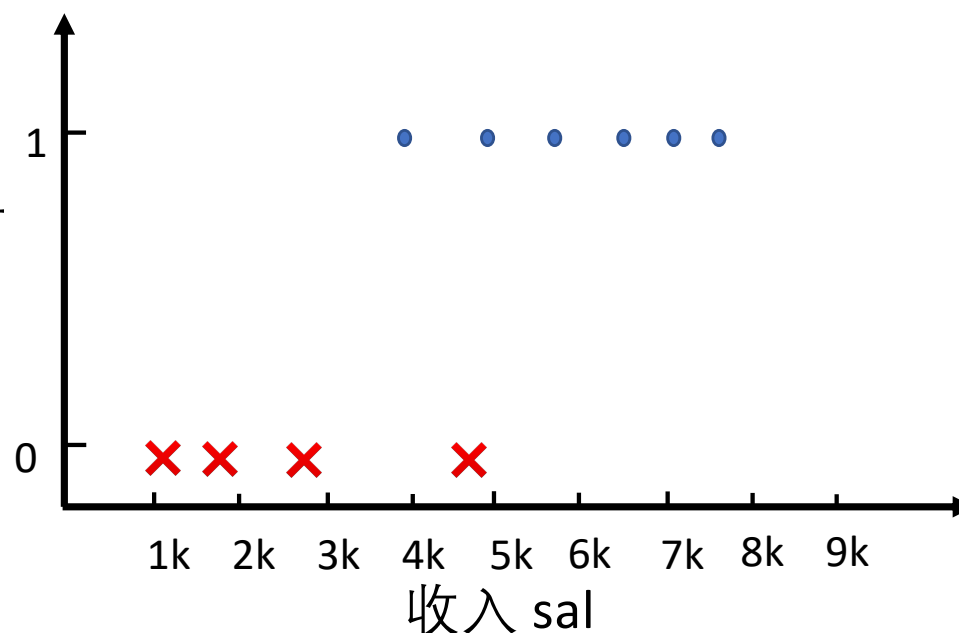
$h_{\omega}(\text{sal})$ 代表什么？

$$h_{\omega}(\text{sal}) = P(y = 1 \mid \mathbf{x}; \boldsymbol{\omega})$$

$$P(y = 0 \mid \mathbf{x}; \boldsymbol{\omega}) = 1 - P(y = 1 \mid \mathbf{x}; \boldsymbol{\omega}) = 1 - h_{\omega}(\mathbf{x})$$

$$P(y = 1 \mid \mathbf{x}; \boldsymbol{\omega}) = h_{\omega}(\mathbf{x})$$

$$P(y = 0 \mid \mathbf{x}; \boldsymbol{\omega}) = 1 - h_{\omega}(\mathbf{x})$$



$$h_{\omega}(\mathbf{x}) = f(\mathbf{x}) = \frac{1}{1 + \exp(-\boldsymbol{\omega}^T \mathbf{x})}$$

注意：此处引入 $x_0 = 1, \omega_0 = b$

Likelihood (似然)

$$h_{\omega}(\mathbf{x}) = f(\mathbf{x}) = \frac{1}{1 + \exp(-\omega^T \mathbf{x})}$$

$$P(y = 1|x; \omega) = h_{\omega}(\mathbf{x})$$

$$P(y = 0|x; \omega) = 1 - h_{\omega}(\mathbf{x})$$

- 假设我们有如下样本:

$$D = \{(x_1, +1), (x_2, 0), \dots, (x_m, 0)\}$$

样本属于其真实标签的概率

$$P(y = 1|x; \omega) \quad P(y = 0|x; \omega) \quad \dots \quad P(y = 0|x; \omega)$$

$$P(y = 1|x; \omega) \quad 1 - P(y = 1|x; \omega) \quad \dots \quad 1 - P(y = 1|x; \omega)$$



$$P(y_i|x_i; \omega) = y_i P(y = 1|x; \omega) + (1 - y_i) P(y = 0|x; \omega)$$

- Likelihood (似然)

$$L(y|\mathbf{x}; \omega) = \prod_{i=1}^m P(y_i|x_i; \omega)$$

- 找到一个合适的 $h_{\omega}(\mathbf{x})$

$$\max L(y|\mathbf{x}; \omega)$$

联乘??



对数似然

$$L(y|x; \omega) = \prod_{i=1}^m P(y_i | \mathbf{x}_i; \omega)$$

- 对数似然：

$$\mathcal{L}(y|x; \omega) = \sum_{i=1}^m \ln P(y_i | \mathbf{x}_i; \omega)$$

- 最大化对数似然：

$$\max \mathcal{L}(y|x; \omega)$$



习惯解决最小化问题

对数几率回归（Logistic Regression）问题

$$\text{Min } \mathcal{L}(y|x; \omega) = - \sum_{i=1}^m \ln P(y_i | \mathbf{x}_i; \omega)$$

交叉熵损失 (Cross Entropy Loss)

$$\text{Min } \mathcal{L}(y|x; \omega) = - \sum_{i=1}^m \ln P(y_i | \mathbf{x}_i; \omega)$$

已知：

$$P(y_i | \mathbf{x}_i; \omega) = y_i P(y = 1 | \mathbf{x}; \omega) + (1 - y_i) P(y = 0 | \mathbf{x}; \omega)$$

已知：

$$P(y = 1 | \mathbf{x}; \omega) = h_{\omega}(\mathbf{x}) \quad P(y = 0 | \mathbf{x}; \omega) = 1 - h_{\omega}(\mathbf{x})$$

已知：

$$h_{\omega}(\mathbf{x}) = \frac{\exp(\omega^T \mathbf{x})}{1 + \exp(\omega^T \mathbf{x})} \quad 1 - h_{\omega}(\mathbf{x}) = \frac{1}{1 + \exp(\omega^T \mathbf{x})}$$

代入

$$\mathcal{L}(y|x; \omega) = - \sum_{i=1}^m \ln \left(y_i \frac{\exp(\omega^T \mathbf{x})}{1 + \exp(\omega^T \mathbf{x})} + (1 - y_i) \frac{1}{1 + \exp(\omega^T \mathbf{x})} \right)$$

进一步整理

$$\mathcal{L}(y|x; \omega) = -\sum_{i=1}^m \ln \left(y_i \frac{\exp(\omega^T \mathbf{x}_i)}{1 + \exp(\omega^T \mathbf{x}_i)} + (1 - y_i) \frac{1}{1 + \exp(\omega^T \mathbf{x}_i)} \right)$$

$$= -\sum_{i=1}^m \ln \frac{y_i \exp(\omega^T \mathbf{x}_i) + 1 - y_i}{1 + \exp(\omega^T \mathbf{x}_i)}$$

$$= \sum_{i=1}^m \left(\ln(1 + \exp(\omega^T \mathbf{x}_i)) - \ln(y_i \exp(\omega^T \mathbf{x}_i) + 1 - y_i) \right)$$

xxx

$$y_i=0 \text{ 时 } \ln 1 = 0$$

阴影部分可统合为 $y_i \omega^T \mathbf{x}_i$

$$y_i=1 \text{ 时 } \ln xxx = \ln \exp(\omega^T \mathbf{x}_i) = \omega^T \mathbf{x}_i$$

$$\text{Min } \mathcal{L}(\omega) = \sum_{i=1}^m (\ln(1 + \exp(\omega^T \mathbf{x}_i)) - y_i \omega^T \mathbf{x}_i)$$

梯度下降

$$\mathcal{L}(\omega) = \sum_{i=1}^m (\ln(1 + \exp(\omega^T \mathbf{x}_i)) - y_i \omega^T \mathbf{x}_i)$$

$$\frac{\partial \mathcal{L}(\omega)}{\partial \omega_j} = \sum_{i=1}^m \left(\frac{\partial \ln((1 + \exp(\omega^T \mathbf{x}_i)))}{\partial ((1 + \exp(\omega^T \mathbf{x}_i)))} \frac{\partial ((1 + \exp(\omega^T \mathbf{x}_i)))}{\partial (\omega^T \mathbf{x}_i)} \frac{\partial (\omega^T \mathbf{x}_i)}{\partial \omega_j} - \frac{\partial (y_i \omega^T \mathbf{x}_i)}{\partial \omega_j} \right)$$

$$= \sum_{i=1}^m \left(\frac{1}{(1 + \exp(\omega^T \mathbf{x}_i))} \exp(\omega^T \mathbf{x}_i) \mathbf{x}_{ij} - y_i \mathbf{x}_{ij} \right)$$

$$= \sum_{i=1}^m (P(y_i = 1 | \mathbf{x}; \omega) \mathbf{x}_{ij} - y_i \mathbf{x}_{ij})$$

For each feature j

$$\omega_j = \omega_j - \eta \frac{1}{m} \sum_{i=1}^m (P(y_i = 1 | \mathbf{x}; \omega) \mathbf{x}_{ij} - y_i \mathbf{x}_{ij})$$



对数几率回归-预测

- $y = 1$ 的概率

$$y = h_{\omega}(x) = \frac{1}{1 + e^{-(\omega^T x + b)}}$$

- 假设 预测 $y = 1$ if

$$\omega^T x + b \geq 0$$

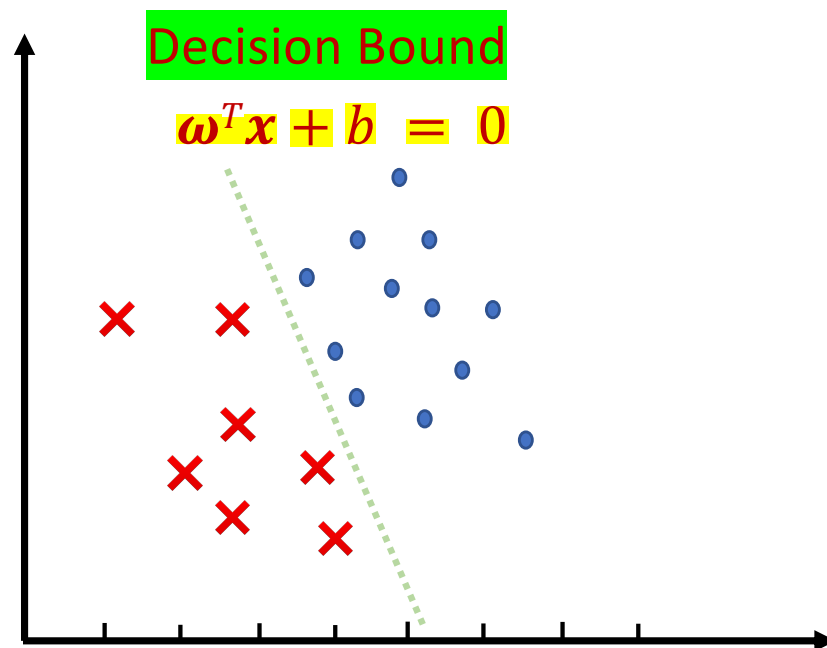
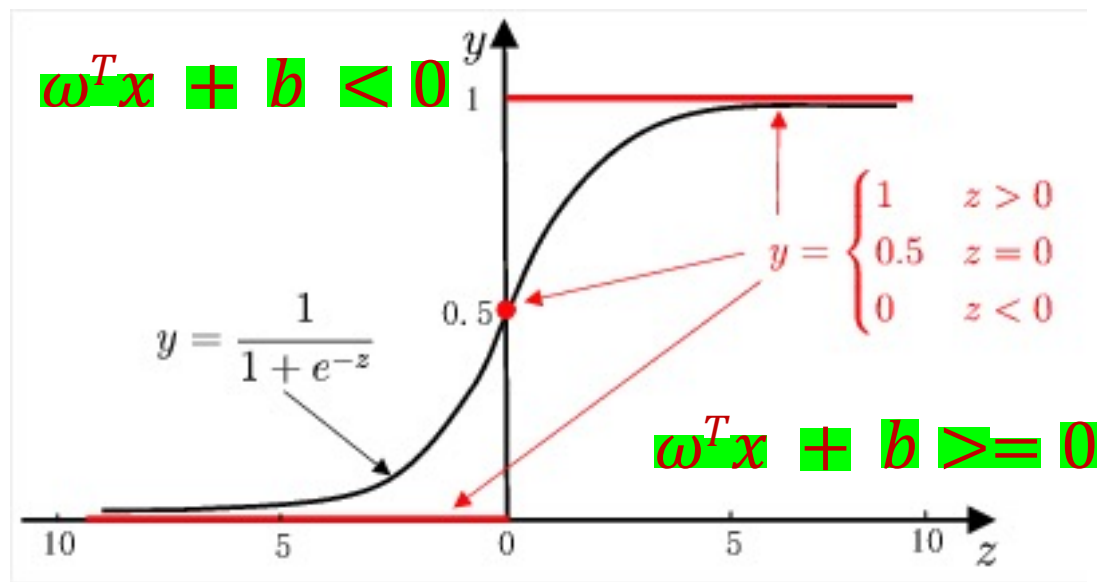
$$h_{\omega}(x) \geq 0.5$$

- 假设 预测 $y = 0$ if

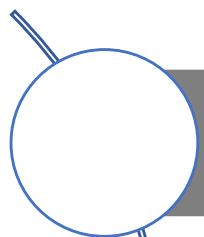
$$h_{\omega}(x) < 0.5 \quad \omega^T x + b < 0$$

$$\ln \frac{y}{1-y} = \omega^T x + b$$

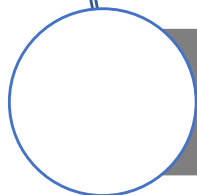
对数几率：样本作为正例的相对可能性的对数



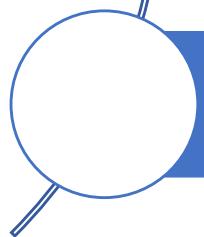
分类问题



(1) 什么是分类问题



(2) 对数几率回归模型建立

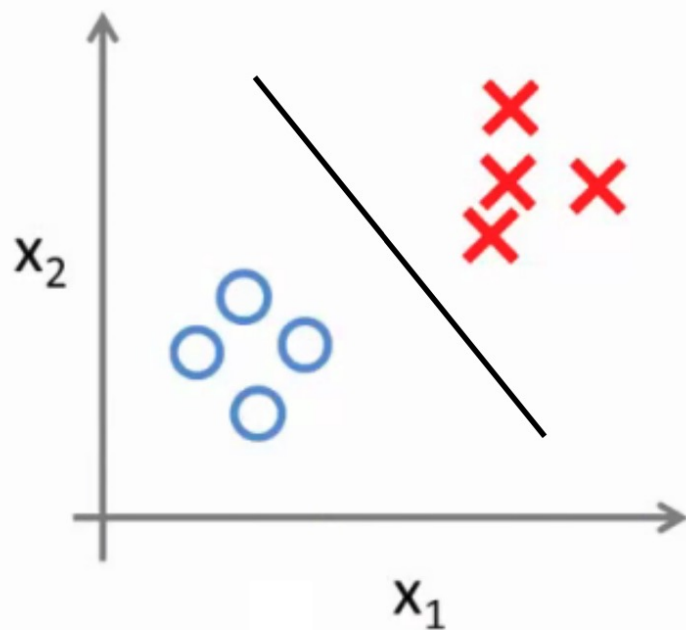


(3) 多分类学习

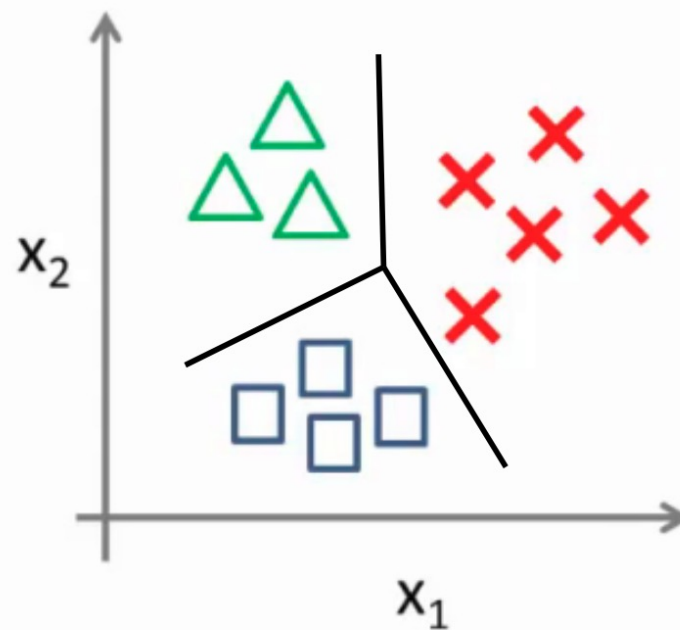
多分类学习

数据: $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$, $y_i \in \{C_1, \dots, C_N\}$

Binary classification:



Multi-class classification:



多分类学习

数据: $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$, $y_i \in \{C_1, \dots, C_N\}$

多分类学习方法

- 二分类学习方法推广到多类（如多项对数几率回归）
- 利用二分类学习器解决多分类问题（常用）
 - 对问题进行拆分，为拆出的每个二分类任务训练一个分类器
 - 对于每个分类器的预测结果进行集成以获得最终的多分类结果

拆分策略

- 一对一（One vs. One, OvO）
- 一对其余（One vs. Rest, OvR）
- 多对多（Many vs. Many, MvM）

多分类学习：一对一

数据： $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$, $y_i \in \{C_1, \dots, C_N\}$

拆分阶段

- N 个类别，两两配对（如 C_i 和 C_j ，将 C_i 作为正例 C_j 作为反例）
 - $N(N - 1)/2$ 个二类任务
- 各个二类任务，学习分类器（注意数据集的切分和使用）
 - $N(N - 1)/2$ 个二类分类器

测试阶段

- 新样本提交给所有分类器预测
 - $N(N - 1)/2$ 个分类结果
- 投票产生最终分类结果
 - 被预测最多的类别为最终类别

多分类学习：一对其余

数据： $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$, $y_i \in \{C_1, \dots, C_N\}$

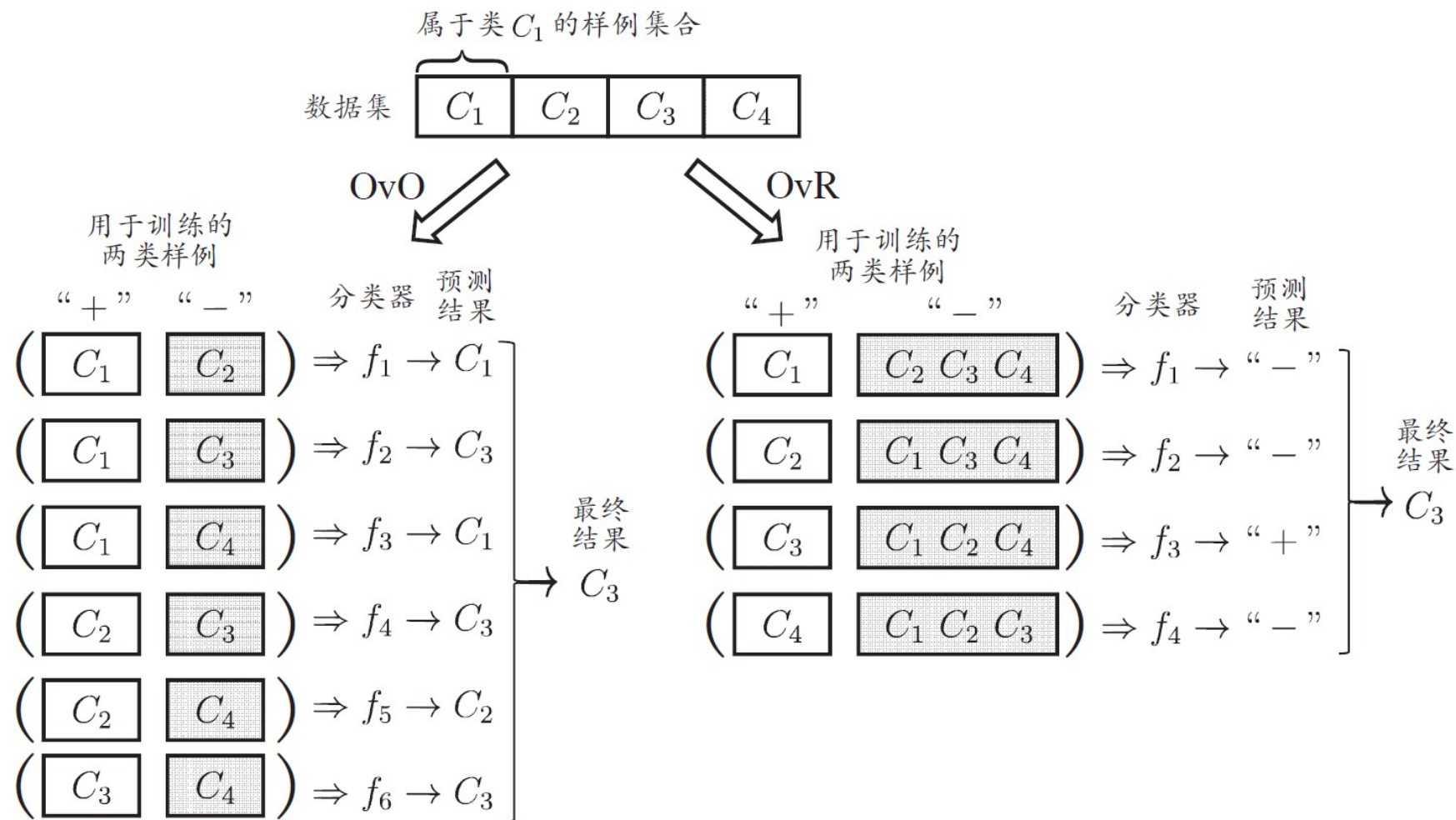
拆分阶段

- N 个类别，某一类作为正例，所有其他类作为反例
 - N 个二类任务
- 各个二类任务，学习分类器（注意数据集的使用）
 - N 个二类分类器

测试阶段

- 新样本提交给所有分类器预测
 - N 个分类结果
- 比较各分类器预测置信度
 - 置信度最大类别作为最终类别

多分类学习：两种策略比较



多分类学习：两种策略比较

一对一

- 训练 $N(N - 1)/2$ 个分类器，存储开销大
- 训练只用两个类的样例，训练时间短
- 测试时间长

一对其余

- 训练 N 个分类器，存储开销小
- 训练用到全部训练样例，训练时间长
- 测试时间短

预测性能取决于具体数据分布，多数情况下两者差不多



谢谢！