MACHINE LEARNING

机器学习

第三章：线性模型

第三章：线性模型

1. 线性回归

- 线性回归、多元线性回归

2. 二分类任务

- 广义线性回归、线性判别分析

3. 多分类任务

4. 类别不平衡问题

第三章：线性模型

重点

- 线性回归
- 线性判别分析
- 类别不平衡问题

难点

- 线性回归求解：最小二乘法
- 线性判别分析

第三章：线性模型

线性回归

- 线性模型、分线性模型
- 线性回归 及其求解（最小二乘法）
- 多元线性回归

二分类任务

多分类任务

类别不平衡问题

线性模型

线性模型

$$f(x) = \omega_1 x_1 + \omega_2 x_2 + \ldots + \omega_d x_d + b$$

例子

- 综合考虑色泽、根蒂和敲声，来判断西瓜好不好

$$f_{\text{好瓜}}(x) = 0.2 \cdot x_{\text{色泽}} + 0.5 \cdot x_{\text{根蒂}} + 0.3 \cdot x_{\text{敲声}} + 1$$

  - 根蒂的系数最大，表明根蒂最要紧
  - 敲声的系数比色泽大，说明敲声比色泽更重要

线性模型

线性模型

$$f(x) = \omega_1 x_1 + \omega_2 x_2 + \ldots + \omega_d x_d + b$$

向量形式

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

其中，$\boldsymbol{x} = (x_1, x_2, \ldots, x_d)$，$\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_d)$

- 综合考虑色泽、根蒂和敲声，来判断西瓜好不好

$$f_{好瓜}(x) = 0.2 \cdot x_{色泽} + 0.5 \cdot x_{根蒂} + 0.3 \cdot x_{敲声} + 1$$

$\boldsymbol{x} = (x_{色泽}, x_{根蒂}, x_{敲声})$，$\boldsymbol{\omega} = (0.2, 0.5, 0.3)$，$b = 1$

线性模型

线性模型 $f(\boldsymbol{x})=\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b$

优点

- 形式简单、易于建模
- 可解释性
  - $\boldsymbol{\omega}$ 直观表达了各属性在预测中的重要性
  - 线性模型有很好的可解释性

非线性模型

线性模型　$f(\boldsymbol{x}) = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b$

非线性模型的基础

- 在线性模型的基础上，引入层级结构或高维映射，构成非线性模型

非线性模型

线性模型 $f(\boldsymbol{x}) = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b$

非线性模型的基础

- 在线性模型的基础上，引入层级结构或高维映射，构成非线性模型
- 非线性模型 可以转换为 线性模型

二次多项式 $ax^2 + bx + c = (mx + p)(nx + q)$

e.g. $6x^2 + 17x + 5 = (2x + 5)(3x + 1)$

指数函数 $e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!}$

非线性模型

线性模型 $f(\boldsymbol{x}) = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b$

非线性模型的基础

- 在线性模型的基础上，引入层级结构或高维映射，构成非线性模型
- 非线性模型 可以转换为 线性模型

泰勒公式

$$f(x) = \frac{f(a)}{0!} + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

线性回归

- 给定数据集 $D=\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots (\boldsymbol{x}_m, y_m)\}$，

  其中 $\boldsymbol{x}_i = (x_{i1}; x_{i2}; \ldots; x_{id})$，$y_i \in \mathbf{R}$

离散属性处理

- 样本x的属性值间，存在 有“序”关系

- 样本x的属性值间，存在 无“序”关系

线性回归

- 给定数据集 $D=\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots (\boldsymbol{x}_m, y_m)\}$，

  其中 $\boldsymbol{x}_i=(x_{i1}; x_{i2}; \ldots; x_{id})$，$y_i \in \mathbf{R}$

离散属性处理

- 样本x的属性值间，存在 有“序”关系
  - 通过连续化，将其转化为连续值

    e.g. 二值属性“身高”的高、矮 可转化为{1,0}

- 样本x的属性值间，存在 无“序”关系

线性回归

- 给定数据集 $D=\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots (\boldsymbol{x}_m, y_m)\}$，

  其中 $\boldsymbol{x}_i = (x_{i1}; x_{i2}; \ldots; x_{id})$，$y_i \in \mathbf{R}$

离散属性处理

- 样本x的属性值间，存在 **有“序”关系**
  - 通过连续化，将其转化为连续值

    e.g. 二值属性“身高”的高、矮 可转化为$\{1,0\}$

- 样本x的属性值间，存在 **无“序”关系**
  - 有k个属性值，则转换为k维向量

    e.g. “瓜类”的取值“西瓜、南瓜、黄瓜”可转化为$\{0,0,1\}$, $\{0,1,0\}$, $\{1,0,0\}$

线性回归

- 给定数据集 $D=\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots (\boldsymbol{x}_m, y_m)\}$，
  其中 $\boldsymbol{x}_i=(x_{i1}; x_{i2}; \ldots; x_{id})$，$y_i \in \mathbf{R}$

线性回归 目的

- 学得一个线性模型 $f(\boldsymbol{x})=\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b$
  - 离散
  - 连续

线性回归

- 给定数据集 $D=\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots (\boldsymbol{x}_m, y_m)\}$，

  其中 $\boldsymbol{x}_i = (x_{i1}; x_{i2}; \ldots; x_{id})$，$y_i \in \mathbf{R}$

线性回归 目的

- 学得一个线性模型 $f(\boldsymbol{x}) = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b$
  - 离散：以尽可能准确地 预测 真实输出标记 y，即

    $$f(x) \cong y$$

  - 连续

线性回归

- 给定数据集 $D=\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots (\boldsymbol{x}_m, y_m)\}$，
  其中 $\boldsymbol{x}_i = (x_{i1}; x_{i2}; \ldots; x_{id})$，$y_i \in \mathbf{R}$

线性回归 目的

- 学得一个线性模型 $f(x) = \omega^{\mathrm{T}}x+b$
  - 离散：以尽可能准确地 预测 真实输出标记 y，即

    $$f(x) \cong y$$

  - 连续：预测输出 $f(x)$ 与真实输出 y 的偏差尽可能小

    $$\min (f(x) - y)^2$$

线性回归 目标

$f(x) = \omega^{\mathrm{T}}x+b$ 使得 $f(x) \cong y$ 或 $\min (f(x) - y)^2$

线性回归

线性回归 目标

$$f(x) = \omega^{\mathrm{T}} x + b \text{ 使得 } f(x) \cong y \text{ 或 } \min\,(f(x) - y)^2$$

单一、连续 属性　　参数/模型 估计：最小二乘法

线性回归

线性回归 目标

$f(x) = \omega^{\mathrm{T}} x + b$ 使得 $f(x) \cong y$ 或 $\min(f(x) - y)^2$

单一、连续 属性　　参数/模型 估计：最小二乘法

- 目标　$\min(f(x) - y)^2$

线性回归

线性回归 目标

$f(x) = \omega^{\mathrm{T}}x+b$ 使得 $f(x) \cong y$ 或 $\min(f(x)-y)^2$

单一、连续 属性　　参数/模型 估计：最小二乘法

- 目标 $\min(f(x)-y)^2$

- m个样本，只有一个属性 x，目标：

$$\min\sum_{i=1}^{m}(f(x_i)-y_i)^2 = \min\sum_{i=1}^{m}(y_i-\omega x_i-b)^2$$

线性回归

线性回归 目标

$f(x)=\omega^{T}x+b$ 使得 $f(x)\cong y$ 或 $\min(f(x)-y)^2$

单一、连续 属性　　参数/模型 估计：最小二乘法

□ 目标　$\min\ (f(x)-y)^2$

- m个样本，只有一个属性 x，目标：

$$\min\sum_{i=1}^{m}(f(x_i)-y_i)^2=\min\sum_{i=1}^{m}(y_i-\omega x_i-b)^2$$

- 试图找到一条直线，使所有样本(m个) 到直线上的欧氏距离之和最小

线性回归

线性回归 目标

$f(x)=\omega^{T}x+b$ 使得 $f(x)\cong y$ 或 $\min\,(f(x)-y)^2$

单一、连续 属性　　参数/模型 估计：最小二乘法

- **目标** $\min\,(f(x)-y)^2$

- **m个样本，只有一个属性 x，目标：**

$$\min\sum_{i=1}^{m}(f(x_i)-y_i)^2=\min\sum_{i=1}^{m}(y_i-\omega x_i-b)^2$$

- **试图找到一条直线，使所有样本(m个) 到直线上的欧氏距离之和最小**
  - **目标：最小化均方误差** $E_{(\omega,b)}=\sum_{i=1}^{m}(y_i-\omega x_i-b)^2$

$$(\omega^*,b^*)=\arg\min_{(\omega,b)}\sum_{i=1}^{m}(f(x_i)-y_i)^2$$
$$=\arg\min_{(\omega,b)}\sum_{i=1}^{m}(y_i-\omega x_i-b)^2$$

线性回归

线性回归 目标

$f(x) = \omega^{\mathrm{T}} x + b$ 使得 $f(x) \cong y$ 或 $\min (f(x) - y)^2$

单一、连续 属性 参数/模型 估计：最小二乘法

- 目标 $\min (f(x) - y)^2$

$$(\omega^*, b^*) = \arg\min_{(\omega,b)} \sum_{i=1}^{m} (y_i - \omega x_i - b)^2$$

线性回归

线性回归 目标

$f(x) = \omega^{\mathrm{T}}x+b$ 使得 $f(x) \cong y$ 或 $\min (f(x)-y)^2$

单一、连续 属性　　参数/模型 估计：最小二乘法

- **目标** $\min\ (f(x) - y)^2$

$$(\omega^*, b^*) = \arg\min_{(\omega,b)} \sum_{i=1}^{m} (y_i - \omega x_i - b)^2$$

- 训练样本 仅为 全部样本的 一个小的采样
  无法求得$E_{(\omega,b)}$最小值
  只能求得$E_{(\omega,b)}$的极小值，即$E_{(\omega,b)}$一阶导数为0

线性回归

线性回归 目标

$f(x) = \omega^{\mathrm{T}} x + b$ 使得 $f(x) \cong y$ 或 min $(f(x) - y)^2$

单一、连续 属性　　参数/模型 估计：最小二乘法

- 目标 min $(f(x) - y)^2$

$$(\omega^*, b^*) = \arg\min_{(\omega,b)} \sum_{i=1}^{m} (y_i - \omega x_i - b)^2$$

求 $E_{(\omega,b)}$ 的极小值，即 $E_{(\omega,b)}$ 一阶导数为0

线性回归

线性回归 目标

$f(x) = \omega^{\mathrm{T}}x + b$ 使得 $f(x) \cong y$ 或 $\min(f(x) - y)^2$

单一、连续 属性　　参数/模型 估计：最小二乘法

□ 目标　$\min(f(x) - y)^2$

$$(\omega^*, b^*) = \arg\min_{(\omega,b)} \sum_{i=1}^{m} (y_i - \omega x_i - b)^2$$

求 $E_{(\omega,b)}$ 的极小值，即 $E_{(\omega,b)}$ 一阶导数为0

- 分别对 ω 和 b 求一阶导，令其等于0可得到 ω 和 b 最优解的闭式解

$$\frac{\partial E_{(\omega,b)}}{\partial \omega} = 2\left(\omega \sum_{i=1}^{m} {x_i}^2 - \sum_{i=1}^{m} (y_i - b) \cdot x_i\right) = 0$$

$$\frac{\partial E_{(\omega,b)}}{\partial b} = 2\left(mb - \sum_{i=1}^{m} (y_i - \omega x_i)\right) = 0$$

线性回归

线性回归 目标

$f(x) = \omega^{T}x + b$ 使得 $f(x) \cong y$ 或 $\min(f(x) - y)^2$

单一、连续 属性 参数/模型 估计：最小二乘法

- 目标 $\min(f(x) - y)^2$

$$(\omega^*, b^*) = \arg\min_{(\omega,b)} \sum_{i=1}^{m} (y_i - \omega x_i - b)^2$$

求 $E_{(\omega,b)}$ 的极小值，即 $E_{(\omega,b)}$ 一阶导数为0

- 分别对 ω 和 b 求一阶导，令其等于0可得到 ω 和 b 最优解的闭式解：

$$\omega = \frac{\sum_{i=1}^{m} y_i (x_i - \bar{x})}{\sum_{i=1}^{m} x_i^2 - \frac{1}{m}\left(\sum_{i=1}^{m} x_i\right)^2} \qquad b = \frac{1}{m}\sum_{i=1}^{m}(y_i - \omega x_i)$$

其中，$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$

第三章：线性模型

1. 线性回归

- 线性回归
- 多元线性回归

2. 二分类任务

- 广义线性回归、线性判别分析

3. 多分类任务

4. 类别不平衡问题

多元线性回归

- 给定数据集 $D=\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots (\boldsymbol{x}_m, y_m)\}$，

  其中 $\boldsymbol{x} = (x_{i1}; x_{i2}; \ldots; x_{id})$，$y_i \in \mathbf{R}$

多元线性回归 目标

- 对于任意的 x_i 满足 $f(x) = \omega x_i + b$ 使得 $f(x) \cong y_i$

多元线性回归

- 给定数据集 $D=\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots (\boldsymbol{x}_m, y_m)\}$，

  其中 $\boldsymbol{x} = (x_{i1}; x_{i2}; \ldots; x_{id})$，$y_i \in \mathbf{R}$

多元线性回归 目标

- 对于任意的 x_i 满足 $f(x) = \omega x_i + b$ 使得 $f(x) \cong y_i$

◆ 数据集 D

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} & 1 \\ x_{21} & x_{22} & \cdots & x_{2d} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{md} & 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^{\mathrm{T}} & 1 \\ \boldsymbol{x}_2^{\mathrm{T}} & 1 \\ \vdots & \vdots \\ \boldsymbol{x}_m^{\mathrm{T}} & 1 \end{pmatrix}$$

其中，矩阵X大小m×(d +1)

每行对应于一个示例（样本），该行前d个元素对应于示例的d个属性值，最后一个元素恒置为1。

多元线性回归

- 给定数据集 $D=\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots (\boldsymbol{x}_m, y_m)\}$，

  其中 $\boldsymbol{x} = (x_{i1}; x_{i2}; \ldots; x_{id})$，$y_i \in \mathbf{R}$

多元线性回归 目标

- 对于任意的 x_i 满足 $f(x) = \omega x_i + b$ 使得 $f(x) \cong y_i$

◆ 数据集 D

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} & 1 \\ x_{21} & x_{22} & \cdots & x_{2d} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{md} & 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^{\mathrm{T}} & 1 \\ \boldsymbol{x}_2^{\mathrm{T}} & 1 \\ \vdots & \vdots \\ \boldsymbol{x}_m^{\mathrm{T}} & 1 \end{pmatrix}$$

◆ 把 ω 和 b 吸收入向量形式 $\hat{\boldsymbol{\omega}}^* = (\omega, b)$

◆ 标记 表示为向量形式 $\boldsymbol{y} = (y_1; y_2; \ldots; y_m)$

多元线性回归

多元线性回归 目标 $f(x)=\omega^{\mathrm{T}}X+b$ 使得 $f(x)\cong y$

参数/模型 估计：最小二乘法

- 最小二乘法 估计 ω 和 b

$$\widehat{\boldsymbol{\omega}}^{*}=\arg\min_{\widehat{\omega}}(\boldsymbol{y}-\mathbf{X}\widehat{\boldsymbol{\omega}})^{\mathrm{T}}(\boldsymbol{y}-\mathbf{X}\widehat{\boldsymbol{\omega}})$$

令 $E_{\widehat{\omega}}=(\boldsymbol{y}-\mathbf{X}\widehat{\boldsymbol{\omega}})^{\mathrm{T}}(\boldsymbol{y}-\mathbf{X}\widehat{\boldsymbol{\omega}})$，对 $\widehat{\omega}$ 求导，令其等于0

$$\frac{\partial E_{\widehat{\omega}}}{\partial\widehat{\boldsymbol{\omega}}}=2\mathbf{X}^{\mathrm{T}}(\mathbf{X}\widehat{\boldsymbol{\omega}}-\boldsymbol{y})=0$$

多元线性回归

多元线性回归 目标 $f(x) = \omega^{\mathrm{T}}X + b$ 使得 $f(x) \cong y$

参数/模型 估计：最小二乘法

最小二乘法 估计 ω 和 b

$$\hat{\boldsymbol{\omega}}^{*} = \arg\min_{\hat{\omega}} (\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\omega}})^{\mathrm{T}} (\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\omega}})$$

$$\frac{\partial E_{\hat{\omega}}}{\partial \hat{\boldsymbol{\omega}}} = 2\mathbf{X}^{\mathrm{T}}(\mathbf{X}\hat{\boldsymbol{\omega}} - \boldsymbol{y}) = 0$$

多元线性回归

多元线性回归 目标 $f(x) = \boldsymbol{\omega}^{\mathrm{T}}X+\boldsymbol{b}$ 使得 $f(x) \cong y$

参数/模型 估计：最小二乘法

- 最小二乘法 估计 ω 和 b

$$\hat{\boldsymbol{\omega}}^* = \arg\min_{\hat{\omega}} (\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\omega}})^{\mathrm{T}}(\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\omega}})$$

$$\frac{\partial E_{\hat{\omega}}}{\partial \hat{\boldsymbol{\omega}}} = 2\mathbf{X}^{\mathrm{T}}(\mathbf{X}\hat{\boldsymbol{\omega}} - \boldsymbol{y}) = 0$$

- $\hat{\boldsymbol{\omega}}^*$最优解的闭式解
  - $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ 是 满秩矩阵或正定矩阵
  - $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ 不是 满秩矩阵

多元线性回归

多元线性回归 目标 $f(x) = \omega^{\mathrm{T}}X+b$ 使得 $f(x) \cong y$

参数/模型 估计：最小二乘法

- 最小二乘法 估计 ω 和 b

$$\hat{\boldsymbol{\omega}}^{*}= \arg\min_{\hat{\omega}}(\boldsymbol{y}-\mathbf{X}\hat{\boldsymbol{\omega}})^{\mathrm{T}}(\boldsymbol{y}-\mathbf{X}\hat{\boldsymbol{\omega}})$$

$$\frac{\partial E_{\hat{\omega}}}{\partial \hat{\boldsymbol{\omega}}} = 2\mathbf{X}^{\mathrm{T}}(\mathbf{X}\hat{\boldsymbol{\omega}}-\boldsymbol{y}) = 0$$

- $\hat{\boldsymbol{\omega}}^{*}$最优解的闭式解
  - $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ 是 满秩矩阵或正定矩阵，则
    - $\hat{\boldsymbol{\omega}}^{*} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{y}$
    - 线性回归模型为 $\boldsymbol{f}(\hat{x}_i) = \hat{x}_i{}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{y}$

多元线性回归

多元线性回归 目标 $f(x) = \omega^{T}X + b$ 使得 $f(x) \cong y$

参数/模型 估计：最小二乘法

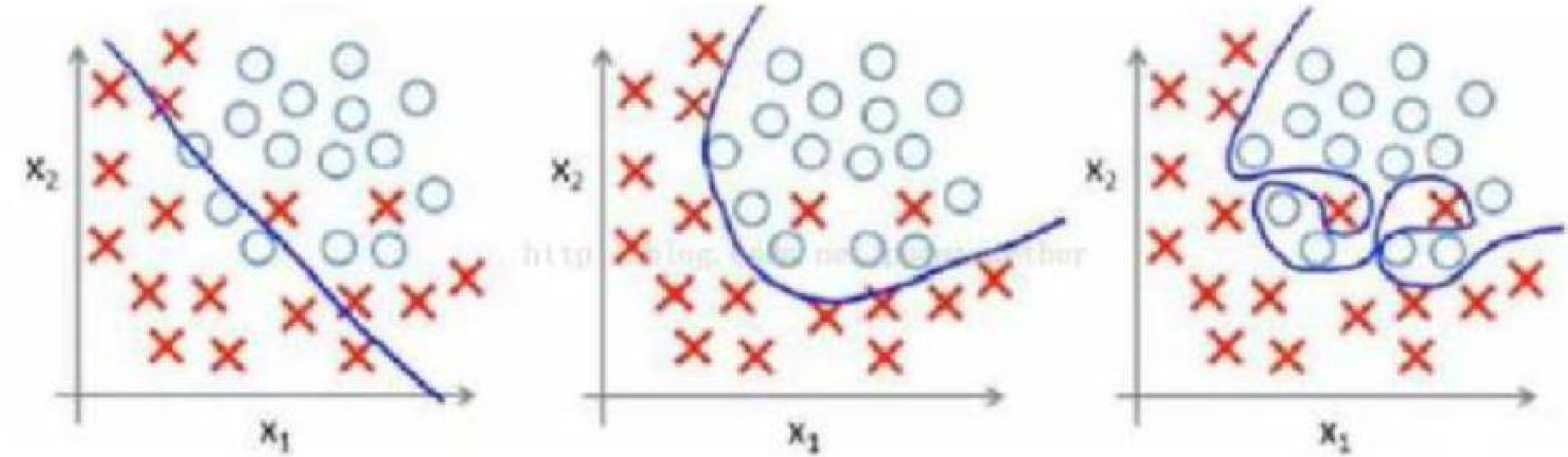
最小二乘法 估计 ω 和 b

$$\hat{\omega}^{*} = \arg\min_{\hat{\omega}} (\mathbf{y} - \mathbf{X}\hat{\omega})^{T}(\mathbf{y} - \mathbf{X}\hat{\omega})$$

$$\frac{\partial E_{\hat{\omega}}}{\partial \hat{\omega}} = 2\mathbf{X}^{T}(\mathbf{X}\hat{\omega} - \mathbf{y}) = 0$$

- $\hat{\omega}^{*}$最优解的闭式解
  - $\mathbf{X}^{T}\mathbf{X}$ 不是 满秩矩阵
    - 例如，在许多任务中我们会遇到大量的变量，其数目甚至超过样本数，导致X的列数多于行数，$\mathbf{X}^{T}\mathbf{X}$显然不满秩。
    - 此时，可解出多个ω，均能使均方误差最小化。
      - ✓ 选择哪一个解作为输出？
      - ✓ 归纳偏好选择解 (1.4节)、引入正则化(6.4、11.4节)

正则化



欠拟合　　　　恰好　　　　过拟合

高偏差　　　　　　　　　　高方差

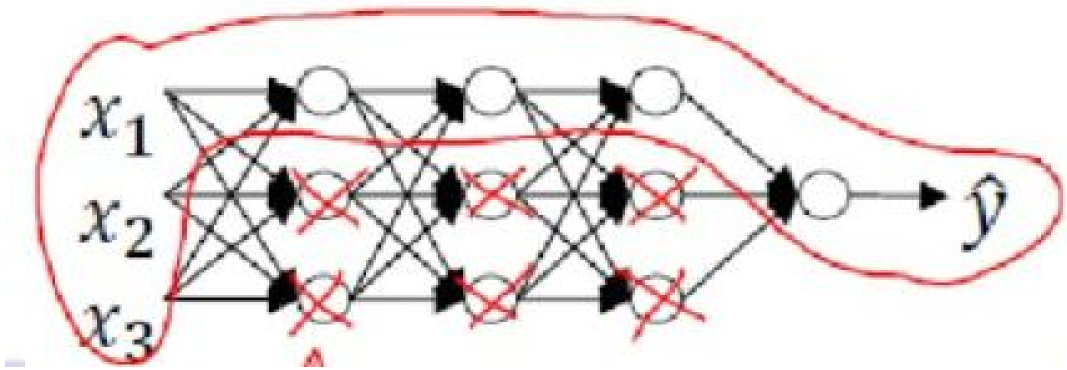
考虑如下一般形式的损失函数：

$$w^* = \arg\min_w \sum L(y_i, f(x_i; w)) + \lambda\Omega(w)$$

我们既要让训练误差（上式第一项）最小，又想让模型尽可能地简单（上式第二项）。

正则化

我们有个朴素的想法：那就让权重W多几个为0（或者接近于0，说明该节点影响很小）不就好了，相当于在神经网络中删掉了一些节点，这样模型就变简单了。



正则化

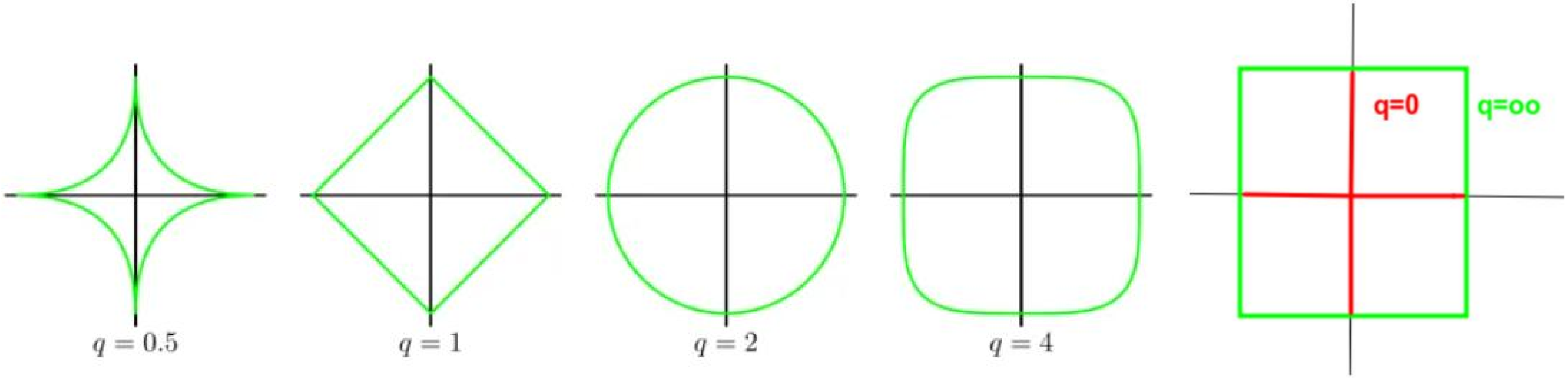
为了让W多几个为0，对于我们的正则化项 Ω（W),定义如下3种范数：

- **L0范数**：$\|w\|_0$，指向量中非0的元素的个数，越小说明0元素越多
- **L1范数**：$\|w\|_1$，指向量中各个元素绝对值之和
- **L2范数**：$\|w\|_2$，即各元素的平方和再开方

$$||x||_1 = \sum_{i=1}^{N} |x_i| , \qquad ||\mathbf{x}||_2 = \sqrt{\sum_{i=1}^{N} x_i^2} ,$$

正则化

$$\frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^2 + \frac{\lambda}{2}\sum_{j=1}^{M}|w_j|^q$$



上图中，可以明显看到一个趋势，即q越小，曲线越贴近坐标轴，q越大，曲线越远离坐标轴，并且棱角越明显。那么 q=0 和 q=oo 时极限情况如何呢？猜猜看。

正则化

- 线性回归+L1正则项：Lasso 回归
- 线性回归+L2正则项：Ridge 回归（岭回归）
- 如果我们用L0范数来正则化一个参数矩阵W的话，就是希望W的大部分元素都是0，让参数W是稀疏的，“压缩感知”、“稀疏编码”就是通过L0来实现的
- L1范数是L0范数的最优凸近似，而且它比L0范数要容易优化求解

$$\begin{array}{c} \min \|x\|_0 \\ \text{s.t. } Ax = b \end{array} \xleftrightarrow{\text{在一定条件下，以概率1意义下等价}} \begin{array}{c} \min \|x\|_1 \\ \text{s.t. } Ax = b \end{array}$$

正则化

- **特征选择**： x_i的大部分元素（也就是特征）都是和最终的输出y_i没有关系或者不提供任何信息的；但在预测新的样本时，这些没用的信息反而会被考虑，从而干扰了对正确y_i的预测。稀疏规则化算子的引入就是为了完成特征自动选择的光荣使命，它会学习地去掉这些没有信息的特征，也就是把这些特征对应的权重置为0。

- **可解释性**： 患病回归模型$y=w_1*x_1+w_2*x_2+\ldots+w_{1000}*x_{1000}+b$，通过学习，如果最后学习到的w*就只有很少的非零元素，例如只有5个非零的wi。也就是说，患不患这种病只和这5个因素有关，那医生就好分析多了。

第三章：线性模型

1. 线性回归

- 线性回归、多元线性回归

2. 二分类任务

- 广义线性回归

单位跃阶函数、对数线性回归、对数几率回归

- 线性判别分析

3. 多分类任务

4. 类别不平衡问题

广义线性模型 解 二分类

广义线性模型

- 一般形式 $y = g^{-1}(\omega^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b})$ $y = \{0,1\}$ $f(\boldsymbol{x}) \cong y$
  - $g(\cdot)$：联系函数，单调可微函数

    将分类标记 y 与线性回归输出 $z = f(\boldsymbol{x}) = \omega^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b}$ 联系起来

广义线性模型 解 二分类

广义线性模型

- 一般形式 $y = g^{-1}(\omega^{\mathrm{T}}x+b)$ $y = \{0,1\}$ $f(x) \cong y$

  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z = f(x) = \omega^{\mathrm{T}}x+b$联系起来

- $g(\cdot)$ 最理想的函数为单位跃阶函数，$y = \begin{cases} 0 & z < 0 \\ 0.5 & z = 0 \\ 1 & z > 0 \end{cases}$

广义线性模型 解 二分类

广义线性模型

- 一般形式 $y = g^{-1}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b)$ $\quad y = \{0,1\}$ $\quad f(\boldsymbol{x}) \cong y$

  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z = f(\boldsymbol{x}) = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b$ 联系起来

- $g(\cdot)$ 最理想的函数为单位跃阶函数，$y = \begin{cases} 0 & z < 0 \\ 0.5 & z = 0 \\ 1 & z > 0 \end{cases}$

- $g(\cdot) = \ln(y)$，对数函数，为对数线性回归

广义线性模型 解 二分类

广义线性模型

- 一般形式 $y = g^{-1}(\omega^{\mathrm{T}}x+b)$ $y = \{0,1\}$ $f(x) \cong y$

  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z=f(x) = \omega^{\mathrm{T}}x+b$ 联系起来

- $g(\cdot)$ 最理想的函数为单位跃阶函数，$y = \begin{cases} 0 & z < 0 \\ 0.5 & z = 0 \\ 1 & z > 0 \end{cases}$

- $g(\cdot) = \ln(y)$，对数函数，为对数线性回归

- $g(\cdot) = \ln \frac{y}{1-y}$，对数几率函数，为对数几率回归

广义线性模型 解 二分类

广义线性模型

- 一般形式 $y = g^{-1}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b})$ $y=\{0,1\}$ $f(\boldsymbol{x}) \cong y$

$g(\cdot)$ 将分类标记 y 与线性回归输出 $z=f(\boldsymbol{x}) = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b}$ 联系起来

- $g(\cdot)$ 最理想的函数为单位跃阶函数

$$y=\begin{cases}0 & z<0 \text{ 即 } \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b<\mathbf{0}\\ 0.5 & z=0 \text{ 即 } \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b=\mathbf{0}\\ 1 & z>0 \text{ 即 } \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b>\mathbf{0}\end{cases}$$

- 预测值 $z>0$, 即 $\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b>0$，判为 正例 $y=\{1\}$
  预测值 $z<0$, 即 $\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b<0$，判为 反例 $y=\{0\}$
  预测值 $z=0$, 即 $\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b=0$ 临界值，则可任意判别

广义 线性模型 解 二分类

广义线性模型

- 一般形式 $y = g^{-1}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b})$ $\quad y = \{0,1\}$ $\quad f(\boldsymbol{x}) \cong y$

  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z = f(\boldsymbol{x}) = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b}$ 联系起来

- $g(\cdot)$ 最理想的函数为**单位跃阶函数**

$$y = \begin{cases} 0 & z < 0 \text{ 即 } \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b < \mathbf{0} \\ 0.5 & z = 0 \text{ 即 } \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b = \mathbf{0} \\ 1 & z > 0 \text{ 即 } \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b > \mathbf{0} \end{cases}$$

- 预测值 $z>0$, 即 $\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b>0$，判为 **正例** $y=\{1\}$

  预测值 $z<0$, 即 $\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b<0$，判为 **反例** $y=\{0\}$

  预测值 $z=0$, 即 $\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b=0$ **临界值**，则可**任意判别**

- **缺点**：不连续，不能直接用作 $g^{-1}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b)$ 中的 $g^{-1}()$

广义 线性模型 解 二分类

广义 线性模型

- 一般形式 $y = g^{-1}(\omega^{\mathrm{T}}x+b)$ $y = \{0,1\}$ $f(x) \cong y$

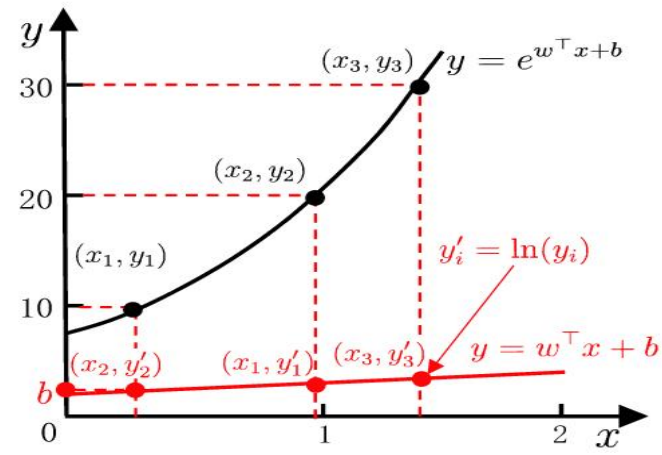
  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z = f(x) = \omega^{\mathrm{T}}x+b$ 联系起来

- $g(\cdot) = \ln(y)$，对数线性回归

$$y = e^{\omega^{\mathrm{T}}x+b}$$

$$\Updownarrow$$

$$\ln y = \omega^{\mathrm{T}}x+b$$



广义 线性模型 解 二分类

广义 线性模型

- 一般形式 $y = g^{-1}(\omega^{T}x+b)$ $y = \{0,1\}$ $f(x) \cong y$

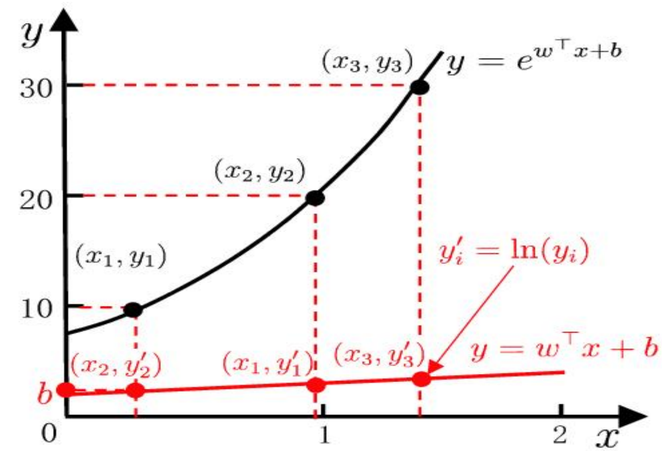
$g(\cdot)$ 将分类标记 y 与线性回归输出 $z = f(x) = \omega^{T}x+b$ 联系起来

- $g(\cdot) = \ln(y)$，对数线性回归

$$y = e^{\omega^{T}x+b}$$

$$\Updownarrow$$

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



- 用线性回归模型的预测结果

$f(x)=\omega^{T}x+b$ 逼近 真实标记 y 的 对数 $\ln(y)$， $y = e^{\omega^{T}x+b}$

- 实质上，是在求取输入空间到输出空间的 非线性函数映射。

广义 线性模型 解 二分类

广义线性模型

- 一般形式 $y=g^{-1}(\omega^{\mathrm{T}}x+b)$ $y=\{0,1\}$ $f(x)\cong y$

  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z=f(x)=\omega^{\mathrm{T}}x+b$ 联系起来

- $g(\cdot)=\ln\frac{y}{1-y}$，对数几率回归

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

◆ 线性回归的预测结果 $f(x)=\omega^{\mathrm{T}}x+b$ 逼近 真实标记 y 的对数几率

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

广义 线性模型 解 二分类

广义线性模型

- 一般形式 $y = g^{-1}(\omega^{\mathrm{T}}x+b)$ $y=\{0,1\}$ $f(x) \cong y$

  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z=f(x)=\omega^{\mathrm{T}}x+b$ 联系起来

- $g(\cdot)=\ln\frac{y}{1-y}$ ，对数几率回归

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

- 线性回归的预测结果 $f(x)=\omega^{\mathrm{T}}x+b$ 逼近 真实标记 y 的对数几率

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

- 对应的模型称为“对数几率回归”（logistic regression)
  - 注意：虽然它的名字是“回归”，实际是一种分类学习方法

广义线性模型解二分类

广义线性模型

- 一般形式　$y=g^{-1}(\omega^{\mathrm{T}}x+b)$　　$y=\{0,1\}$　　$f(x)\cong y$

  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z=f(x)=\omega^{\mathrm{T}}x+b$ 联系起来

- $g(\cdot)=\ln\frac{y}{1-y}$，对数几率回归

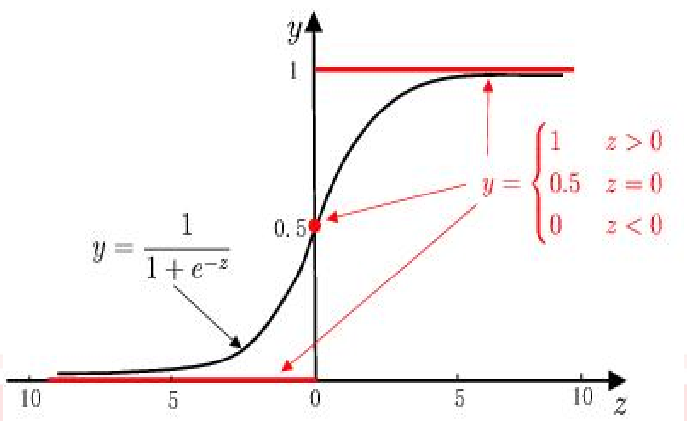
◆ 线性回归的预测结果 $f(x)=\omega^{\mathrm{T}}x+b$ 逼近 真实标记 y 的对数几率

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

◆ 特点

- 单调可微、任意阶可导
- 将 $z=f(x)$ 值转化为一个接近0或1的 y 值
- 输出值在 $z=f(x)=0$ 附近变化很陡

单位阶跃函数 VS 对数几率函数



广义 线性模型 解 二分类

广义线性模型

- **一般形式** $y = g^{-1}(\omega^{\mathrm{T}} x + b)$ $y=\{0,1\}$ $f(x) \cong y$

  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z = f(x) = \omega^{\mathrm{T}} x + b$ 联系起来

- $g(\cdot) = \ln\frac{y}{1-y}$，**对数几率回归**

◆ **物理含义**

- y ：样本 x 作为**正例**的可能性
- $1\text{-}y$：样本 x 是**反例**的可能性
- 对数几率 $\ln\frac{y}{1-y}$ 描述了样本 x 作为 **正例** 的相对**可能性**

广义 线性模型 解 二分类

广义线性模型

- 一般形式 $y = g^{-1}(\omega^T x + b)$ $y = \{0,1\}$ $f(x) \cong y$

  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z = f(x) = \omega^T x + b$ 联系起来

- $g(\cdot) = \ln\frac{y}{1-y}$，对数几率回归

◆ 对数几率 $\ln\frac{y}{1-y}$ 描述了 样本 作为 正例 的相对可能性

$$\ln\frac{y}{1-y} = \omega^T x + b \Rightarrow \ln\frac{P(y=1|x)}{P(y=0|x)} = \omega^T x + b$$

$$\Rightarrow \begin{cases} P(y=1|x) = \dfrac{e^{\omega^T x + b}}{1 + e^{\omega^T x + b}} \\ P(y=0|x) = \dfrac{1}{1 + e^{\omega^T x + b}} \end{cases}$$

广义 线性模型 解 二分类

广义线性模型

□ 一般形式 $y = g^{-1}(\omega^{\mathrm{T}}x+b)$ $y = \{0,1\}$ $f(x) \cong y$

$g(\cdot)$ 将分类标记 y 与线性回归输出 $z = f(x) = \omega^{\mathrm{T}}x+b$ 联系起来

□ $g(\cdot) = \ln\frac{y}{1-y}$ ，对数几率回归

样本为正例的相对可能性

即 $\ln\frac{\mathrm{P}(y=1|x)}{\mathrm{P}(y=0|x)} = \omega^{\mathrm{T}}x + b$ ⇒ $\mathrm{P}(y=1|x) = \frac{e^{\omega^{\mathrm{T}}x+b}}{1+e^{\omega^{\mathrm{T}}x+b}}$

◆ 求解: 极大似然法 $f(x) \cong y \rightarrow$ 估计 ω 和 b，数据集 $\{(x_i, y_i)\}_{i=1,\ldots,m}$

广义 线性模型 解 二分类

广义线性模型

- 一般形式 $y = g^{-1}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b})$ $\quad y = \{0,1\}$ $\quad f(x) \cong y$

  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z = f(x) = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b}$ 联系起来

- $g(\cdot) = \ln\frac{y}{1-y}$， 对数几率回归

样本 为 正例 的相对可能性

即 $\ln\frac{\mathrm{P}(y=1|\boldsymbol{x})}{\mathrm{P}(y=0|\boldsymbol{x})} = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x} + b \Rightarrow \mathrm{P}(y=1|\boldsymbol{x}) = \frac{e^{\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b}}}{1+e^{\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b}}}$

◆ 求解: 极大似然法 $f(x) \cong y \rightarrow$ 估计 ω 和 b， 数据集 $\{(x_i, y_i)\}_{i=1,\dots,m}$

目标函数： 最大化 样本标记属于其真实标记 的概率

即 max P{ 样本标记 ⊂ 真实标记 } 或 min P{ 样本 错误标记 }

- 最大化 对数似然函数 $\ell(\boldsymbol{\omega}, b) = \sum_{i=1}^{m} \ln p(y_i|x_i; \omega_i, b)$

即 每个样本标记属于其真实标记的概率越大越好

广义 线性模型 解 二分类

广义线性模型

- 一般形式 $y = g^{-1}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b})$　　$y = \{0,1\}$　　$f(x) \cong y$

$g(\cdot)$ 将分类标记 y 与线性回归输出 $z = f(x) = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b}$ 联系起来

- $g(\cdot) = \ln\frac{y}{1-y}$，对数几率回归

样本 为 正例 的相对可能性

即 $\ln\frac{\mathrm{P}(y=1|\boldsymbol{x})}{\mathrm{P}(y=0|\boldsymbol{x})} = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x} + b \Rightarrow \mathrm{P}(y=1|\boldsymbol{x}) = \frac{e^{\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b}}{1+e^{\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}+b}}$

◆ 求解: 极大似然法 $f(x) \cong y \rightarrow$ 估计 ω 和b，数据集$\{(x_i, y_i)\}_{i=1,\ldots,m}$

目标函数：最大化 样本标记属于其真实标记 的概率

- 最大化 对数似然函数 $\ell(\boldsymbol{\omega}, b) = \sum_{i=1}^{m} \ln p(y_i|x_i; \omega_i, b)$

求解：凸优化理论中的数值优化算法，如梯度下降法、牛顿法

广义 线性模型 解 二分类

广义线性模型

- 一般形式 $y = g^{-1}(\omega^{\mathrm{T}}x+b)$ $y=\{0,1\}$ $f(x) \cong y$

  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z=f(x)=\omega^{\mathrm{T}}x+b$ 联系起来

- $g(\cdot) = \ln\frac{y}{1-y}$，对数几率回归

◆ 线性回归$f(x)=\omega^{\mathrm{T}}x+b$ 逼近 真实标记 y 的对数几率

◆ 对数几率描述了：样本作为 正例 的相对可能性

广义 线性模型 解 二分类

广义线性模型

- 一般形式　$y = g^{-1}(\omega^{T}x+b)$　　$y=\{0,1\}$　　$f(x) \cong y$

  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z=f(x)=\omega^{T}x+b$ 联系起来

- $g(\cdot) = \ln\frac{y}{1-y}$，对数几率回归

◆ 线性回归 $f(x)=\omega^{T}x+b$ 逼近 真实标记 y 的对数几率

◆ 对数几率描述了：样本作为 正例 的相对可能性

◆ 优点

- 直接对分类可能性进行建模，无需事先假设数据分布，避免了假设分布不准确所带来的问题

广义 线性模型 解 二分类

广义线性模型

- 一般形式 $y = g^{-1}(\omega^T x + b)$ $y = \{0,1\}$ $f(x) \cong y$

  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z = f(x) = \omega^T x + b$ 联系起来

- $g(\cdot) = \ln \frac{y}{1-y}$，对数几率回归

- 线性回归 $f(x)=\omega^T x+b$ 逼近 真实标记 y 的对数几率
- 对数几率 描述了：样本作为 正例 的相对可能性
- 优点
  - 不仅预测出“类别”，且可得到“类别”的近似概率预测，对许多需利用概率辅助决策的任务很有用。

广义 线性模型 解 二分类

广义线性模型

- 一般形式　$y = g^{-1}(\omega^{\mathrm{T}}x + b)$　　$y = \{0,1\}$　　$f(x) \cong y$

  $g(\cdot)$ 将分类标记 y 与线性回归输出 $z = f(x) = \omega^{\mathrm{T}}x + b$ 联系起来

- $g(\cdot) = \ln \frac{y}{1-y}$，对数几率回归

◆ 线性回归$f(x)=\omega^{\mathrm{T}}x+b$ 逼近 真实标记 y 的对数几率

◆ 对数几率 描述了：样本作为 正例 的相对可能性

◆ 优点

- 对数几率函数，是任意阶可导的凸函数，可直接应用现有数值优化算法求取最优解

第三章：线性模型

1. 线性回归

- 线性回归、多元线性回归

2. 二分类任务

- 广义线性回归

单位跃阶函数、对数线性回归、对数几率回归

- **线性判别分析**

3. 多分类任务

4. 类别不平衡问题

线性判别分析 LDA

线性判别分析

- Linear Discriminant Analysis，LDA
- 经典的线性学习方法，最早由[Fisher, 1936] 提出，称Fisher判别分析

线性判别分析 LDA

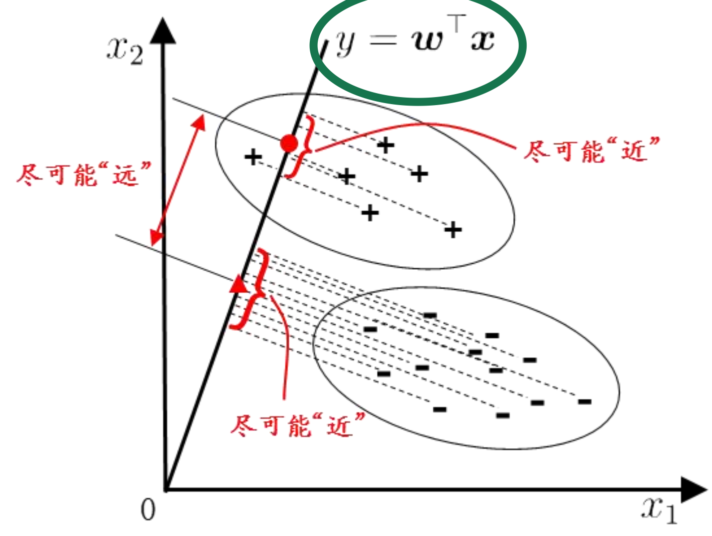
线性判别分析

- 给定训练样本集，设法将样本投影到一条直线上，使得
  - 同类样本的投影点，尽可能接近
  - 异类样本的投影点，尽可能远离

线性判别分析 LDA

线性判别分析

- 给定训练样本集，设法将样本投影到一条直线上，使得
  - 同类样本的投影点，尽可能接近
  - 异类样本的投影点，尽可能远离
- 对新样本进行分类时，
  - 将新样本投影到同样的这条直线上
  - 根据投影点的位置确定新样本的类别



- LDA可视为监督降维技术，求解二分类问题

线性判别分析 LDA

LDA 思想

- 同类样本的投影点，尽可能接近
  - 投影点的协方差尽可能小
  - 协方差矩阵 $\sum_i$ 尽可能小
- 异类样本的投影点，尽可能远离
  - 类中心之间的距离，尽可能大
  - 异类样本的均值向量 μ_i之间的距离尽可能大

线性判别分析 LDA

LDA 思想

- 同类样本接近，协方差矩阵 Σ_i 尽可能小
- 异类样本远离，均值向量 μ_i 尽可能大

线性判别分析 LDA

LDA 思想

- ◆同类样本接近，协方差矩阵 Σ_i 尽可能小
- ◆异类样本远离，均值向量 μ_i 尽可能大

一些变量

- 第i类样本的集合 X_i
- 第i类样本的均值向量 μ_i
- 第i类样本的协方差矩阵 Σ_i
- 两类样本中心在直线上的投影：$\omega^{\mathrm{T}}\mu_0$　$\omega^{\mathrm{T}}\mu_1$
- 两类样本在直线上的投影的协方差：$\omega^{\mathrm{T}}\Sigma_0\omega$ 和 $\omega^{\mathrm{T}}\Sigma_1\omega$

线性判别分析 LDA

同类样本接近，协方差矩阵 Σ_i 尽可能小

- 最小化类内散度矩阵

$$\mathbf{S}_\omega = \Sigma_0 + \Sigma_1 = \sum_{x\in X_0}(x-\mu_0)(x-\mu_0)^{\mathrm{T}} + \sum_{x\in X_1}(x-\mu_1)(x-\mu_1)^{\mathrm{T}}$$

线性判别分析 LDA

同类样本接近，协方差矩阵 Σ_i 尽可能小

- 最小化类内散度矩阵

$$\mathbf{S}_\omega = \boldsymbol{\Sigma}_\mathbf{0} + \boldsymbol{\Sigma}_\mathbf{1} = \sum_{x \in X_0} (x - \mu_0)(x - \mu_0)^\mathrm{T} + \sum_{x \in X_1} (x - \mu_1)(x - \mu_1)^\mathrm{T}$$

异类样本远离，均值向量 μ_i 尽可能大

- 最大化类间散度矩阵

$$\mathbf{S}_b = (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^\mathrm{T}$$

线性判别分析 LDA

同类样本接近，协方差矩阵 Σ_i 尽可能小

- 最小化类内散度矩阵

$$\mathbf{S}_\omega = \Sigma_0 + \Sigma_1 = \sum_{x \in X_0} (x-\mu_0)(x-\mu_0)^{\mathrm{T}} + \sum_{x \in X_1} (x-\mu_1)(x-\mu_1)^{\mathrm{T}}$$

异类样本远离，均值向量 μ_i 尽可能大

- 最大化类间散度矩阵

$$\mathbf{S}_b = (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^{\mathrm{T}}$$

最大化 广义瑞利商

$$J = \frac{\omega^{\mathrm{T}} \mathbf{S}_b \omega}{\omega^{\mathrm{T}} \mathbf{S}_\omega \omega} = \frac{\omega^{\mathrm{T}} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^{\mathrm{T}} \omega}{\omega^{\mathrm{T}} (\Sigma_0 + \Sigma_1) \omega}$$

线性判别分析 LDA

LDA 求解多分类任务

同类样本接近，协方差矩阵 $\sum_i$ 尽可能小

异类样本远离，均值向量 μ_i 尽可能大

线性判别分析 LDA

LDA 求解多分类任务

同类样本接近，协方差矩阵 $\sum_i$ 尽可能小

- 最小化每个类别内散度矩阵

$$\mathbf{S}_{\omega}=\sum_{i=1}^{N}\mathbf{S}_{\omega_i} \qquad \mathbf{S}_{\omega_i}=\sum_{x\in X_i}(x-\mu_i)(x-\mu_i)^{\mathrm{T}}$$

线性判别分析 LDA

LDA 求解多分类任务

同类样本接近，协方差矩阵 Σ_i 尽可能小

- 最小化每个类别内散度矩阵

$$\mathbf{S}_\omega = \sum_{i=1}^{N} \mathbf{S}_{\omega_i} \qquad \mathbf{S}_{\omega_i} = \sum_{x \in X_i} (x - \mu_i)(x - \mu_i)^{\mathrm{T}}$$

异类样本远离，均值向量 μ_i 尽可能大

- 最大化不同类别间散度矩阵

$$\mathbf{S}_b = \mathbf{S}_t - \mathbf{S}_\omega = \sum_{i=1}^{N} m_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^{\mathrm{T}}$$

线性判别分析 LDA

LDA 求解多分类任务

同类样本接近，协方差矩阵 Σ_i 尽可能小

- 最小化 每个类别内 散度矩阵

$$\mathbf{S}_\omega = \sum_{i=1}^{N} \mathbf{S}_{\omega_i} \qquad \mathbf{S}_{\omega_i} = \sum_{x \in X_i} (x - \mu_i)(x - \mu_i)^{\mathrm{T}}$$

异类样本远离，均值向量 μ_i 尽可能大

- 最大化 不同类别间 散度矩阵

$$\mathbf{S}_b = \mathbf{S}_t - \mathbf{S}_\omega = \sum_{i=1}^{N} m_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^{\mathrm{T}}$$

最大化全局 散度矩阵

$$\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_\omega = \sum_{i=1}^{m} (x_i - \boldsymbol{\mu})(x_i - \boldsymbol{\mu})^{\mathrm{T}}$$

线性判别分析 LDA

LDA 求解多分类任务

- 最大化全局 散度矩阵 $\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_\omega = \sum_{i=1}^{m} (x_i - \boldsymbol{\mu})(x_i - \boldsymbol{\mu})^{\mathrm{T}}$

- 转化为：广义特征值 求解

$$\max_{\mathrm{W}} \frac{\operatorname{tr}(\mathrm{W}^{\mathbf{T}}\mathbf{S}_b\mathrm{W})}{\operatorname{tr}(\mathrm{W}^{\mathbf{T}}\mathbf{S}_\omega\mathrm{W})} \Rightarrow \mathbf{S}_b\mathrm{W} = \lambda\mathbf{S}_\omega\mathrm{W}$$

- ω的闭式解
  - $S_\omega^{-1}S_b$的前d’个最大广义特征值 所对应的 特征向量组成的矩阵

线性判别分析 LDA

LDA 求解多分类任务

- ω的闭式解：
  - $S_{\omega}^{-1}S_b$的d’个最大广义特征值所对应的特征向量矩阵

- 若将ω 视为投影矩阵，则多分类LDA 将样本投影到 N-1 维空间

- 可通过该投影 减小样本点的维数，且投影过程中 使用了类别信息。因此，LDA也常被视为一种经典的监督降维技术

线性模型优化总结

回归/分类**模型优化**的**目标**和**参数**的**优化方法**

- **线性回归**：最小二乘法，即最小化均方误差
  求解：解析数学
- **对数几率回归**：最大化 样本分布似然
  求解：凸优化梯度下降、牛顿法
- **线性判别分析**：投影空间内，最小(大)化 类内(间)散度；
  求解：矩阵论、广义瑞利商

第三章：线性模型

1. 线性回归

- 线性回归、多元线性回归

2. 二分类任务

- 广义线性回归、线性判别分析

3. 多分类任务

- 一对一、一对其余、多对多

4. 类别不平衡问题

多分类学习

多分类学习方法

- 二分类学习方法 推广到 多分类
- 利用二分类学习器，解决多分类问题（常用）
  - 将多分类问题进行拆分，拆出的每个二分类任务训练对应一个二分类分类器
  - 对于每个二分类分类器的预测结果进行集成，以获得最终的多分类结果

多分类学习

多分类学习方法

- 利用二分类学习器，解决多分类问题（常用）
  - 将多分类问题拆分为若干个二分类任务
  - 集成二分类器的结果，并获得最终的多分类结果

拆分策略

- 一对一：$N(N-1)/2$ 个二类分类器
- 一对其余：N 个二类分类器
- 多对多

多分类学习

一对一　　　　一对其余

拆分阶段

决策阶段

多分类学习

一对一　　　　一对其余

拆分阶段

- N个类别两两配对
  - N(N-1)/2 二分类任务

决策阶段

- 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，作为最终类别

多分类学习

一对一

拆分阶段

- N个类别两两配对
  - N(N-1)/2 二分类任务

决策阶段

- 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，作为最终类别

一对其余

拆分阶段

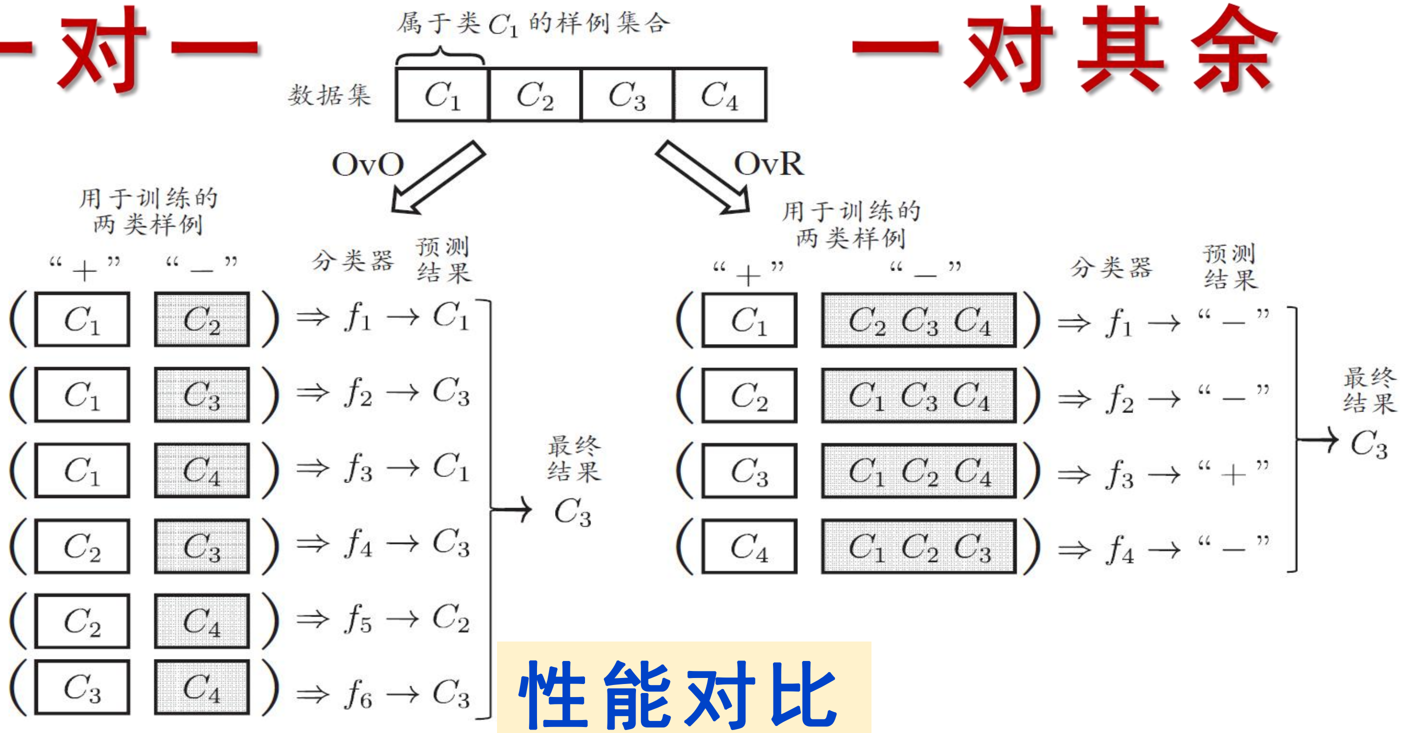
- 某一类作为正例，其他类作为反例
  - N 二分类任务

决策阶段

- 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

一对一　　一对其余

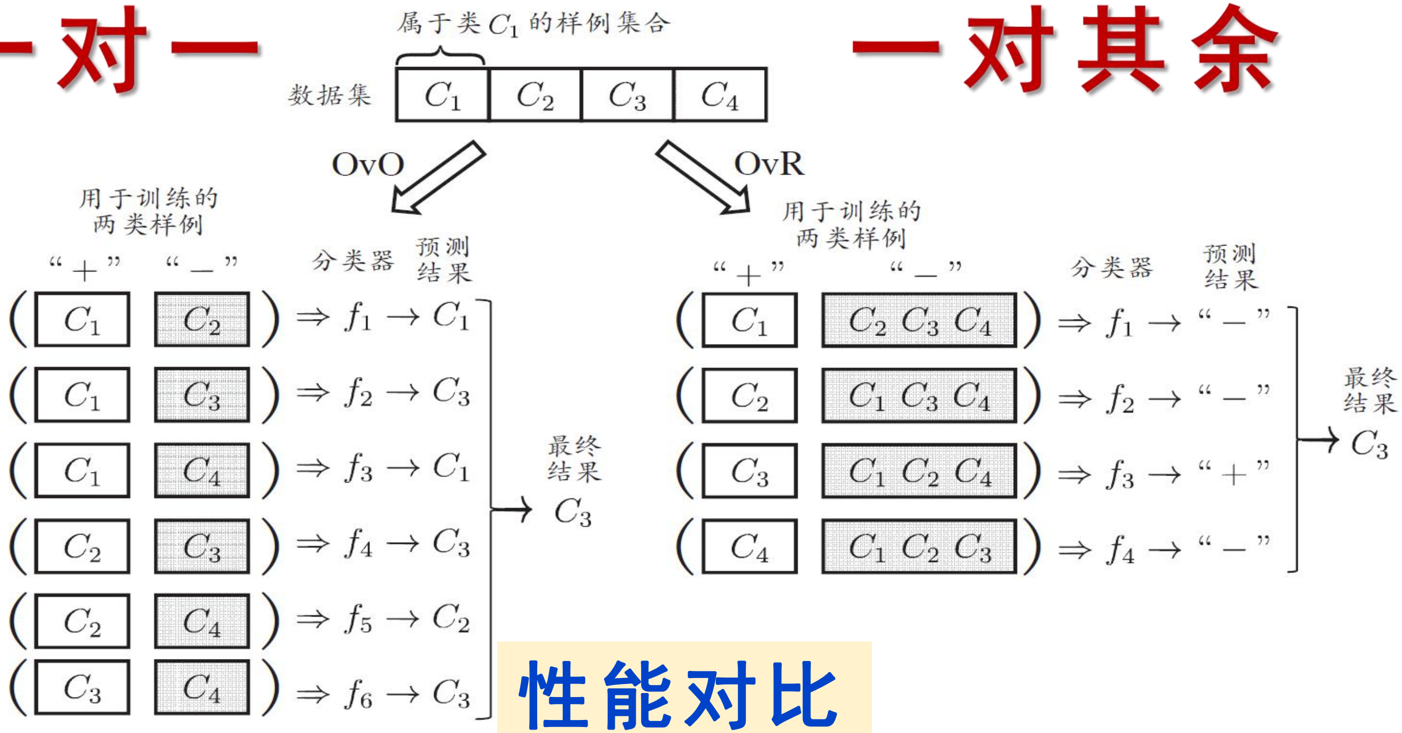


性能对比

- N(N-1)/2 个分类器
  存储开销大和测试时间长
- 每个分类器训练，只用两个类样本，训练时间短

多分类学习

一对一　　　一对其余



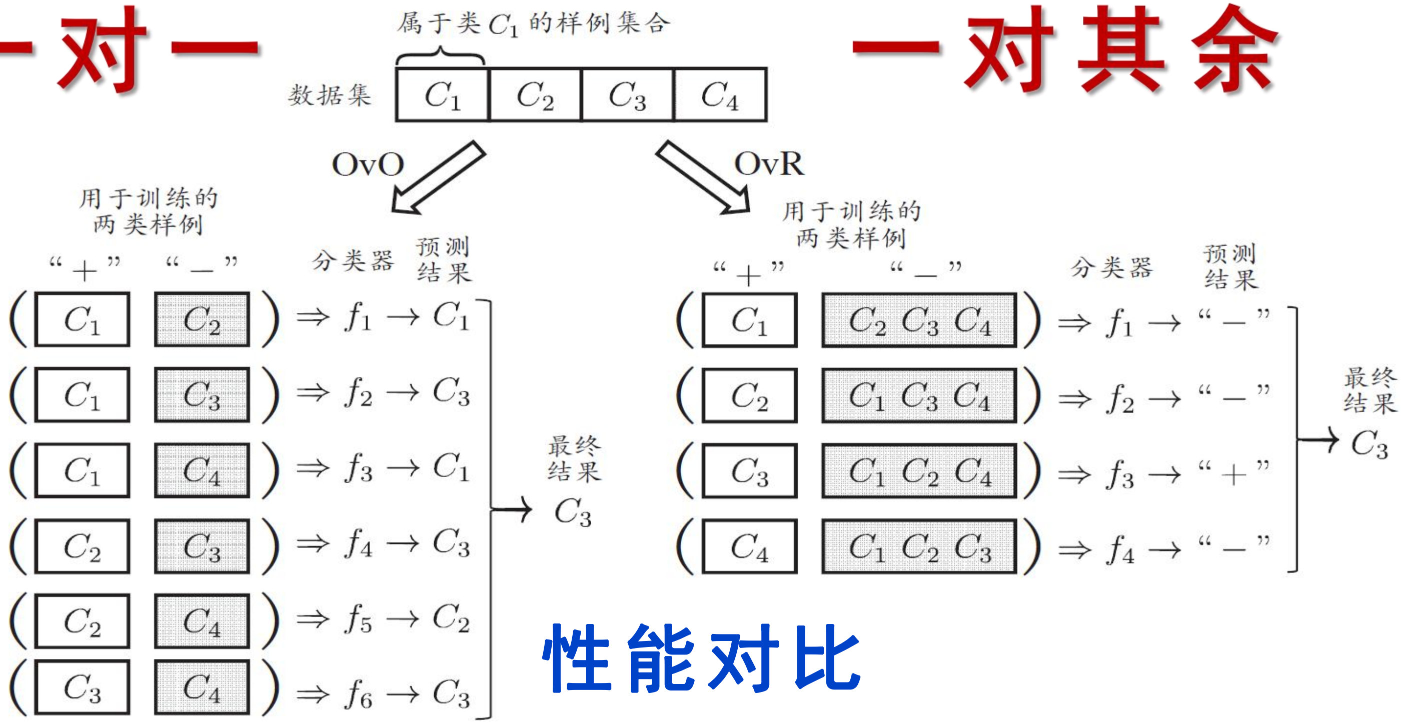
性能对比

- N(N-1)/2 个分类器
  存储开销**大**和测试时间**长**
- 每个分类器训练，只用**两个类样本**，训练时间**短**

- N 个分类器
  存储开销**小**和测试时间**短**
- 每个分类器训练，用到**全部训练样本**，训练时间**长**

多分类学习

一对一　　　　　　　　　　一对其余



性能对比

- **预测性能**:取决于具体数据分布，多数情况下两者差不多
- 不同类别的样本，**差异性较小**时，**一对其余**，效果更好

第三章：线性模型

1. 线性回归

- 线性回归、多元线性回归

2. 二分类任务

- 广义线性回归、线性判别分析

3. 多分类任务

4. 类别不平衡问题

类别不平衡问题

- 前面的分类学习方法，都有一个共同的基本假设 不同类别 的训练样本 数目相当
- 如果不同类别的训练样本数目 稍有差别，通常影响不大
- 但是，若不同类别的 训练样本数目差别很大，则会对学习过程造成困扰

类别不平衡问题

- 不同类别的训练样本数目差别很大，则会对学习过程造成困扰
- 现实分类学习任务中，经常会遇到类别不平衡

类别不平衡问题

- 不同类别的训练样本数目差别很大，则会对学习过程造成困扰
- 现实分类学习任务中，经常会遇到类别不平衡
  - 例如：公安系统中，犯罪分子的甄别问题：有998个反例（合法公民），但正例只有2个（犯罪分子）
  - 那么，学习方法，只需返回一个永远将新样本预测为反例（合法公民）的学习器，就能达到99.8%的精度
  - 然而，这样的学习器往往没有价值，因为它不能预测出任何正例

类别不平衡问题

- 不同类别的训练样本数目差别很大，则会对学习过程造成困扰
- 现实分类学习任务中，经常会遇到类别不平衡
  - 通过拆分法，解决多分类问题时，

    即使 原始问题中，不同类别的训练样本数目相当，

    使用一对多、多对多策略后，产生的二分类任务仍可能出现类别不平衡现象

类别不平衡问题

解决方式:再缩放

- **欠采样**：去除一些反例
- **过采样**：增加一些正例
- **阈值移动**

类别不平衡问题

解决方式：再缩放

◆ 阈值移动

- 对于线性分类器，用 $y=\omega^T x+b$ 对 新样本 进行分类
  - y：表达了正例的可能性。通常 $y>0.5$，判别为正例，否则为反例。
  - 几率 $\frac{y}{1-y}$：反映了正例可能性与反例可能性之比值

类别不平衡问题

解决方式:再缩放

◆ 阈值移动

- 对于线性分类器，用 $y = \omega^T x + b$ 对 新样本 进行分类
  - 分类器的预测 几率 > 观测 几率，判为正例，

    即 $\dfrac{y}{1-y} > \dfrac{m^+}{m^-}$ 预测为正例

    其中，m^+为正例样本个数，m^-为反例样本个数

正例预测

类别平衡 $\dfrac{y}{1-y} > 1$

类别不平衡 $\dfrac{y}{1-y} > \dfrac{m^+}{m^-}$

第三章：线性模型

1. 线性回归

- 线性回归、多元线性回归

2. 二分类任务

- 广义线性回归、线性判别分析

3. 多分类任务

4. 类别不平衡问题

第三章：线性模型总结

1. **线性回归** **目的**：学得一个线性模型以尽可能准确地预测实值输出标记
2. **最小二乘法**：试图 找到一条直线，使所有样本到直线上的欧氏距离之和最小
3. **线性判别分析 LDA**：是一种经典的**线性学习**方法。
   - ✓给定训练样本集，设法将样本投影到**一条直线**上，使得 **同类样本**的投影点尽可能**接近**，**异类样本**的投影点尽可能**远离**。
   - ✓在对新样本进行分类时，将其投影到同样的这条直线上，再根据投影点的位置来确定新样本的类别.
4. **类别不平衡问题**：不同类别训练样本数相差很大情况（正类为小类）