人机交互技术：
BCI中的机器学习--线性模型

- 线性回归
  - 最小二乘法
- 二分类任务
  - 对数几率回归
  - 线性判别分析
- 多分类任务
  - 一对一
  - 一对其余
- 类别不平衡问题

基本形式

- 线性模型一般形式：

$$f(\mathbf{x}) = w_1 x_1 + w_2 x_2 + \ldots + w_d x_d + b$$

$\mathbf{x} = (x_1; x_2; \ldots; x_d)$ 是由属性描述的示例，其中 x_i 是 $\boldsymbol{x}$ 在第 i 个属性上的取值.

- 向量形式:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$$

其中

$$\mathbf{w} = (w_1; w_2; \ldots; w_d)$$

线性模型优点

- 形式简单、易于建模
- 可解释性
- 非线性模型的基础
  - 引入层级结构或高维映射
- 一个例子:
  - 综合考虑色泽、根蒂和敲声来判断西瓜好不好

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

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

线性回归

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

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

- 线性回归（linear regression）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
- 离散属性处理
  - ❖ 有“序”关系
    - ➢ 连续化为连续值
  - ❖ 无“序”关系
    - ➢ 有k个属性值，则转换为k维向量

线性回归

- 单一属性的线性回归目标:

  $f(x) = wx_i + b$ 使得 $f(x_i) \simeq y_i$

- 参数/模型估计：最小二乘法（least square method）

$$
\begin{aligned}
(w^*, b^*) &= \underset{(w,b)}{\arg\min} \sum_{i=1}^{m} \left(f(x_i) - y_i\right)^2 \\
&= \underset{(w,b)}{\arg\min} \sum_{i=1}^{m} \left(y_i - wx_i - b\right)^2
\end{aligned}
$$

线性回归 - 最小二乘法

最小化均方误差: $E_{(w,b)} = \sum_{i=1}^{m} (y_i - wx_i - b)^2$

分别对 *w* 和 *b* 求导，可得:

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

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

得到闭式（closed-form）解:

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

$$b = \frac{1}{m}\sum_{i=1}^{m} (y_i - wx_i)$$

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

多元线性回归

- 给定数据集

$$D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, y_m)\}$$

$$\boldsymbol{x}_i = (x_{i1}; x_{i2}; \ldots; x_{id}) \quad y_i \in \mathbb{R}$$

- 多元线性回归目标:

$$f(\boldsymbol{x}_i) = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i + b \text{ 使得 } f(\boldsymbol{x}_i) \simeq y_i$$

多元线性回归

把 w 和 b 吸收入向量形式 $\hat{\boldsymbol{w}} = (\boldsymbol{w}; b)$，数据集表示为:

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

$$\boldsymbol{y} = (y_1; y_2; \ldots; y_m)$$

多元线性回归 - 最小二乘法

最小二乘法（least square method）：

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

令 $E_{\hat{w}} = (\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})^{\mathrm{T}}(\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{w}})$ ，对 $\hat{\boldsymbol{w}}$ 求导得到:

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

令上式为零可得 $\hat{\boldsymbol{w}}$ 最优解的闭式解

多元线性回归 - 满秩讨论

- $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ 是满秩矩阵，则

$$\hat{\boldsymbol{w}}^* = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{y}$$

其中 $\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$ 是 $\mathbf{X}^{\mathrm{T}}\mathbf{X}$的逆矩阵，线性回归模型为

$$f\left(\hat{\boldsymbol{x}}_i\right) = \hat{\boldsymbol{x}}_i^{\mathrm{T}}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{y}$$

- $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ 不是满秩矩阵
  - ❖ 根据归纳偏好选择解
  - ❖ 引入正则化

正则化：岭回归与LASSO

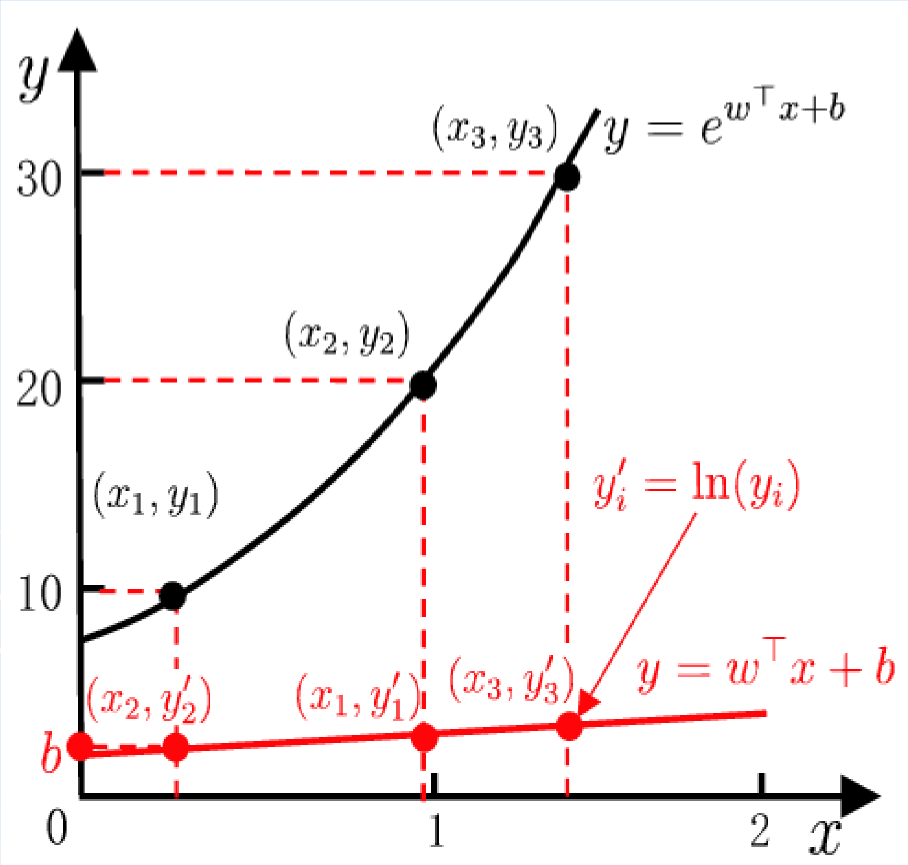
- 岭回归$\hat{\beta}(k) = (X'X + kI)^{-}X'y$可以看做对下式的解析求解

  $$\hat{\beta}(k) = \text{argmin}_{\beta}\{(y - X\beta)^T(y - X\beta) + k\|\beta\|^2\}$$

  - 其中$\|\beta\|^2$称作正则化项
    - 使得模型参数有向0收缩的趋势
    - 通过部分参数向0收缩来缓解过拟合的问题
  - 模型优化中常采用参数p阶范式$\|W\|^p$做正则约束来防止过拟合
    - $p = 1$时称作LASSO（The Least Absolute Shrinkage and Selectionator operator），倾向于稀疏参数或特征选择

对数线性回归

线性模型逼近的目标: 输出标记的对数



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

$$y = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x} + b$$

大纲

- 线性回归
  - 最小二乘法
- **二分类任务**
  - 对数几率回归
  - 线性判别分析
- 多分类任务
  - 一对一
  - 一对其余
- 类别不平衡问题

二分类任务

- 预测值与输出标记

$$z = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b \qquad y \in \{0, 1\}$$

- 寻找函数将分类标记与线性回归模型输出联系起来

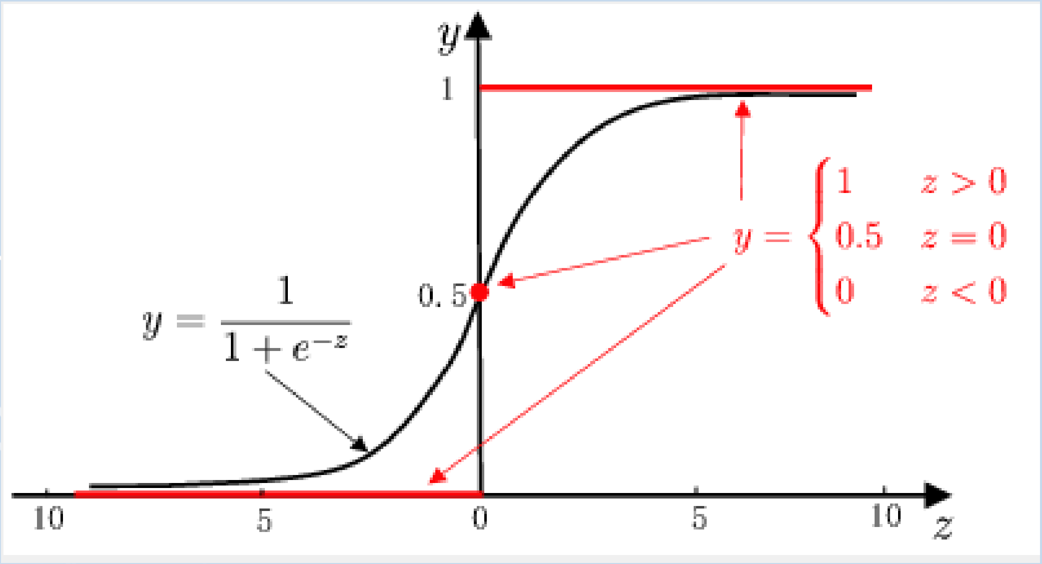
$$y = \begin{cases} 0, & z < 0; \\ 0.5, & z = 0; \\ 1, & z > 0, \end{cases}$$

- 最理想的函数: 单位阶跃函数

二分类任务

- 单位阶跃函数缺点: 不连续
- 替代函数——对数几率函数（logistic function）
  - 单调可微、任意阶可导 $$y = \frac{1}{1+e^{-z}}$$

单位阶跃函数与对数几率函数的比较



对数几率回归

- 运用对数几率函数

$$y = \frac{1}{1+e^{-z}} \quad \text{变为} \quad y = \frac{1}{1+e^{-(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b)}}$$

- 对数几率（log odds）
  - ❖ 样本作为正例的相对可能性的对数：$\ln \frac{y}{1-y}$

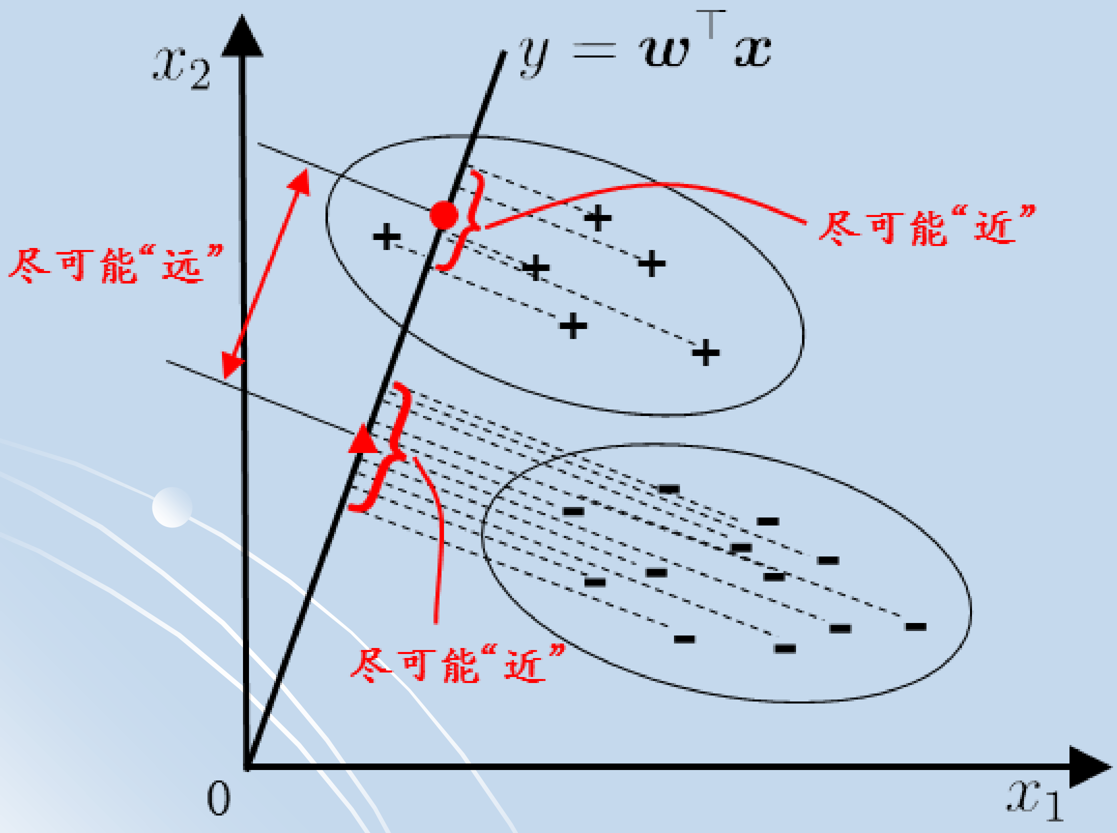
- 对数几率回归优点
  - ❖ 无需事先假设数据分布
  - ❖ 可得到“类别”的近似概率预测
  - ❖ 可直接应用现有数值优化算法求取最优解

大纲

- 线性回归
  - 最小二乘法
- 二分类任务
  - 对数几率回归
  - **线性判别分析**
- 多分类任务
  - 一对一
  - 一对其余
- 类别不平衡问题

二分类任务– 线性判别分析

线性判别分析（Linear Discriminant Analysis）



LDA也可被视为一种监督降维技术

二分类任务– 线性判别分析

- LDA的思想:
  - 欲使同类样例的投影点尽可能接近，可以让同类样例投影点的协方差尽可能小
  - 欲使异类样例的投影点尽可能远离，可以让类中心之间的距离尽可能大
- 一些变量:
  - 第 i 类示例的集合 X_i
  - 第 i 类示例的均值向量 $\boldsymbol{\mu}_i$
  - 第 i 类示例的协方差矩阵 $\mathbf{\Sigma}_i$
  - 两类样本的中心在直线上的投影：$\mathbf{w}^T\boldsymbol{\mu}_0$ 和 $\mathbf{w}^T\boldsymbol{\mu}_1$
  - 两类样本的协方差：$\mathbf{w}^T\mathbf{\Sigma}_0\mathbf{w}$ 和 $\mathbf{w}^T\mathbf{\Sigma}_1\mathbf{w}$

二分类任务– 线性判别分析

- 最大化目标

$$J=\frac{\left\|\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\mu}_0-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\mu}_1\right\|_2^2}{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Sigma}_0\boldsymbol{w}+\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Sigma}_1\boldsymbol{w}}
=\frac{\boldsymbol{w}^{\mathrm{T}}\left(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1\right)\left(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1\right)^{\mathrm{T}}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_0+\boldsymbol{\Sigma}_1\right)\boldsymbol{w}}$$

- 类内散度矩阵

$$\begin{aligned}\mathbf{S}_w&=\boldsymbol{\Sigma}_0+\boldsymbol{\Sigma}_1\\&=\sum_{\boldsymbol{x}\in X_0}\left(\boldsymbol{x}-\boldsymbol{\mu}_0\right)\left(\boldsymbol{x}-\boldsymbol{\mu}_0\right)^{\mathrm{T}}+\sum_{\boldsymbol{x}\in X_1}\left(\boldsymbol{x}-\boldsymbol{\mu}_1\right)\left(\boldsymbol{x}-\boldsymbol{\mu}_1\right)^{\mathrm{T}}\end{aligned}$$

- 类间散度矩阵 $\mathbf{S}_b=\left(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1\right)\left(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1\right)^{\mathrm{T}}$

二分类任务– 线性判别分析

- 广义瑞利商（generalized Rayleigh quotient）

$$J=\frac{\boldsymbol{w}^{\mathrm{T}}\mathbf{S}_b\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\mathbf{S}_w\boldsymbol{w}}$$

- 令 $\boldsymbol{w}^{\mathrm{T}}\mathbf{S}_w\boldsymbol{w}=1$，最大化广义瑞利商等价形式为

$$\min_{\boldsymbol{w}} \; -\boldsymbol{w}^{\mathrm{T}}\mathbf{S}_b\boldsymbol{w}$$
$$\text{s.t. } \boldsymbol{w}^{\mathrm{T}}\mathbf{S}_w\boldsymbol{w}=1$$

- 运用拉格朗日乘子法 $\mathbf{S}_b\boldsymbol{w}=\lambda\mathbf{S}_w\boldsymbol{w}$

二分类任务– 线性判别分析

- 同向向量 $\mathbf{S}_b\boldsymbol{w} = \lambda(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)$

  同向向量

- 结果 $\boldsymbol{w} = \mathbf{S}_w^{-1}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)$

- 求解
  - ❖ 奇异值分解 $\mathbf{S}_w = \mathbf{U}\Sigma\mathbf{V}^T$

- LDA的贝叶斯决策论解释
  - ❖ 两类数据同先验、满足高斯分布且协方差相等时，LDA达到最优分类

大纲

- 线性回归
  - 最小二乘法
- 二分类任务
  - 对数几率回归
  - 线性判别分析
- **多分类任务**
  - 一对一
  - 一对其余
- 类别不平衡问题

多分类学习

- 二分类学习方法推广到多类
- 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

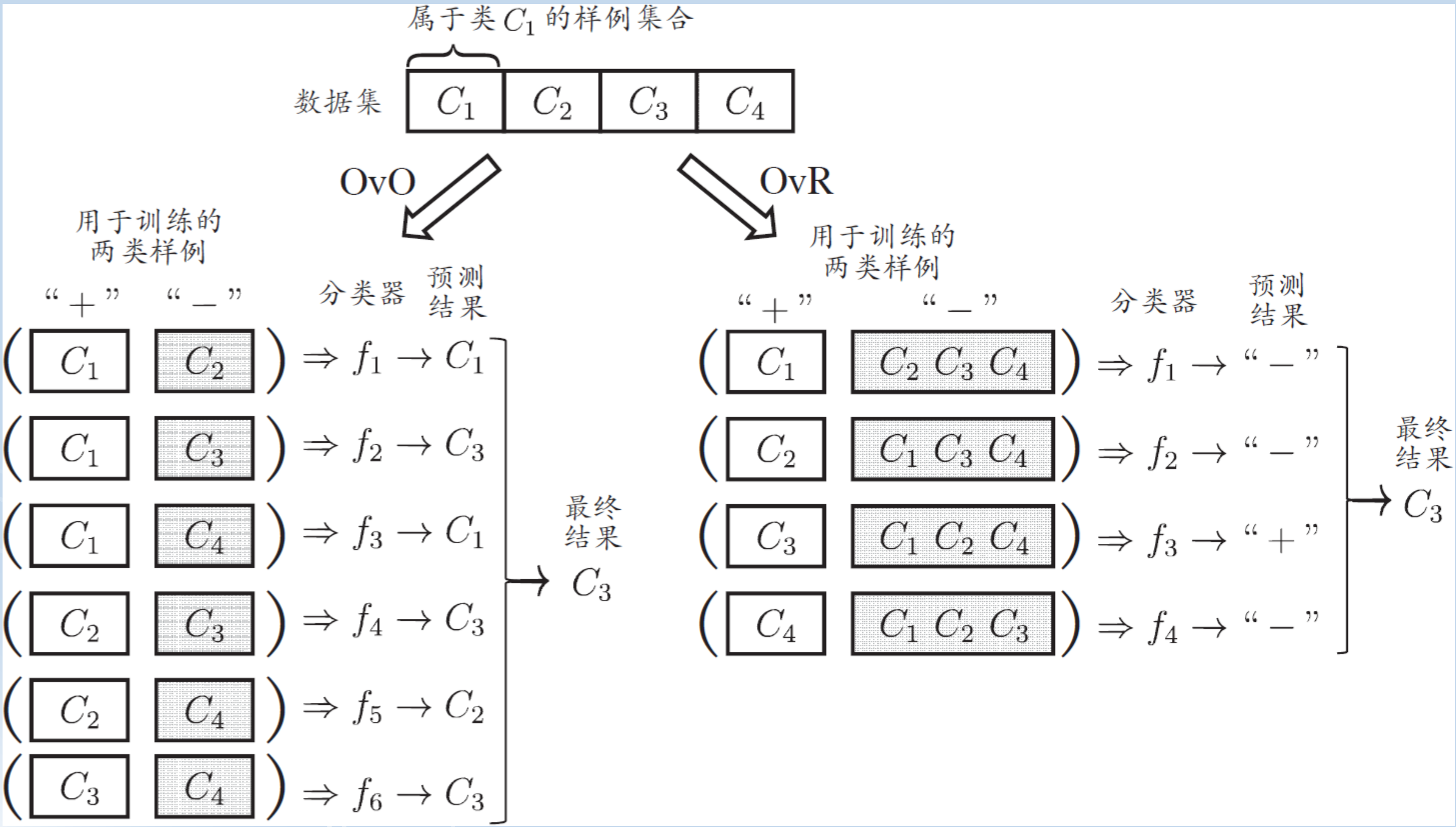
- 拆分阶段
  - N个类别两两配对: N(N-1)/2 个二类任务
  - 各个二类任务学习分类器: N(N-1)/2 个二类分类器
- 测试阶段
  - 新样本提交给所有分类器预测
    - N(N-1)/2 个分类结果
  - 投票产生最终分类结果
    - 被预测最多的类别为最终类别

多分类学习– 一对其余

- 任务拆分
  - ❖ 某一类作为正例，其他反例: N 个二类任务
  - ❖ 各个二类任务学习分类器: 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，存储开销和测试时间大
- 训练只用两个类的样例，训练时间短

一对其余:

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

预测性能取决于具体数据分布，多数情况下两者差不多

多分类学习– 多对多

- 多对多（Many vs Many, MvM）
  - 若干类作为正类，若干类作为反类
- 纠错输出码（Error Correcting Output Code, ECOC）

编码：对N个类别做M次划分，每次划分将一部分类别划为正类，一部分划为反类

M个二类任务
各个类别长度为M的编码

距离最小的类别为最终类别

解码：测试样本交给M个分类器预测

长度为M的编码预测

多分类学习– 多对多

- 纠错输出码(Error Correcting Output Code, ECOC)

	f_1	f_2	f_3	f_4	f_5	海明距离	欧氏距离
C_1	−1	+1	−1	+1	+1	3	$2\sqrt{3}$
C_2	+1	−1	−1	+1	−1	4	4
C_3	−1	+1	+1	−1	+1	1	2
C_4	−1	−1	+1	+1	−1	2	$2\sqrt{2}$
测试示例	−1	−1	+1	−1	+1		

(a) 二元 ECOC 码

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	海明距离	欧氏距离
C_1	−1	−1	+1	+1	−1	+1	+1	4	4
C_2	−1	0	0	0	+1	−1	0	2	2
C_3	+1	+1	−1	−1	−1	+1	−1	5	$2\sqrt{5}$
C_4	−1	+1	0	+1	−1	0	+1	3	$\sqrt{10}$
测试示例	−1	+1	+1	−1	+1	−1	+1		

(b) 三元 ECOC 码

[Dietterich and Bakiri,1995]

[Allwein et al. 2000]

- ECOC编码对分类器错误有一定容忍和修正能力，编码越长、纠错能力越强
- 对同等长度的编码，理论上来说，任意两个类别之间的编码距离越远，则纠错能力越强

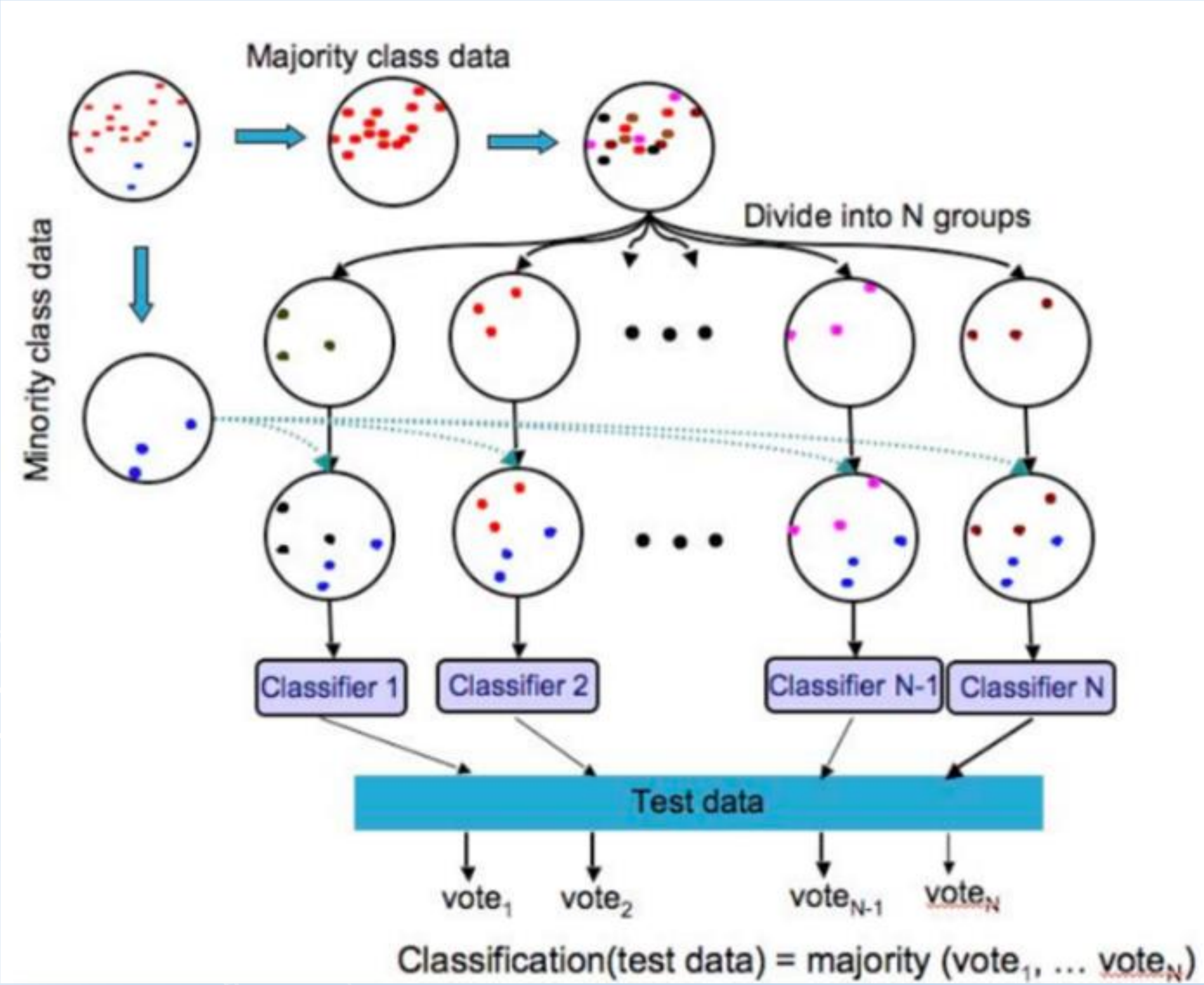
类别不平衡问题

- 类别不平衡（class imbalance）
  - 不同类别训练样例数相差很大情况（正类为小类）

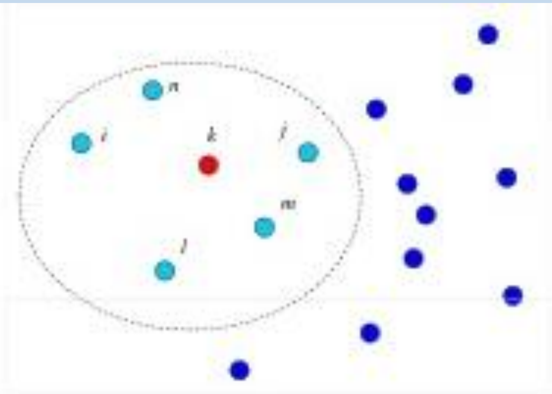
类别平衡正例预测 $\frac{y}{1-y} > 1$ → $\frac{y}{1-y} > \frac{m^+}{m^-}$ 正负类比例

- 再缩放
  - 欠采样（undersampling）
    - 去除一些反例使正反例数目接近（EasyEnsemble）
  - 过采样（oversampling）
    - 增加一些正例使正反例数目接近（SMOTE）
  - 阈值移动（threshold-moving）

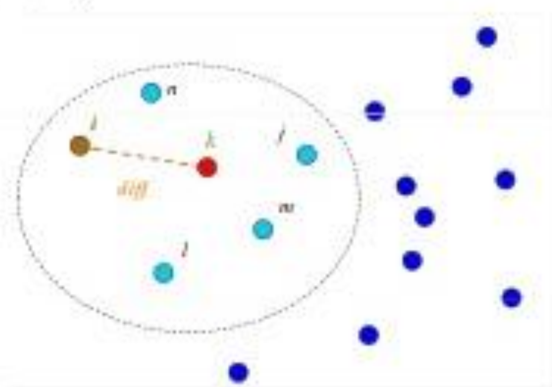
EasyEnsemble



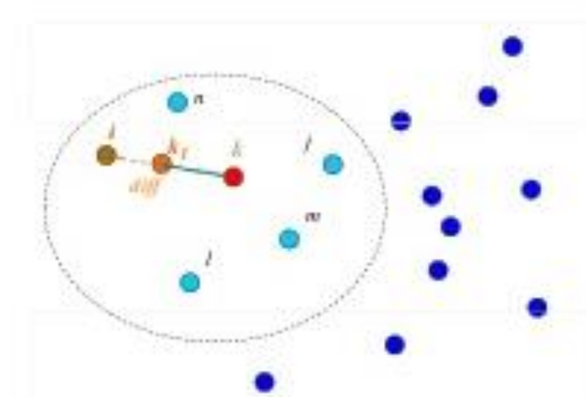
SMOTE



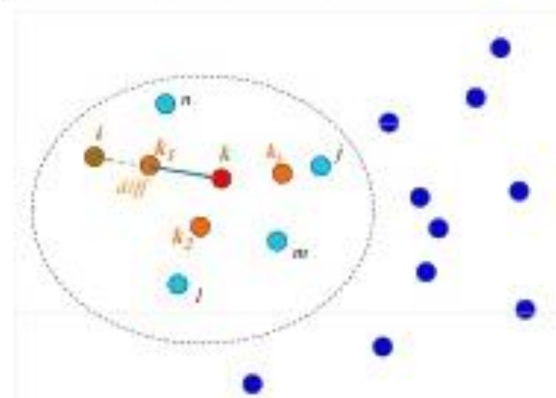
1. For each minority example k compute nearest minority class examples (i, j, l, n, m)



2. Randomly choose an example out of 5 closest points



3. Synthetically generate event k_1, such that k_1 lies between k and i



4. Dataset after applying SMOTE 3 times

优化提要

- 各任务下（回归、分类）各个模型优化的目标
  - 最小二乘法：最小化均方误差
  - 对数几率回归：最大化样本分布似然
  - 线性判别分析：投影空间内最小（大）化类内（间）散度
- 参数的优化方法
  - 最小二乘法：线性代数
  - 对数几率回归：凸优化梯度下降、牛顿法
  - 线性判别分析：矩阵论、广义瑞利商

总结

- 线性回归: 最小二乘法（最小化均方误差）
- 二分类任务
  - 对数几率回归
    - 单位阶跃函数、对数几率函数、极大似然法
  - 线性判别分析
    - 最大化广义瑞利商
- 多分类学习
  - 一对一
  - 一对其余
- 类别不平衡问题
  - 基本策略：再缩放