人工智能的诞生

- **四位学者在1956年提出了人工智能这一术语及研究范畴**
  - John McCarthy（时任Dartmouth数学系助理教授，1971年度图灵奖获得者）
  - Marvin Lee Minsky（时任哈佛大学数学系和神经学系Junior Fellow，1969年度获图灵奖）
  - Claude Shannon（Bell Lab, 信息理论之父）
  - Nathaniel Rochester（IBM, 第一代通用计算机701主设计师）
- 让机器能像人那样认知、思考和学习，即用计算模拟人的智能


A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence

August 31, 1955

当年提出AI概念的建议书

AI概念提出50年后，建议人合影

- 人工智能（Artificial Intelligence）是以机器为载体所展示的人类智能，因此人工智能也被称为机器智能（Machine Intelligence）

人工智能三大学派

符号主义学派

- 又称逻辑学派、计算机学派
- 认为“人的认知基元是符号，认知过程即符号操作过程”
- 认为人和计算机都是物理符号系统， 可以用计算机来模拟人 的智能行为
- 认为人工智能的核心是知识表示、知识推理和知识运用

联结主义学派

- 又称仿生学派或生理学派
- 认为人的思维基元是神经元，而不是符号处理过程
- 认为人脑不同于电脑
- 原理：神经网络及神经网络间的连接机制和学习算法

行为主义学派

又称进化主义或控制论学派

- 认为智能取决于感知和行动
- 主张利用机器对环境作用后的响应或反馈为原型来实 现智能化
- 认为人工智能可以像人类智能一样通过进化、学习来 逐渐提高和增强

第四章　机器学习

- 基本概念
  - 线性回归/线性分类的定义
  - 损失函数：期望风险/经验风险/结构风险最小定义式
  - L1、L2正则的特点
  - 欠拟合/过拟合的定义，多种改善措施
- 监督学习：
  - SVM原理、松弛因子C调整对判别结果的影响
- 半监督学习：
  - 经典代表方法及对应的假设
- 无监督学习：
  - K-means聚类的流程、适用条件
- 集成学习：
  - Bagging和Boosting方法特点、异同

经验风险最小化/结构风险最小化

■期望风险最小化（Expected Risk Minimization）

$$R_{exp} = \int L(f; x, y) dP(x, y)$$

■经验风险最小化(Empirical Risk Minimization，ERM)

$$R_{emp}(f) = \frac{1}{N} \sum_{i=1}^{N} L(f; \boldsymbol{x}^i, y^i)$$

■结构风险最小化(Structural Risk Minimization, SRM)

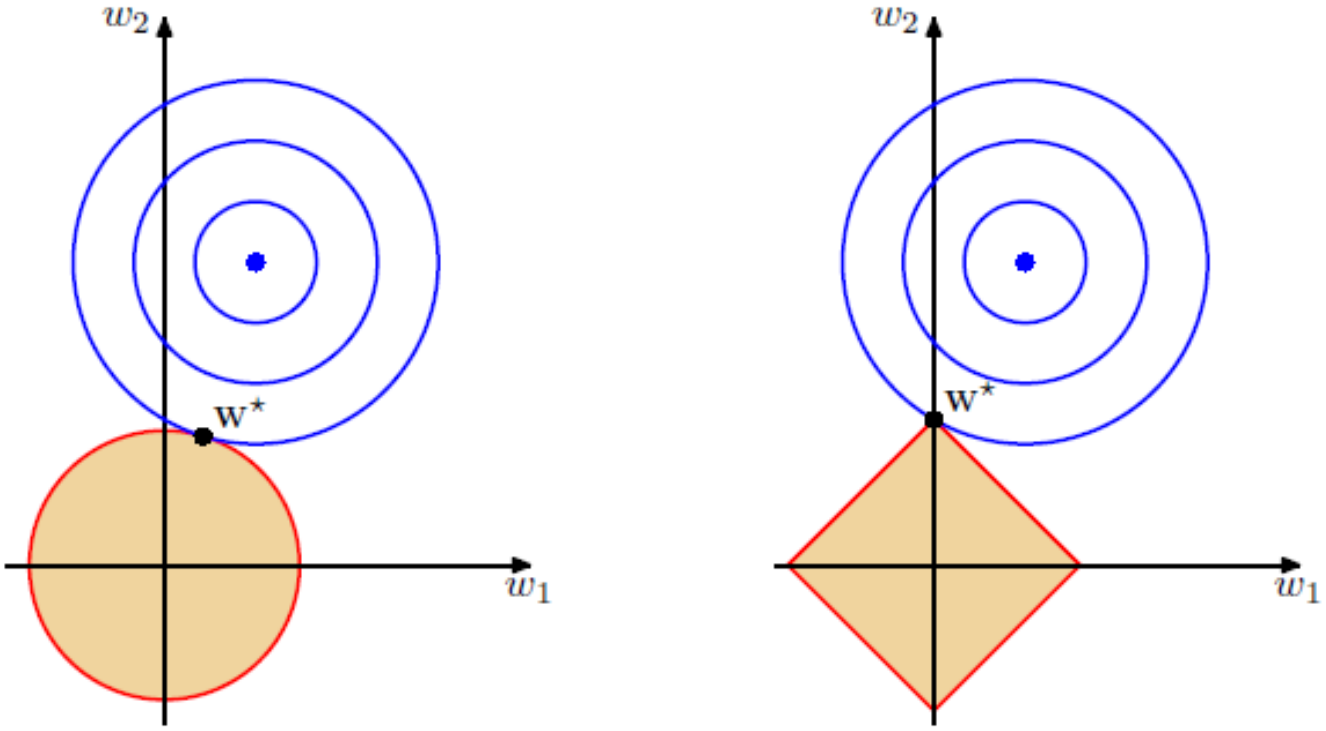
$$R_{srm}(f) = \frac{1}{N} \sum_{i=1}^{N} L(f; \boldsymbol{x}^i, y^i) + \lambda J(f)$$

正则项（Regularizer）/ 惩罚函数（Penalty function）

缓解“过拟合”

24春季 1800860854O4P2001H 计算机科学与技术学院

L1 和 L2 范数的效果



- 使用L_2范数直接拉向原点
- 使用L_1范数会拉向坐标轴，即尝试将某些坐标设置为0.
- 其它损失函数: Hinge loss, Huber loss, Cross-entropy, Exponential loss, quadratic loss,…

24春季 1800860854O4P2001H 计算机科学与技术学院

提高模型性能

- 欠拟合：当模型处于欠拟合状态时，根本的办法是增加模型复杂度。
  - 增加模型的迭代次数
  - 更多特征
  - 降低模型正则化水平
- 过拟合：当模型处于过拟合状态时，根本的办法是降低模型复杂度。
  - 及早停止迭代
  - 减少特征数量
  - 提高模型正则化水平
  - 扩大训练集

计算机科学与技术学院 24春季 1800860854O4P2001H

小结：线性SVM

■输入：线性可分的训练数据集 $S=\{(\boldsymbol{x}^i,y^i),i=1,...,N\}$

■输出：判别函数及决策/判别界面

•通过求解如下最优化问题来得到最优分类器的参数 $(\boldsymbol{w}^*,b^*)$

$$\min_{\boldsymbol{w},b}\frac{1}{2}\|\boldsymbol{w}\|_2^2$$

$$s.t.\ \ y^i\left(\boldsymbol{w}^T\boldsymbol{x}^i+b\right)\geq 1,i=1,...,N$$

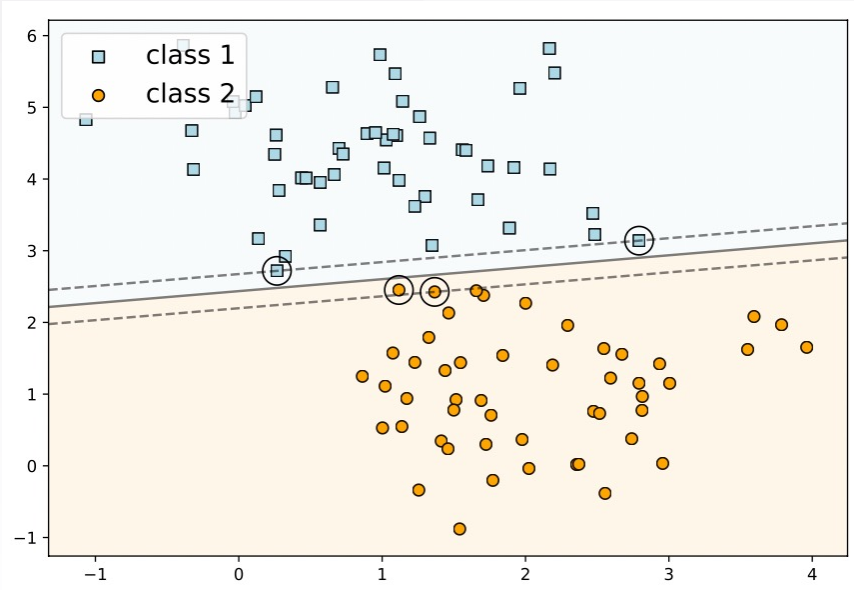
•判别/分离超平面：$(\boldsymbol{w}^*)^T\boldsymbol{x}+b^*=0$

•判别函数：$f_{w,b}(\boldsymbol{x})=sign((\boldsymbol{w}^*)^T\boldsymbol{x}+b^*)$

理论保证：对于线性可分的训练数据集，最大间隔分类器存在且唯一。

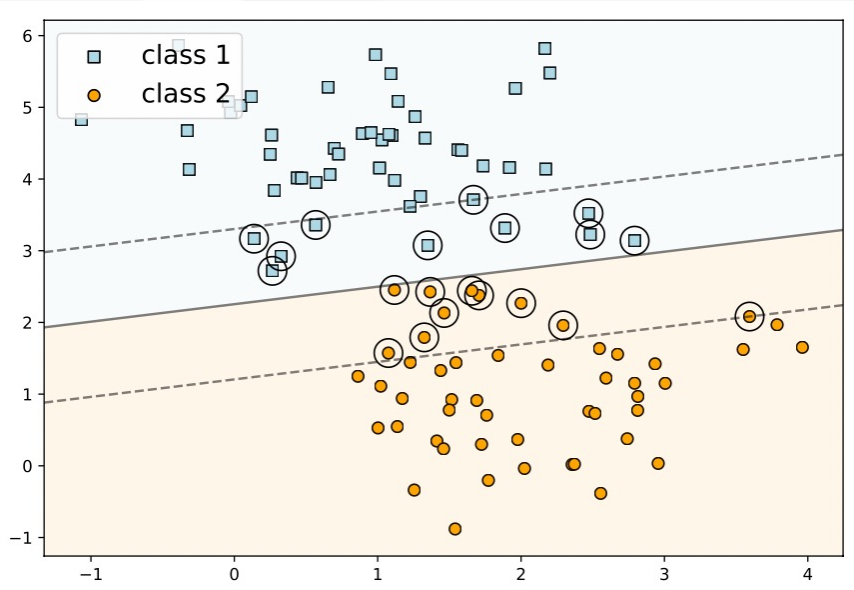
小练习2: 软间隔SVM

$$\min_{w,b} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{N}\xi_i$$



$C = 10.0$

C较大，
错误分类误差容忍低，
较小的间隔



$C = 0.1$

C较小，
错误分类误差的容忍高，
较大的间隔

小结 半监督学习

- 通用想法: 同时从有标注和无标注数据学习
- 假设:
  - 平滑假设 (生成式)
  - 流形假设 (基于图)
  - 聚类/低密度分割假设 (S^3VM)
  - 独立假设 (协同学习)
- 使用无标注数据的方式:
  - 引入损失函数 (S^3VM,协同学习)
    优化方法很重要
  - 正则化 (图方法)
    图的构建很关键

24春季 1800860854O4P2001H 计算机科学与技术学院

K-Means聚类

使用条件：数值型数据

适合于球型簇

簇大小相近的数据集

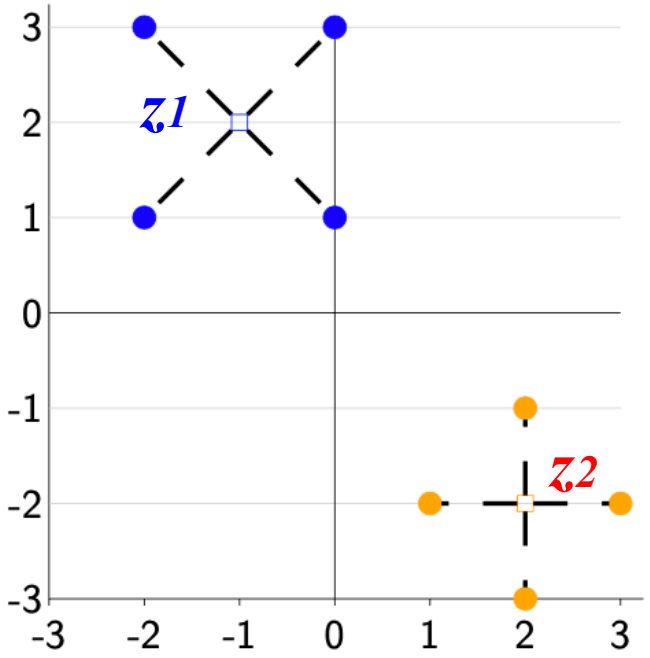
数据量适中，因为数据量太会影响效率

■损失函数

$$\text{Loss}_{\text{kmeans}}(\mathbf{z}, \boldsymbol{\mu}) = \sum_{i=1}^{n} \|\phi(x_i) - \mu_{z_i}\|^2$$

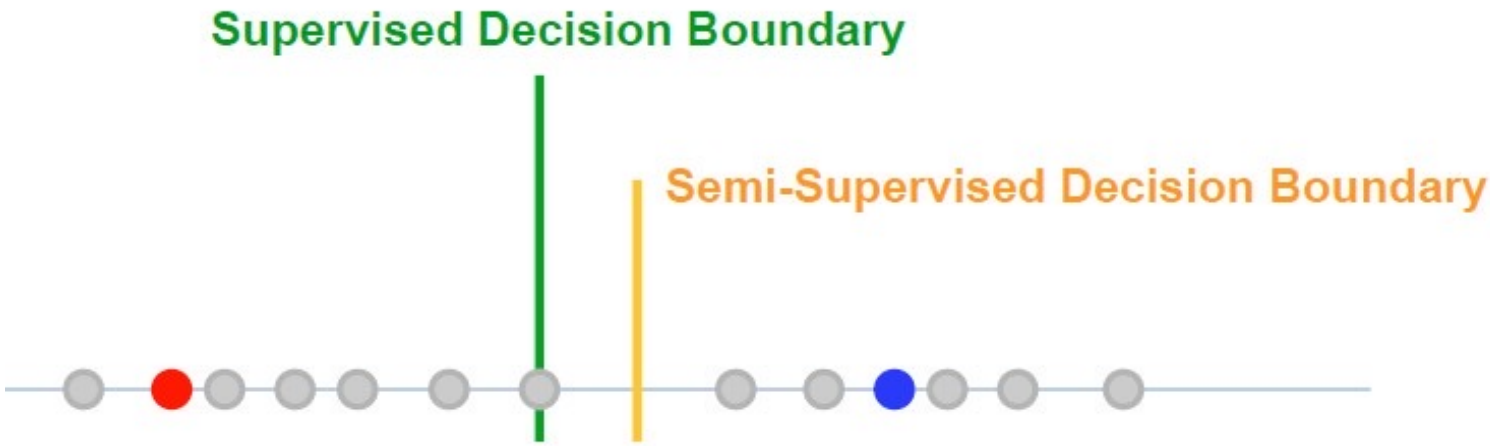
■优化目标

$$\min_{\mathbf{z}} \min_{\boldsymbol{\mu}} \text{Loss}_{\text{kmeans}}(\mathbf{z}, \boldsymbol{\mu})$$



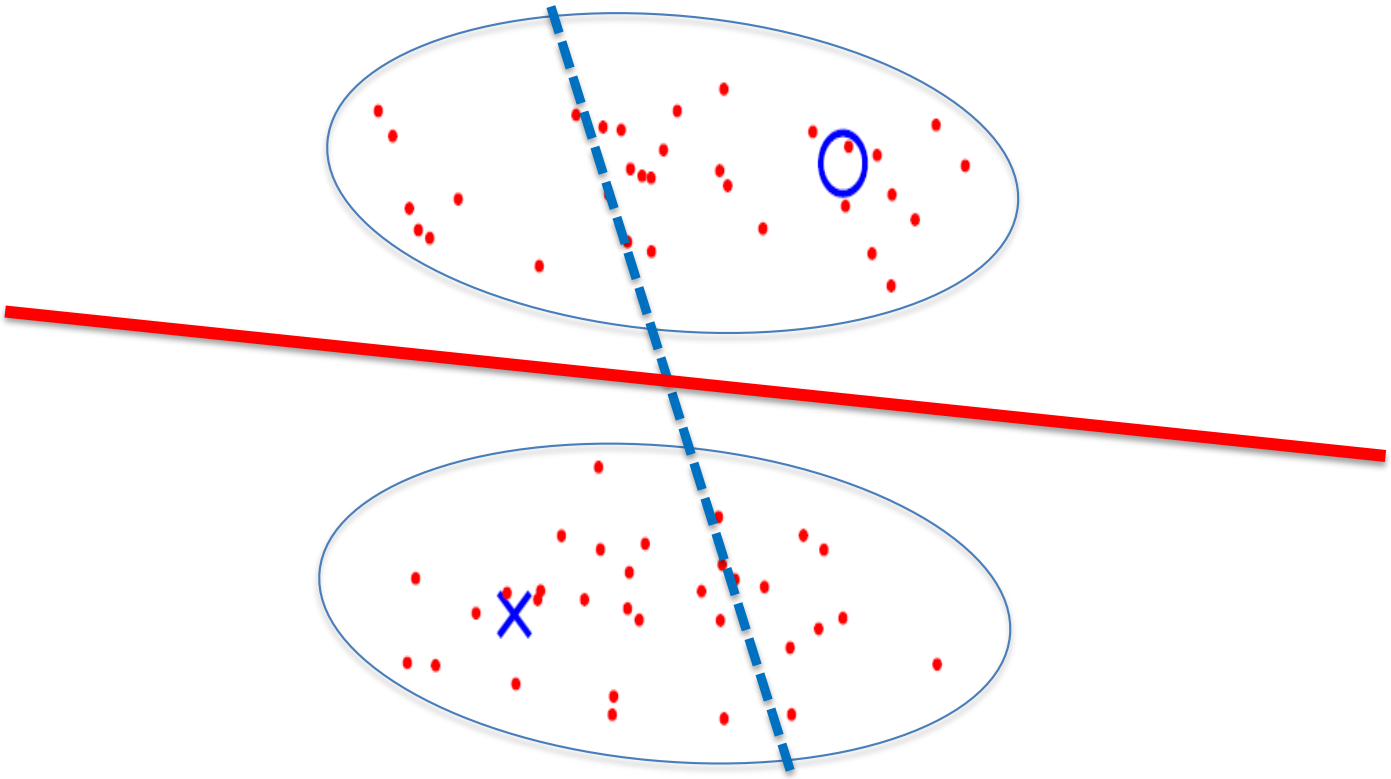
平滑假设(smoothness assumption)

- 半监督学习的平滑假设:
  - 如果高密度空间中两个点$\mathbf{x}_1, \mathbf{x}_2$距离较近, 那么对应的输出$y_1, y_2$也应接近
- 监督学习的平滑假设(用于对比):
  - 如果空间中两个点$\mathbf{x}_1, \mathbf{x}_2$距离较近, 那么对应的输出$y_1, y_2$也应接近



聚类假设（cluster assumption）

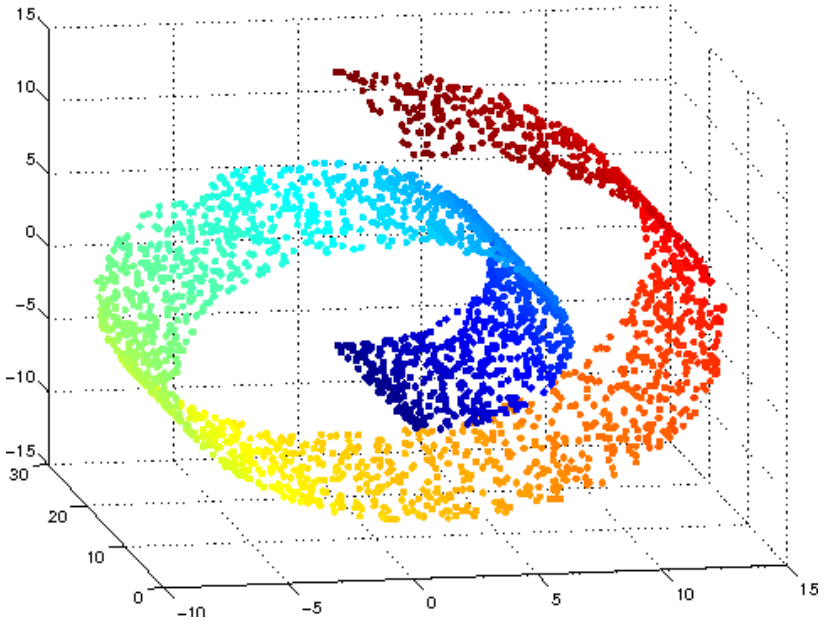
- 聚类假设
  - 如果两点在同一个簇，那么它们很有可能属于同一个类
- 等价表述: 低密度分割
  - 决策边界应该在低密度区域

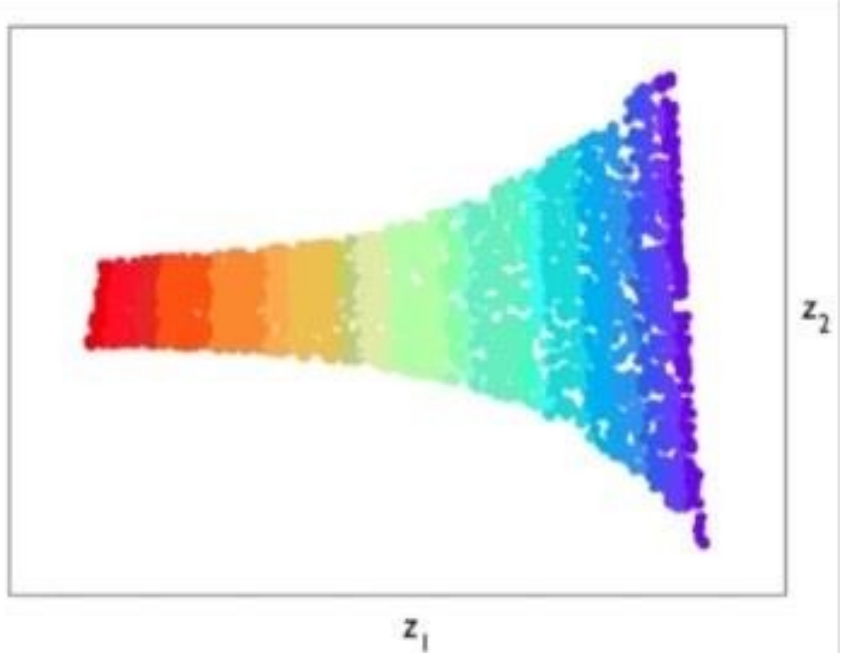


流形假设(manifold assumption)

流形假设

- 高维数据大致会分布在一个低维的流形上
- 邻近的样本拥有相似的输出
- 邻近的程度常用“相似”程度来刻画





计算机科学与技术学院
24春季 1800860854O4P2001H

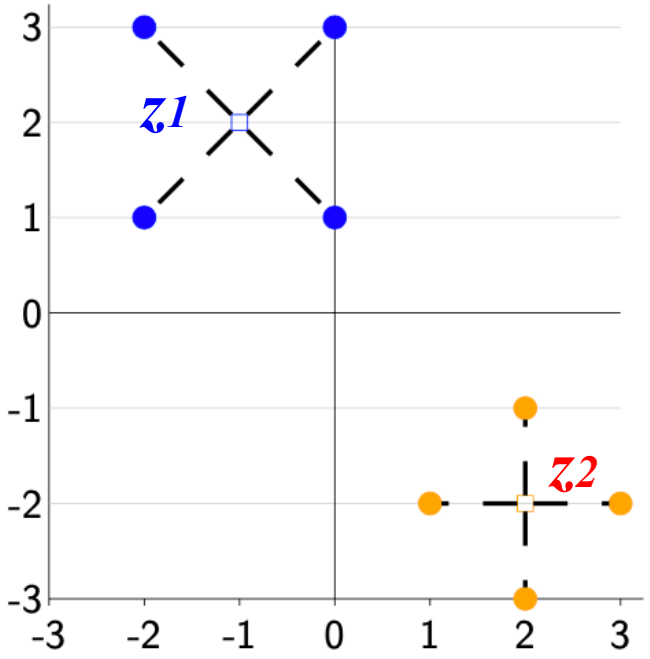
K-Means聚类

■损失函数

$$\text{Loss}_{\text{kmeans}}(\mathbf{z}, \boldsymbol{\mu}) = \sum_{i=1}^{n} \|\phi(x_i) - \mu_{z_i}\|^2$$

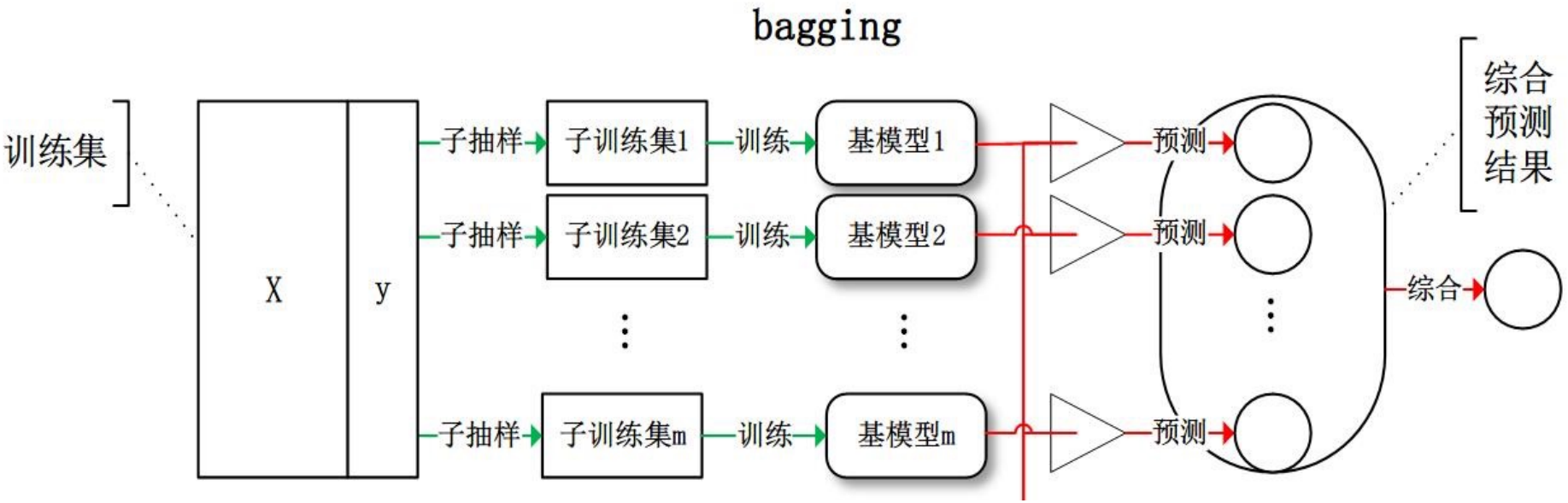
■优化目标

$$\min_{\mathbf{z}} \min_{\boldsymbol{\mu}} \text{Loss}_{\text{kmeans}}(\mathbf{z}, \boldsymbol{\mu})$$



Bagging

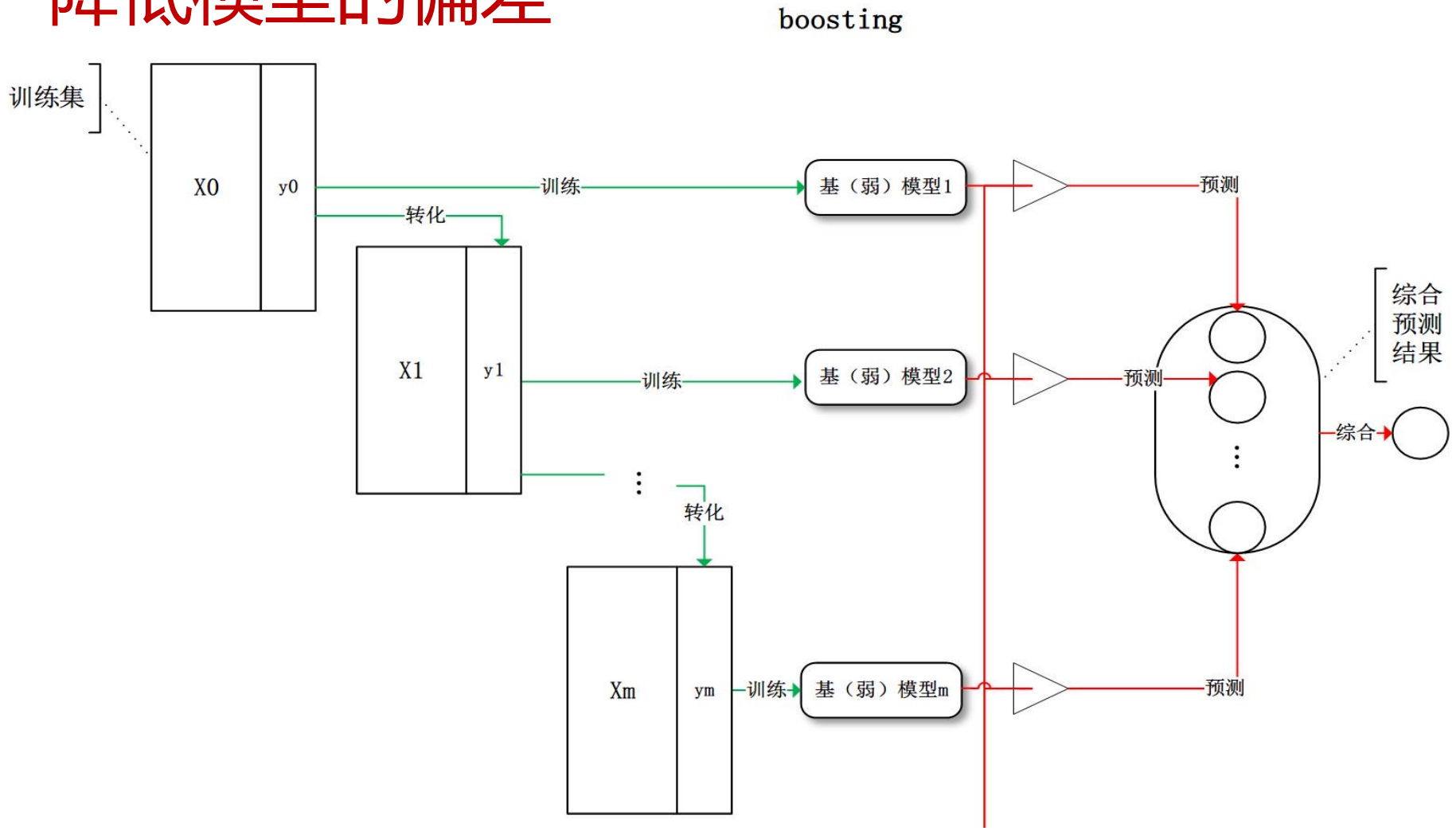
- 降低模型的方差



https://www.cnblogs.com/jasonfreak/p/5657196.html

Boosting

- 降低模型的偏差



Bagging和Boosting是两种常见的集成学习技术，它们通过结合多个弱学习器（例如决策树）来提高模型的性能。对应随机森林，GBDT

相同点
集成学习方法：两者都是集成学习方法，旨在通过组合多个弱学习器来构建一个强大的预测模型。
提高模型性能：两者都旨在减少模型的偏差和方差，从而提高预测性能和泛化能力。
多样性：通过引入弱学习器的多样性来提高整体模型的性能。

不同点

Bagging：并行训练，

最终结果通过平均（回归）或投票（分类）来结合各个模型的预测结果。

主要减少模型的方差，提高稳定性和准确性。

boosting，串行训练,每个模型试图纠正前一个模型的错误。

样本权重调整：每次训练后，根据前一个模型的预测错误调整样本的权重，错误样本的权重增加，使后续模型更加关注这些难分类的样本。

最终结果通过加权平均或加权投票来结合各个模型的预测结果。

通过逐步减少模型的偏差，同时可能也会减少方差。