

机器学习：Day1实训内容

李猛 2017-08-10
V1.0.0.0

The World's Local Training Provider

目录

MENU

1

序言

2

机器学习概论

3

分类算法介绍

4

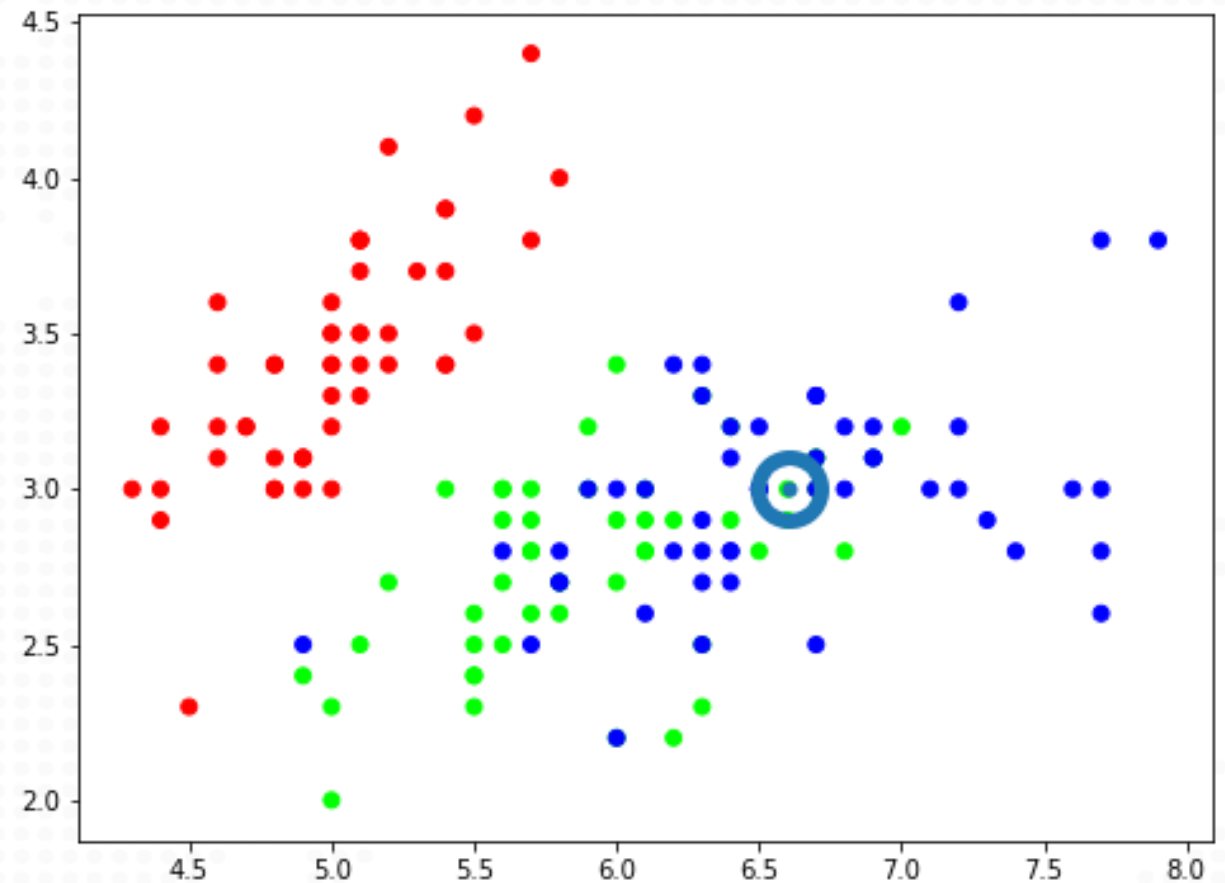
总结Q&A

5

结语

3.1.1: NN近邻: 模型概念介绍

- NN近邻介绍: Nearest Neighbor
- 模型原则/模型来源
- 说明: 在一个多维度的几何空间中, 寻找距离目标点最近的其它点, 基于其它点的特征标记目标点
- 如图: 目标点圆圈点, 周边的点有3中不同的颜色



3.1.2: NN近邻: 算法策略

- 欧氏空间距离定理, 如右图

$$d(x, y) := \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

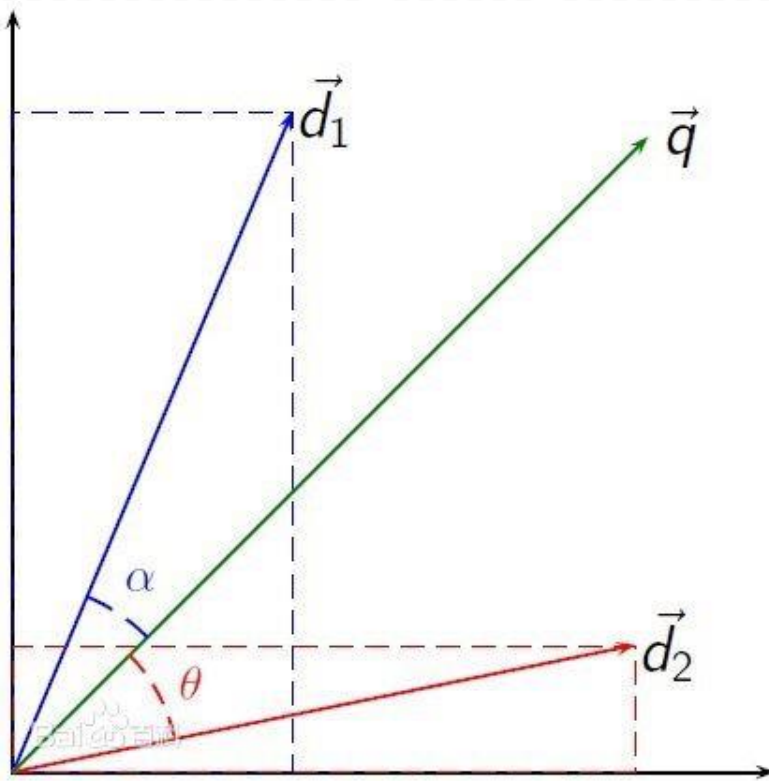
- 曼哈顿距离, 如右图

$$d_{12} = |x_1 - x_2| + |y_1 - y_2|$$

- 余弦夹角定理, 如右图

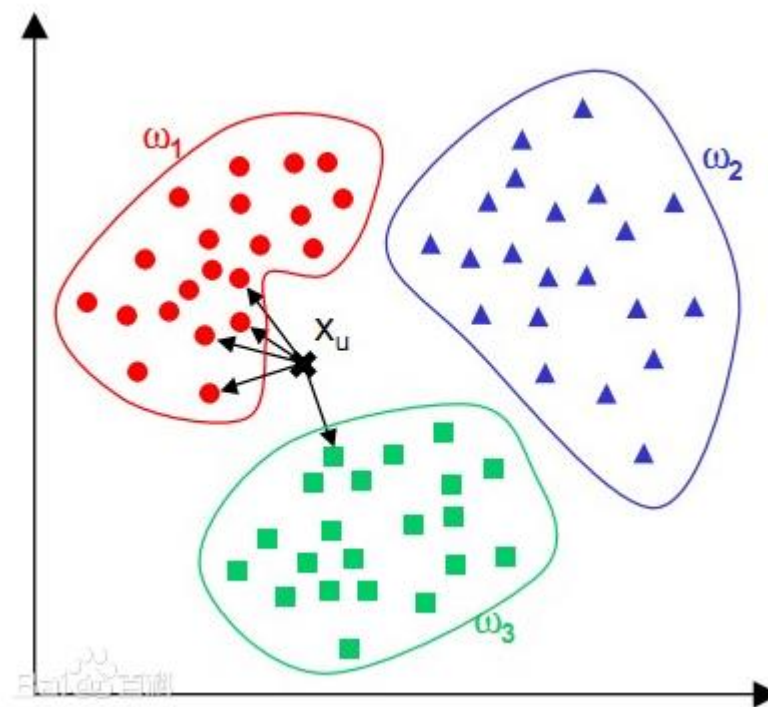
- 闵式距离, 一组距离

- 互动Q&A



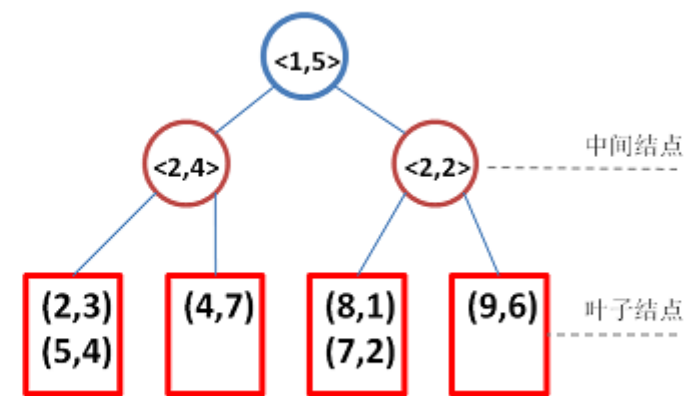
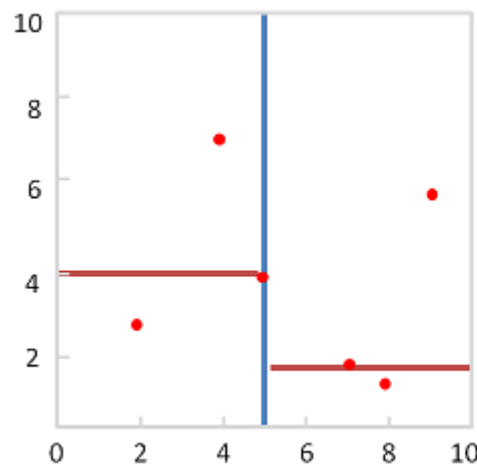
3.1.3: NN近邻: K-NN

- K-NN近邻介绍
- 定义: 基于周边K个已经标记分类的点确认目标点的分类标记
- 算法: 欧氏距离
- 算法优缺点
- 案例: jupyter
- Q&A: 总结感想



3.2.2: NN近邻: KD-Tree

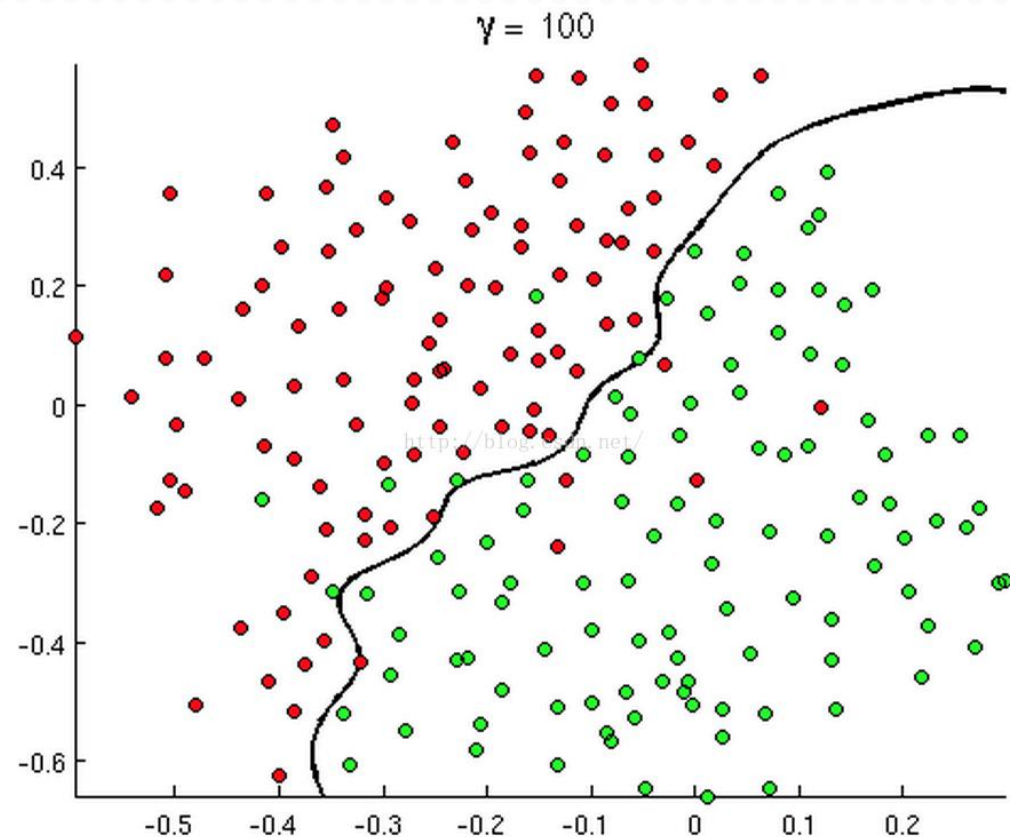
- KD-Tree介绍
- 模型原则
- 模型思路: 建立索引数据, 在多维空间上基于二叉树划分数据点; 查询时, 按照构建流程先找到点附近的空间
- 算法选择: 欧氏距离



- 案例: jupyter
- Q&A: 互动感想

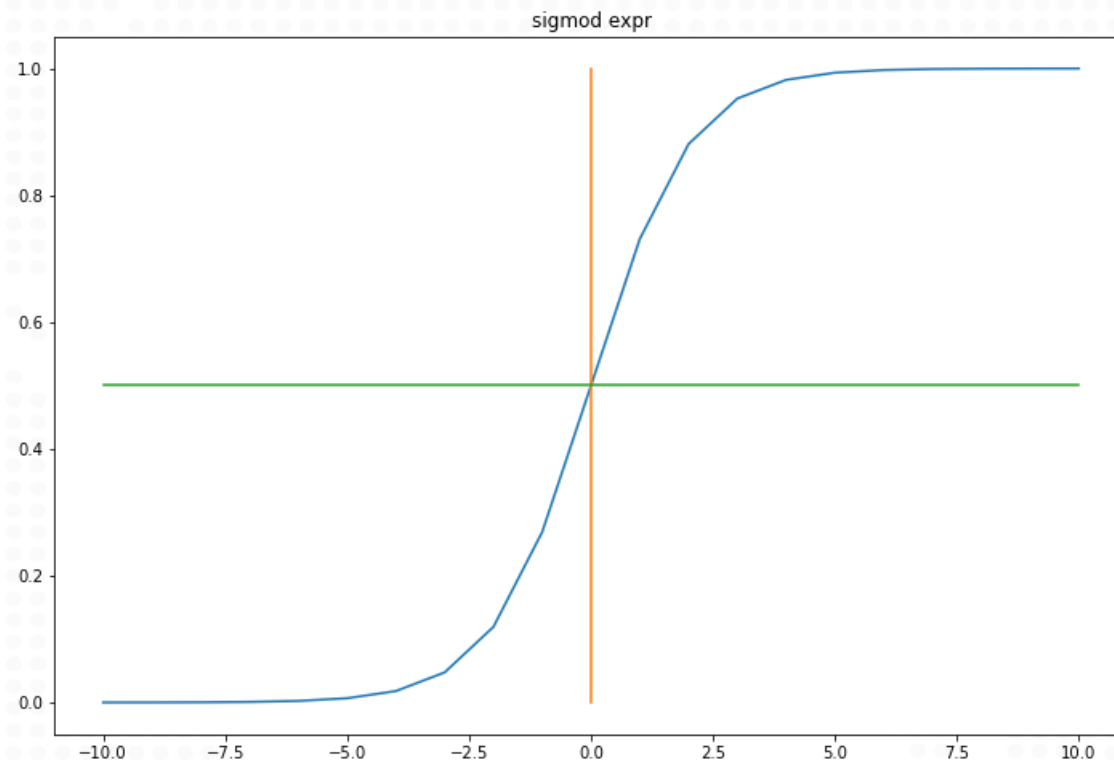
3.2.1: LR逻辑回归

- 逻辑回归介绍
 - 利用线性回归做分类，如右图
 - Logistic函数
 - 自变量（特征X）、因变量（标记Y）
-
- 案例：jupyter
 - Q&A：互动交流



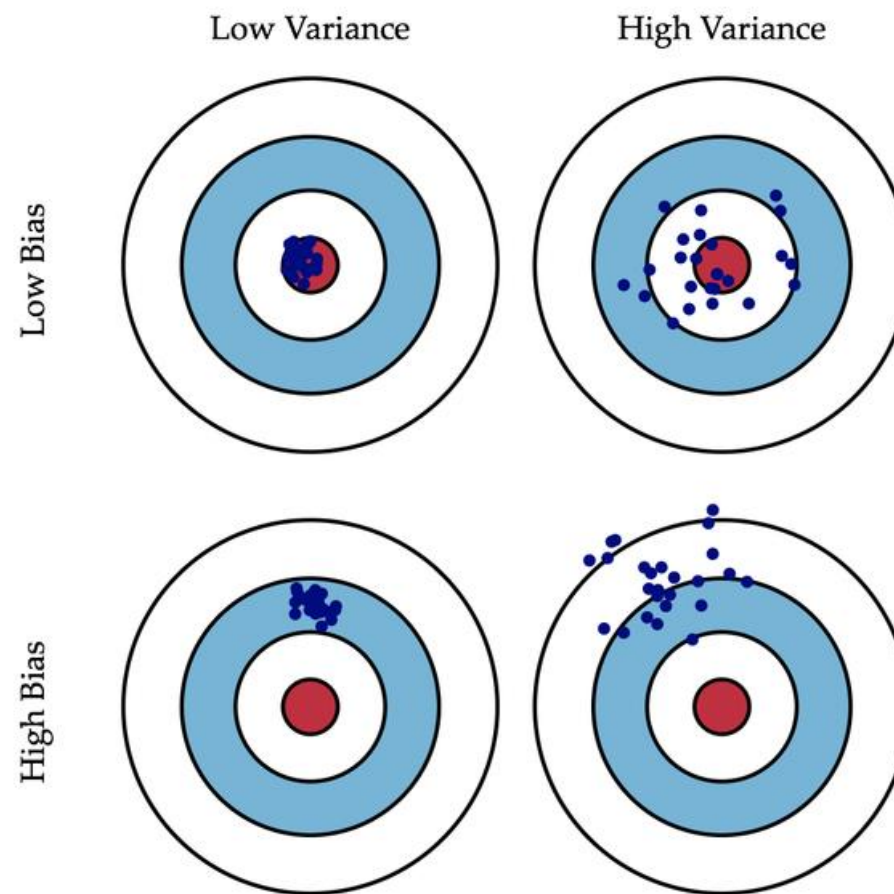
3.2.2: LR逻辑回归

- 概率范围[0,1]之间, ≥ 0.5 或者XX
 - $P(y=\text{分类1})+P(y=\text{分类2})=1$
 - 概率分布函数: `sigmod`
 - 损失函数
 - 最小二乘法
 - 惩罚函数
 - 梯度下降gd
 - 随机梯度下降sgd
-
- 案例: `jupyter`
 - Q&A: 与其它平台对比



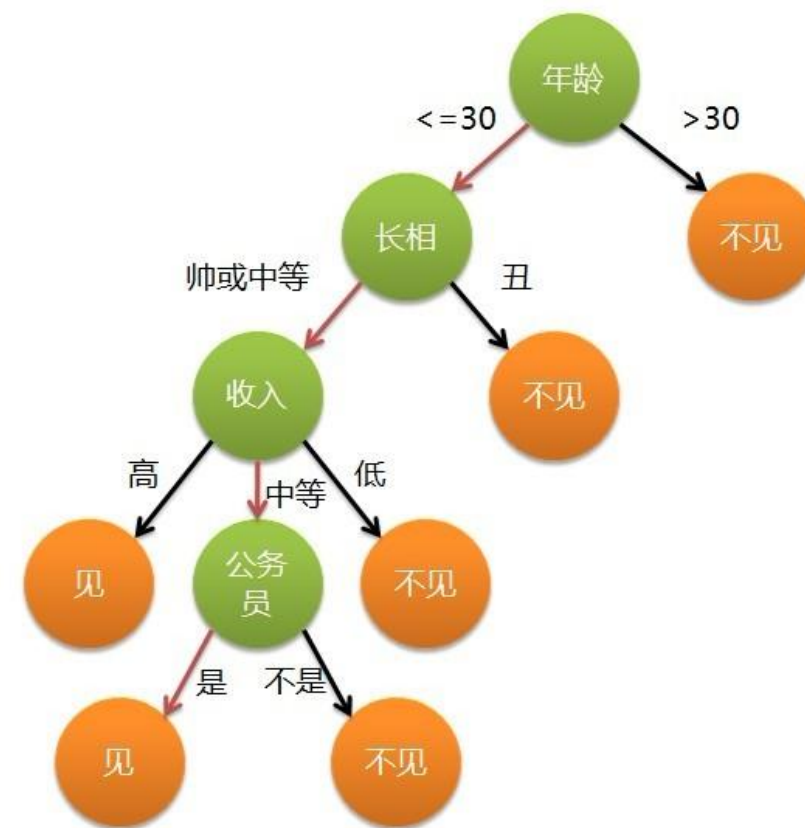
3.2.3: LR逻辑回归: 偏差+方差

- 偏差, 命中准确
- 方差, 密度分布
- 案例: 射击打靶
- 案例: jupyter
- Q&A: 与其它平台对比



3.3.1: DT决策树

- 决策树介绍
- 决策树模型，如图
- 决策树算法3要素：
 - 特征选择
 - 决策树生成
 - 决策树剪枝
- 优点
- 缺点
- 互动Q&A



3.3.2: DT决策树: 构造决策树

- 熵(Entropy)
- 信息熵与条件熵
- 香农教授: 信息论
- ID: 信息增益
- C4.5: 信息增益率
- Cart: gini系数
- 案例: jupyter
- Q&A: 互动交流

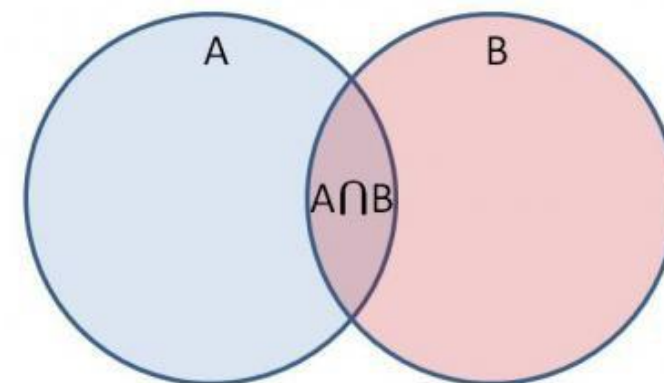
3.4.1: Bayes 贝叶斯: 朴素贝叶斯

- 贝叶斯介绍
- 朴素贝叶斯
- 先验概率/边缘概率: $P(A)$
- 后验概率/条件概率: $P(B|A)$

A已发生的条件下B发生的概率:

$$P(B | A) = P(A \cap B) / P(A)$$

$$P(A \cap B) = P(B | A) \times P(A)$$

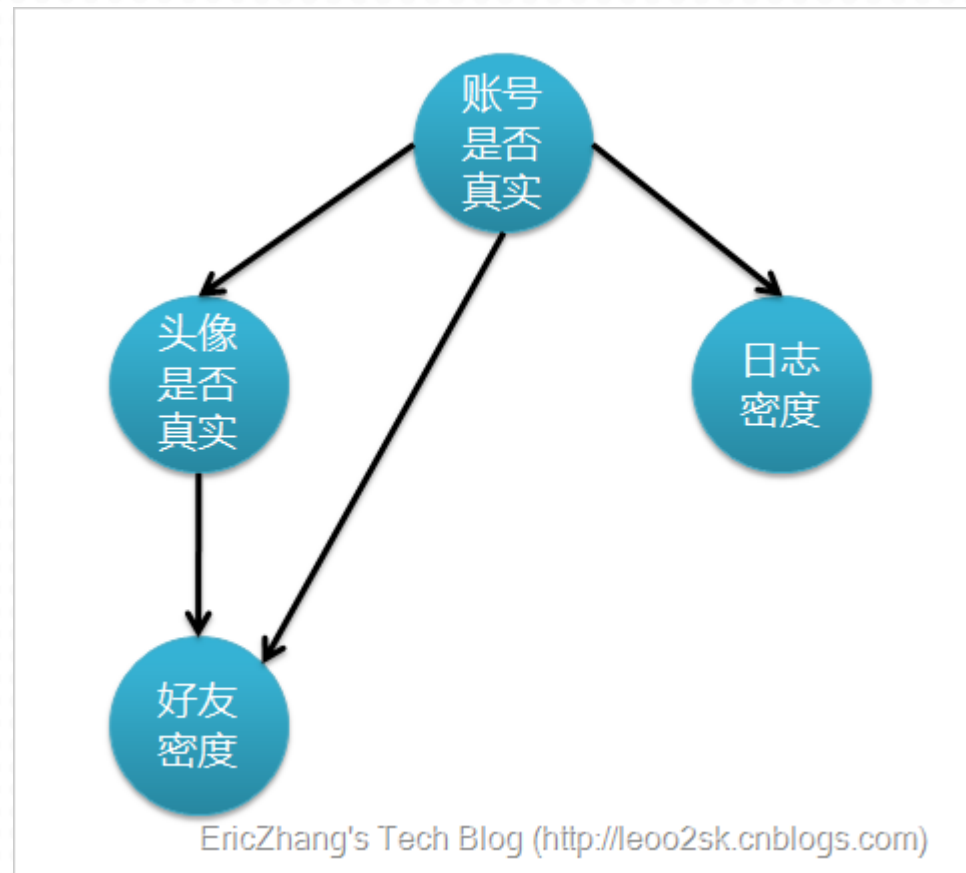


Barco 白蜡

- 案例: jupyter
- Q&A: 互动交流

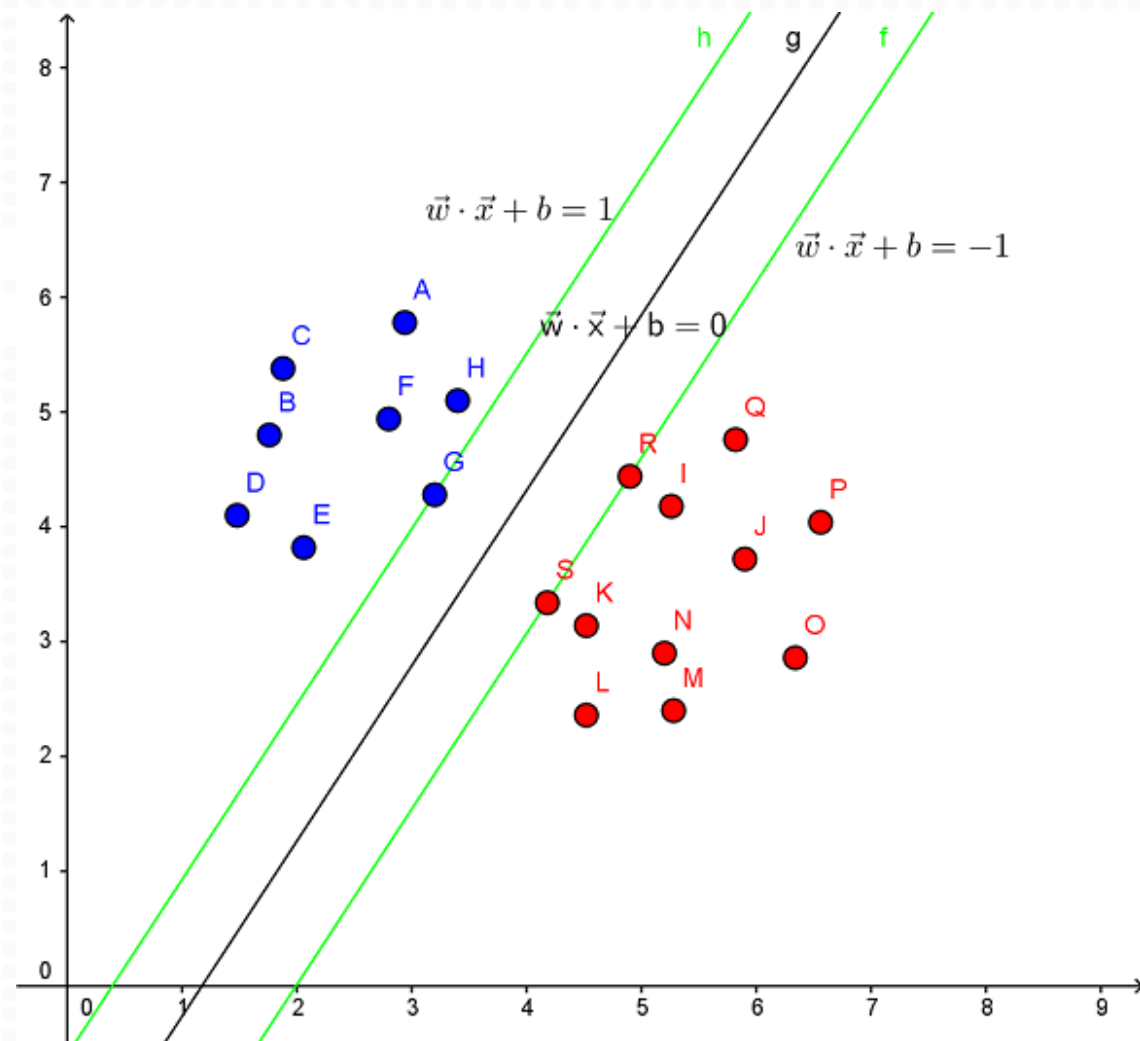
3.4.2: Bayes 贝叶斯: 贝叶斯信念网络

- 贝叶斯信念网络
- 案例: 好友密度估计
- Q&A: 交流



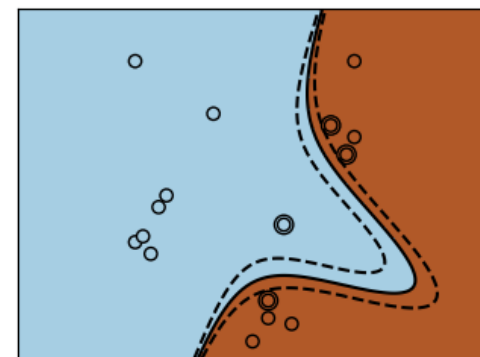
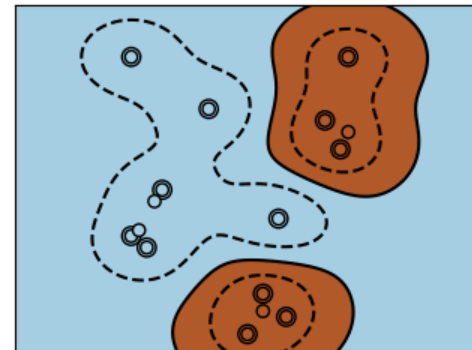
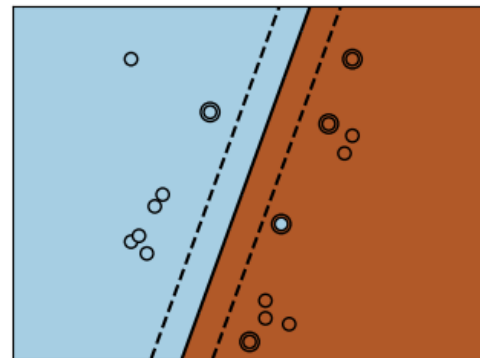
3.5.1: SVM支持向量机

- 支持向量机介绍
- 支持向量
- Q&A: 互动交流



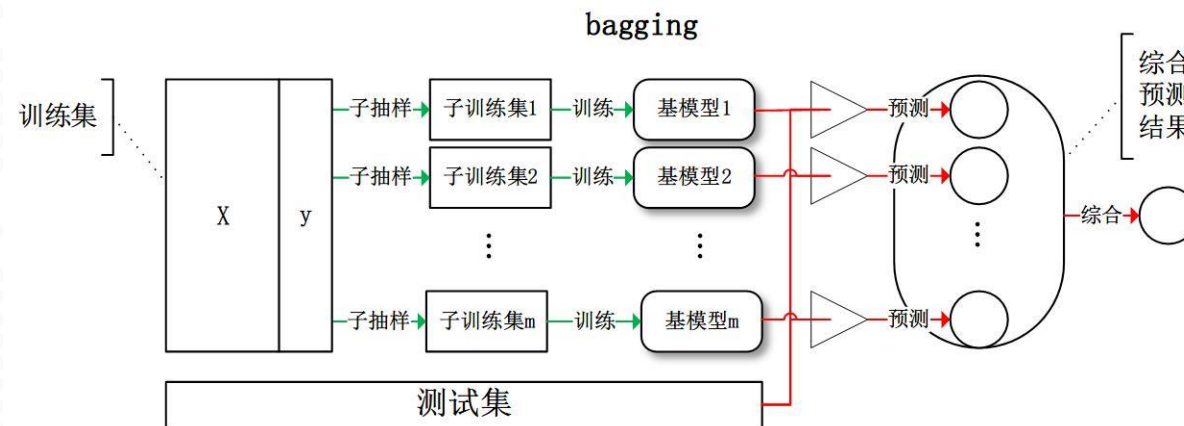
3.5.2: SVM支持向量机

- 线性支持向量机
 - 核函数Kernel
 - Poly多边形核函数
 - Linear线性核函数
 - RBF径向基核函数
-
- 案例: jupyter
 - Q&A: 互动交流



3.6.1: Ensemble集成模型：概念介绍

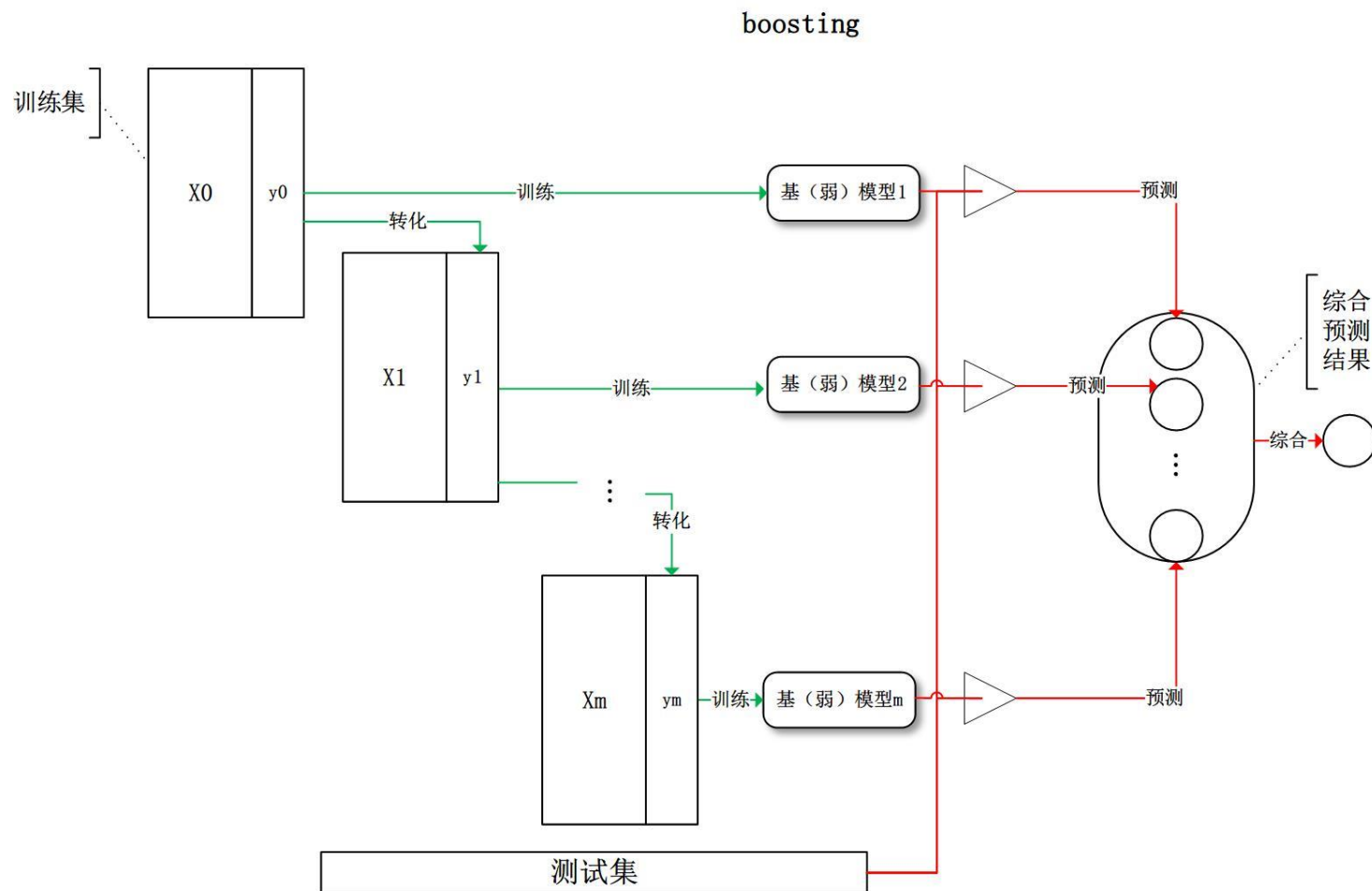
- 集成算法理论基础
- Bagging



- 案例: jupyter
- Q&A: 互动交流

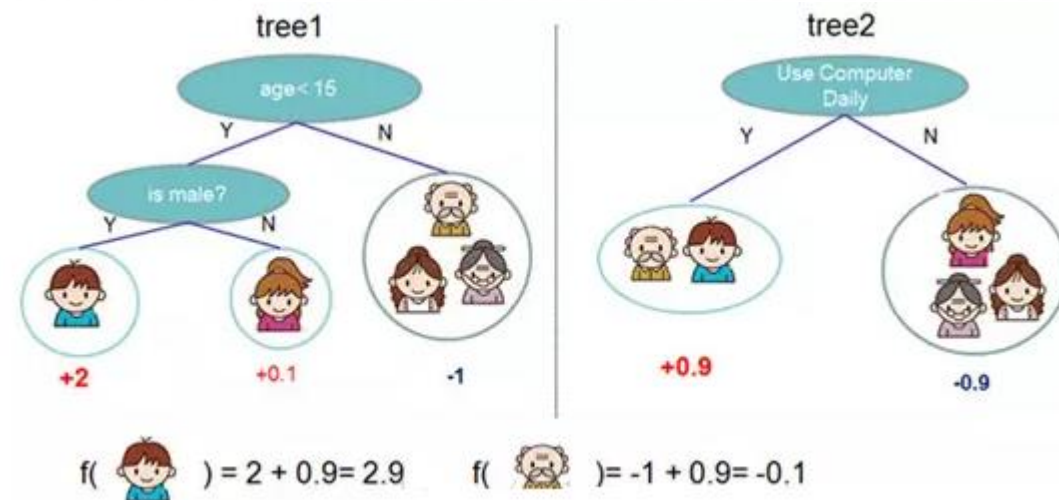
3.6.2: Ensemble集成模型: adaboost模型

- Adaboost模型介绍
 - 弱学习
 - 强学习
-
- 案例: jupyter
 - Q&A: 互动交流



3.6.3: Ensemble集成: 决策树森林

- 模型介绍
- 算法介绍
- 案例: jupyter
- Q&A: 互动交流



4.1.1: 分类总结：模型对比

- 模型对比
- 预测对比
- 性能对比
- 最优模型选择
- SVM支持向量机
- Ensemble集成模型
- Q&A: 互动交流

4.1.2: 分类总结：综合练习

- 练习：同一数据+同一模型+不同参数
- 练习：同一数据+不同模型
- 练习：同一数据+不同模型+不同参数
- 案例：jupyter
- Q&A：互动交流

5.1.1: 结语

- 什么模型，如何看待模型
 - 什么是策略，如何看待策略
 - 什么是算法，如何看待算法
 - 数学基础、如何看待数学
-
- 书籍：数学之美/算法帝国
 - Q&A：互动交流



“给我一个支点，我就能撬起整个地球。”

▲ 阿基米德

联系方式

- Mail:1789909854@qq.com
- QQ:1789909854
- [Tel:17621063575](tel:17621063575)
- Wechat:ynuosoftware

