

分类和回归算法的总结归纳

肖易敏

目 录

CONTENTS

01

分类算法

03

小结

02

回归算法

分类算法



回归算法



Part

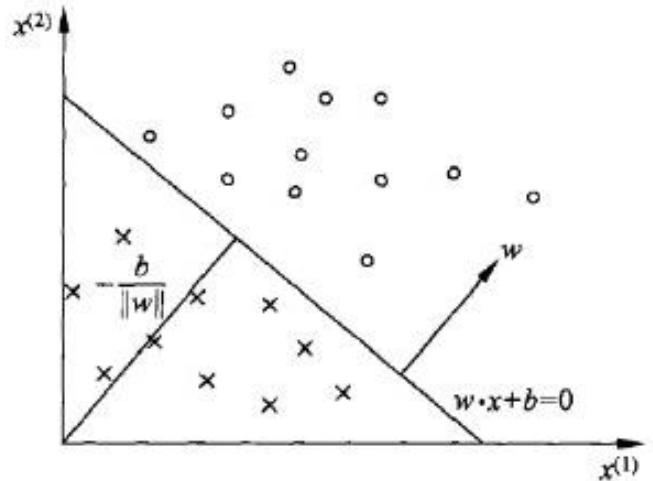
01

分类算法

1.1 感知机

□ 感知机

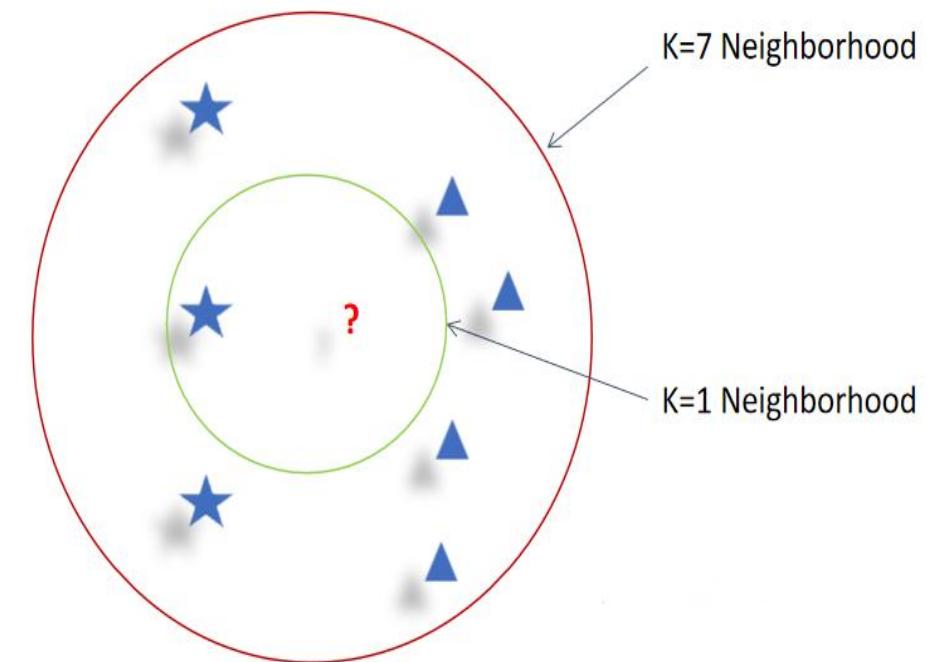
- 模型特点：分离超平面
- 模型类型：判别模型 $f(x) = \text{sign}(w \cdot x + b)$
- 学习策略：极小化误分点到超平面距离
- 学习算法：随机梯度下降
- 解决非线性可以考虑多层感知机



1.2 k近邻法

□ k近邻法

- 模型特点：特征空间和样本点
- 模型类型：判别模型
- 要素：k值的选择、距离度量及分类决策规则



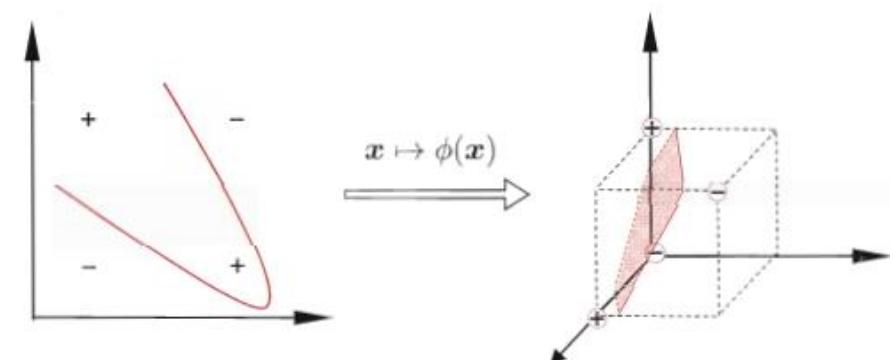
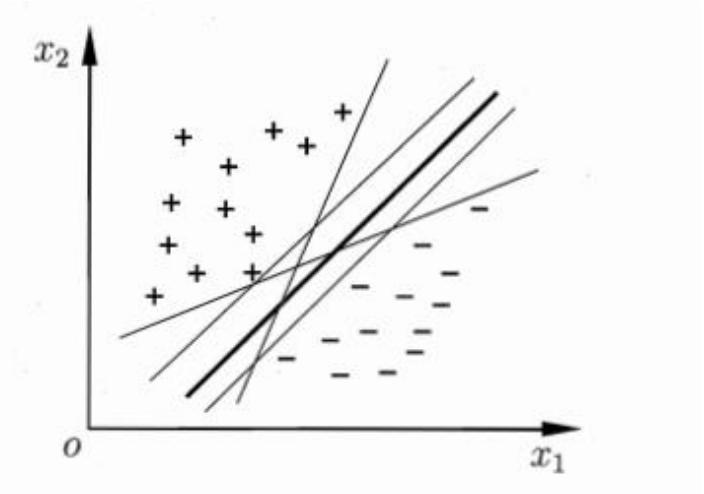
口 朴素贝叶斯法

- 模型特点：特征与类别的联合概率分布，条件独立假设
- 模型类型：生成模型 $y = \arg \max_{c_k} P(Y = c_k) \prod_j P(X^{(j)} = x^{(j)} | Y = c_k)$
- 学习策略：极大似然估计
- 学习算法：概率计算公式

1.4 支持向量机

□ 支持向量机

- 模型特点：分离超平面，核技巧
- 模型类型：判别模型
- 学习：对偶问题求解



1.5 决策树

□ 决策树

- 模型特点：分类树
- 模型类型：判别模型
- 推理过程容易理解，决策推理过程

可以表示成If Then形式

- 学习过程：特征选择、树的生成、树的剪枝

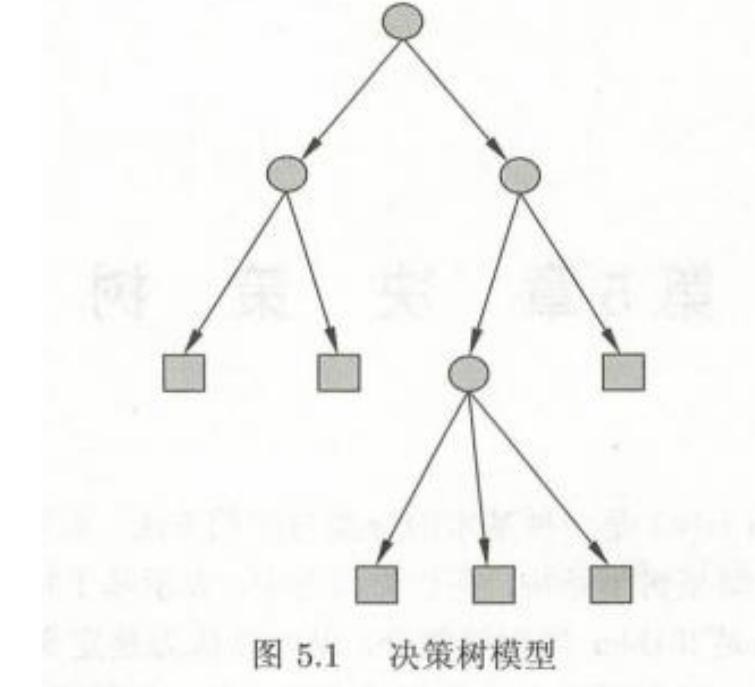


图 5.1 决策树模型

1.6 提升方法

口 提升方法

- 模型特点：弱分类器的线性组合
- 模型类型：判别模型
- Adaboost：提高那些被前一轮

$$\left. \begin{array}{l} h_1(x) \in \{-1, +1\} \\ h_2(x) \in \{-1, +1\} \\ \vdots \\ h_T(x) \in \{-1, +1\} \end{array} \right\} \quad \text{Weak classifiers} \quad \text{strong classifier}$$
$$H_T(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x) \right)$$

弱分类器错误分类样本的权值，降低那些被正确分类样本的权值；加权多数表决

Part

02

回归算法

2.1 线性回归

□ 线性回归

- 模型特点：线性
- 模型类型：判别模型 $f(x_i) = w^T x + b$, 使得 $f(x_i) \simeq y_i$.
- 对数线性回归：非线性 $\ln y = w^T x + b$.

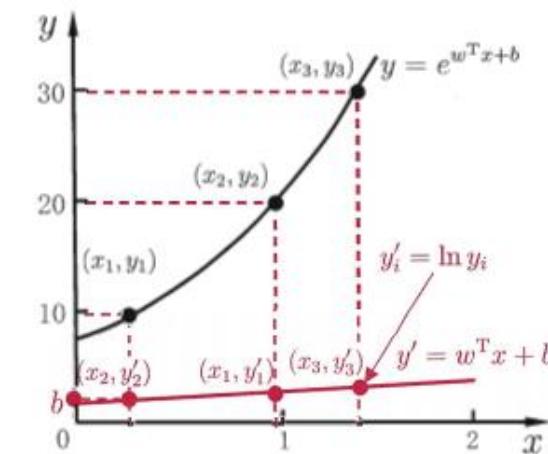


图 3.1 对数线性回归示意图

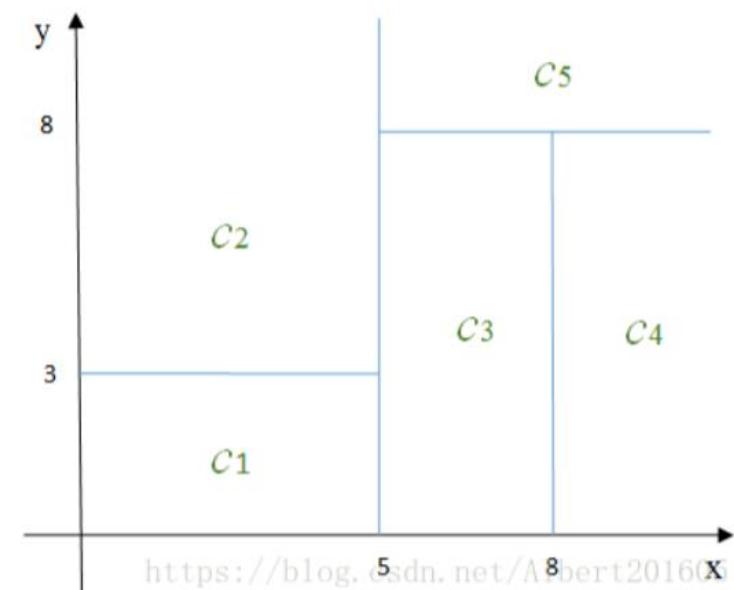
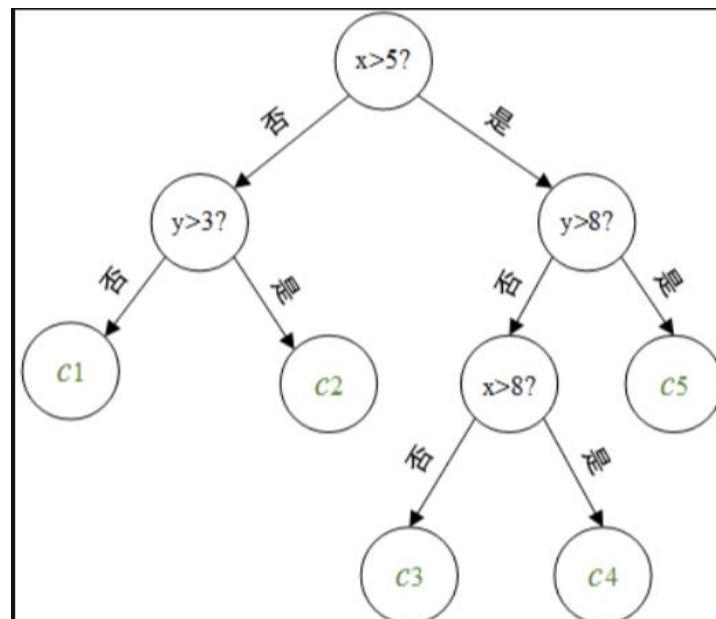
口 k近邻回归

- 回归的主要思想：根据邻居的标签值的平均值，作为测试样本的预测标签值
- 优点：k近邻回归模型对数据分布没有假设，适用于各种类型的数据
- 缺点：k近邻回归模型需要计算测试样本与训练样本之间的距离，计算量随着训练样本数量的增加而增加

2.3 决策树回归

口 决策树回归

- 回归的主要思想：回归树就是将特征空间划分成若干单元，每一个划分单元有一个特定的输出
- 针对目标变量是连续的



2.4 支持向量回归

口 支持向量回归

- 回归的主要思想：寻找具有最大边界的回归面
- 问题形式：

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \ell_i(f(x_i) - y_i)$$

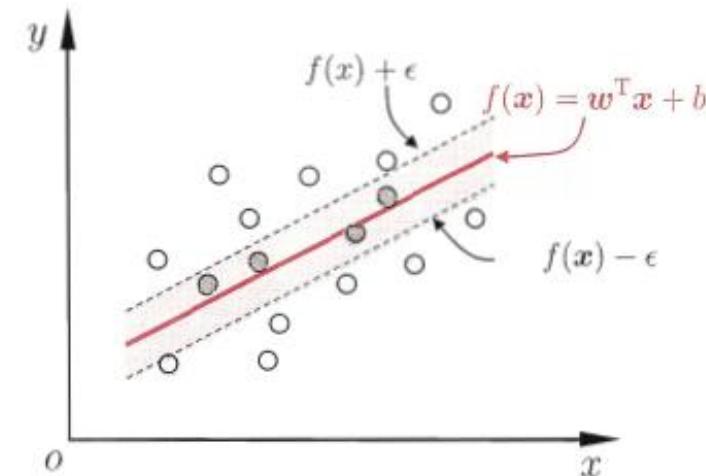


图 6.6 支持向量回归示意图。红色显示出 ϵ -间隔带，落入其中的样本不计算损失。

Part

03

小结

3.1 小结

- 分类和回归算法：有监督方法
- 分类方法可以扩展到回归：决策树回归、支持向量回归

thanks