



## 6.1 机器学习中的回归问题（下）

HUAWEI TECHNOLOGIES CO., LTD.

[www.huawei.com](http://www.huawei.com)



# 目录

Contents

1

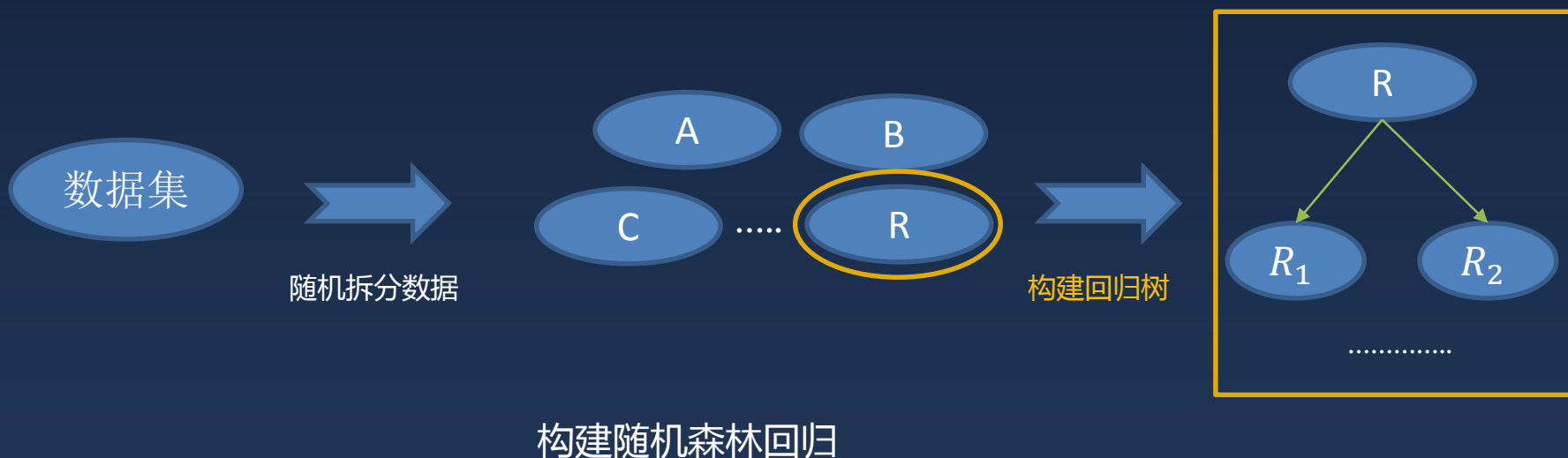
随机森林回归

2

梯度提升树

# 回归树

前面的课程我们提到了决策树和随机森林分类算法，今天我们讨论的是随机森林回归算法。随机森林的相关基本概念可查看之前的课程（机器学习中的分类问题（上））。随机森林用于分类还是回归取决于每颗决策树是分类树还是回归树。当为回归树时，树结点采用的分裂原则是**最小均方差**。即对于任意划分特征A，对应的任意划分点s两边划分成的数据集 $R_1$ 和 $R_2$ ，求出使 $R_1$ 和 $R_2$ 各自集合的均方差最小，同时 $R_1$ 和 $R_2$ 的均方差之和最小所对应的特征和特征划分点。



回归树构建过程：

- 1.考虑数据集  $R$  上的所有特征  $j$ ，遍历每一个特征下所有可能的取值或者切分点  $s$ ，将数据集  $R$  划分成两部分  $R_1$  和  $R_2$
- 2.分别计算上述两个子集的平方误差和，选择最小的平方误差对应的特征与分割点，生成两个子节点。
- 3.对上述两个子节点递归调用步骤1和2,直到满足停止条件。

# 回归树构建的过程

构建二叉决策树：

1.选择最优切分变量j与切分点s，求解：

$$\min_{j,s} \left[ \min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right]$$

遍历变量j，对固定的切分变量j扫描切

分点s，选择使上式达到最小值得对(j, s)。

2.用选定的对(j, s)划分区域并决定相应的输出值：

$$R_1(j,s) = \{x | x^{(j)} \leq s\}, \quad R_2(j,s) = \{x | x^{(j)} > s\}$$
$$\hat{c}_m = \frac{1}{N_m} \sum_{x_i \in R_m(j,s)} y_i, \quad x \in R_m, \quad m=1,2$$

3.继续对两个子区域调用步骤1和2，直至满足停止条件。

4.将输入空间划分为M个区域 $R_1, R_2, \dots, R_M$ ，生成决策树：

$$f(x) = \sum_{m=1}^M \hat{c}_m I(x \in R_m)$$

# 目录

Contents

1

随机森林回归

2

梯度提升树

# 梯度提升

梯度提升树回归是一种基于决策树的迭代回归算法。该算法采用**迭代**的思想不断地构建**回归决策树**模型，每棵树都是通过梯度优化损失函数而构建，从而达到从基准值到目标值的逼近。首先，我们以之前学习的线性回归为例来介绍**梯度提升**的概念，如下图所示，我们每一次的迭代建模都是：后一次模型都是针对前一次模型预测出错的情况进行修正，模型随着迭代不断地改进，从而获得比较好的预测效果。

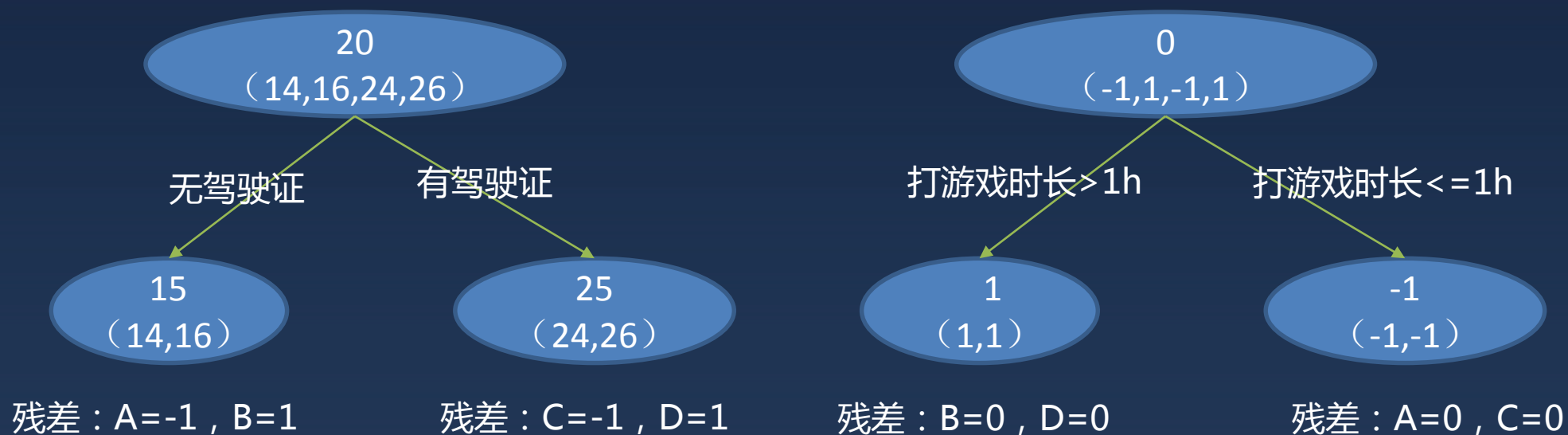


梯度提升步骤：

- 1.我们首先用简单的模型对数据进行建模并分析数据中的错误。
- 2.这些错误表示难以用简单模型拟合的数据点。
- 3.然后对于以后的模型，我们特别关注那些难以拟合的数据点，以使他们正确。
- 4.最后，我们通过给每个预测变量赋予一些权重来组合所有预测变量。

# 梯度提升树回归问题

这里我们用一个事例来介绍**梯度提升树回归**算法。假设我们的数据集有4个人：A、B、C、D。，年龄分别为（14，16，24，26）其中A、B是学生；C，D是上班族；A、C是女生；B，D是男生。由于数据少，我们限定叶子节点最多有两个，即每棵树只有一个分枝，并限定只学两棵树，我们会得到如下结果：



注：残差的意思是：残差 = 实际值 - 预测值

# 梯度提升树回归问题

上面的例子可以翻译为以下句子：

1.14岁女学生，没有驾驶证打游戏少；预测年龄为： $15-1=14$

2.16岁男学生，没有驾驶证打游戏多；预测年龄为： $15+1=16$

3.24岁女员工，有驾驶证打游戏少；预测年龄为： $25-1=24$

4.26岁男员工，有驾驶证打游戏多；预测年龄为： $25+1=26$

上面的过程：通过有无驾驶证把4人分为了AB、CD两组，每组用平均年龄作为预测值。此时A的残值为 $15-14=1$ 。以此类推得到ABCD的残差分别为-1,1-1,1。然后我们那残差代替ABCD的原值，进行第二颗树的训练。第二颗树，我们通过上网时长，将四个人分为两类，残差都为0，即每个人都得到了最真实的预测值。





# Thank You.

**Copyright©2016 Huawei Technologies Co., Ltd. All Rights Reserved.**

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

华为云机器学习服务MLS  
[www.huaweicloud.com/product/mls.html](http://www.huaweicloud.com/product/mls.html)