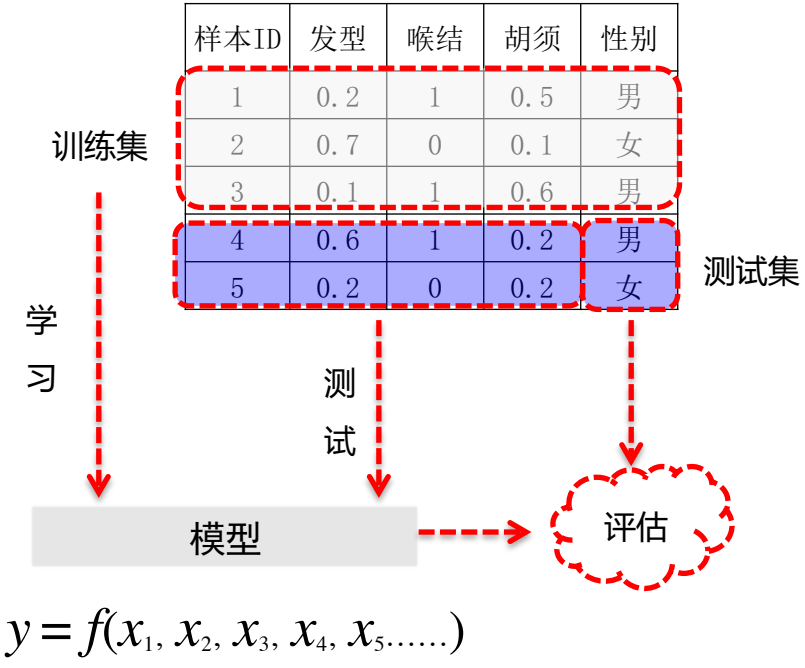
模型评估与选择

目录

基本术语



$$y = f(x_1, x_2, x_3, x_4, x_5 \ldots\ldots)$$

目录

评估方法

- 留出法
- 交叉验证
- 自助法

训练集

样本ID	发型	喉结	胡须	性别
1	0.2	1	0.5	男
2	0.7	0	0.1	女
3	0.1	1	0.6	男
4	0.6	1	0.2	男
5	0.2	0	0.2	女

测试集

评估方法

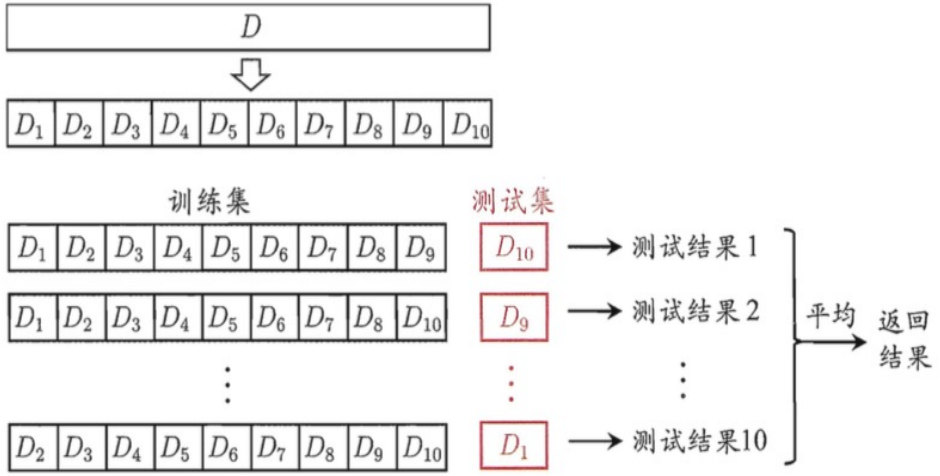
留出法

- 单次留出与多次留出
- 多次留出法：如对专家样本随机进行100次训练集 / 测试集划分，评估结果取平均

评估方法

交叉验证法

- K折交叉验证：将专家样本等份划分为K个数据集，轮流用K - 1个用于训练，1个用于测试
- P次K折交叉验证



评估方法

自助法

留出法与交叉验证法的训练集数据少于样本数据

- 给定m个样本的数据集D，从D中有放回随机取m次数据，形成训练集D’
- 用D中不包含D’ 的样本作为测试集
- D中某个样本不被抽到的概率：$\left(1-\frac{1}{m}\right)^m$
- 测试集数据量：$\lim_{m\mapsto\infty}\left(1-\frac{1}{m}\right)^m \mapsto \frac{1}{e}\approx 0.368$
- 缺点：改变了初始数据集的分布

目录

回归性能度量

评价方法与评价标准

平均绝对误差、均方误差和中值绝对误差的值越靠近0，模型性能越好。可解释方差值和R方值则越靠近1，模型性能越好。

方法名称	最优值
平均绝对误差	0.0
均方误差	0.0
中值绝对误差	0.0
可解释方差值	1.0
R方值	1.0

$$E(f;D)=\frac{1}{m}\sum_{i=1}^{m}\left(f\left(\boldsymbol{x}_i\right)-y_i\right)^2$$

分类性能度量

错误率与精度

- 错误率：分类错误样本数占总样本数比例

$$E(f;D)=\frac{1}{m}\sum_{i=1}^{m}\mathbb{I}(f(\boldsymbol{x}_i)\neq y_i)$$

- 精度：1 - 错误率，分类正确样本数占总样本数比例

$$\begin{aligned}\mathrm{acc}(f;D) &= \frac{1}{m}\sum_{i=1}^{m}\mathbb{I}(f(\boldsymbol{x}_i)=y_i)\\ &= 1-E(f;D)\ .\end{aligned}$$

样本ID	real	pre
1	男	男
2	女	女
3	男	男
4	男	女
5	女	女

分类性能度量

混淆矩阵

混淆矩阵就是汇总分类模型中分类正确和不正确的样本数目的矩阵。对于简单的二分类问题的混淆矩阵如下表：

		分类结果	
		正	负
实际结果	正	TP	FN
	负	FP	TN

		预测值	
		1	0
实际值	1	TP	FN
	0	FP	TN

分类性能度量

查准率 / 准确率

表示的是**分类为正类**的样本中**实际为正类**的样本所占的比例，精确率越高，模型效果越好。

		分类结果	
		正	负
实际结果	正	TP	FN
	负	FP	TN

查准率 / 准确率（precision）： P = TP/(TP+FP)

分类性能度量

查全率 / 召回率 / 灵敏度

表示**实际为正类**的样本中被预测正确的比例，召回率越高，表示模型将正类误概率越低，模型效果越好。

		分类结果	
		正	负
实际结果	正	TP	FN
	负	FP	TN

查全率 / 召回率 / 灵敏度（recall）： R = TP/(TP+FN)

分类性能度量

F1值

F-Measure（又称为F-Score）**综合考虑精确度与召回率**，其中P指精确率，R指召回率。F-Measure是精确度和召回率的**加权调和平均**：

$$F = \frac{(\alpha^2 + 1) * P * R}{\alpha^2 * (P + R)}$$

当参数α=1时，就是最常见的F1值，即：

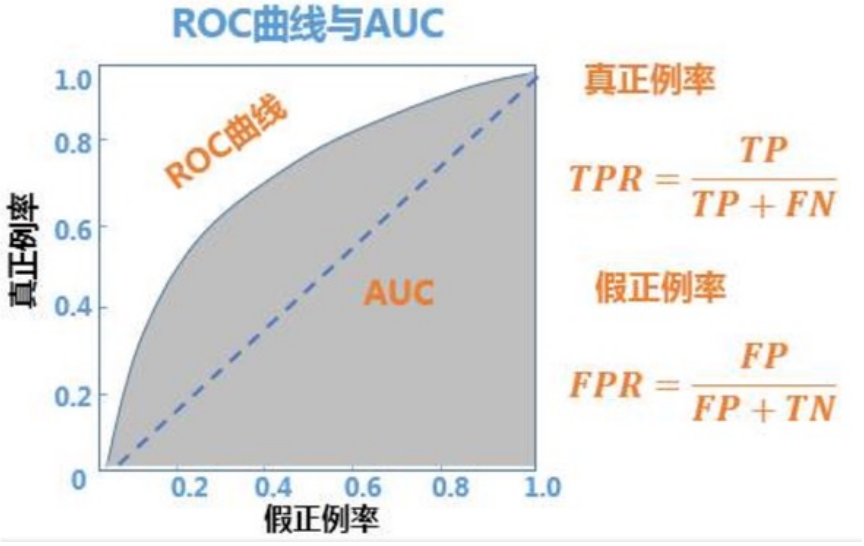
$$F1 = \frac{2 * P * R}{P + R}$$

P = TP/(TP+FP)

R = TP/(TP+FN)

分类性能度量

- **ROC曲线：**接收者操作特征曲线（Receiver Operating Characteristic curve）是一种非常有效的模型评价方法，可为选定临界值给出定量提示。
- ROC曲线图以真正率为纵坐标，以假正率为横坐标。AUC值是ROC曲线下方的面积，面积的大小与模型的优劣密切相关，可反映分类器正确分类的统计概率。因此，AUC 的值越大，说明分类模型的性能越好 。

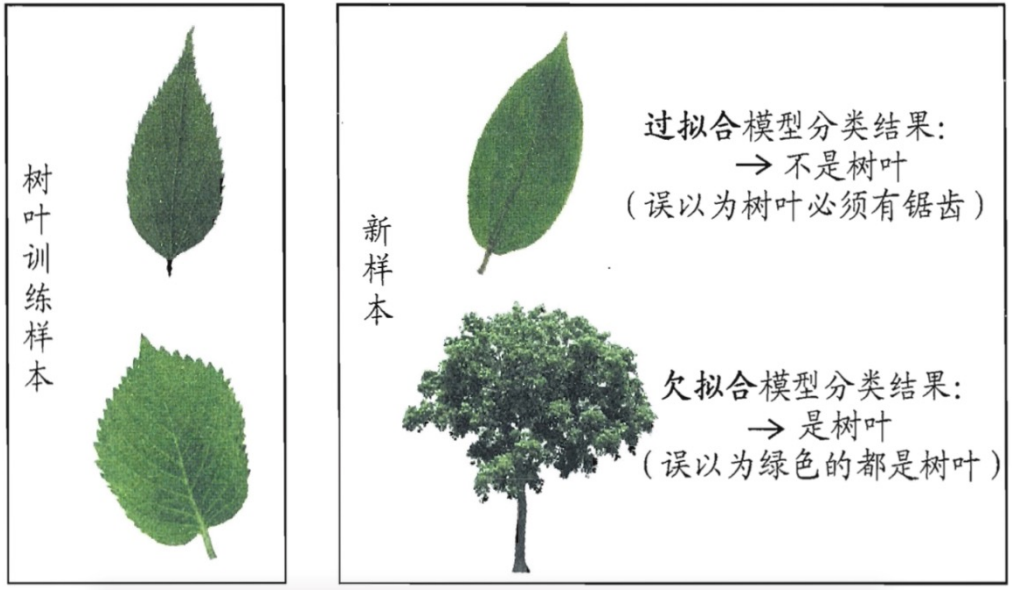


目录

经验误差与过拟合

“过”与“不及”

- 过拟合：用力过猛
- 欠拟合：用力不足



经验误差与过拟合

缓解过拟合问题：

- 要想解决过拟合问题，就要显著减少测试误差而不过度增加训练误差，从而提高模型的泛化能力。我们可以使用正则化（Regularization）方法。**即通过修改学习算法，使其降低泛化误差而非训练误差。**
- 让机器学习或深度学习模型泛化能力更好的办法就是使用更多的数据进行训练。当我们拥有的数据量是有限的。解决这个问题的一种方法就是**创建“假数据”并添加到训练集中——数据集增强。**通过增加训练集的额外副本来增加训练集的大小，进而改进模型的泛化能力。
- **降低特征的数量**

  对于一些特征工程而言，可以降低特征的数量——删除冗余特征。
- **Dropout层**

  Dropout是在训练网络时用的一种技巧，相当于在隐藏单元增加了噪声。
- **Early stopping（提前终止）**

经验误差与过拟合

缓解欠拟合问题：

- 增加或者构造属性，**使输入数据具有更强的表达能力**
- **模型复杂化**

对同一个算法复杂化。例如回归模型添加更多的高次项，增加决策树的深度，增加神经网络的隐藏层数和隐藏单元数等

- **降低正则化约束**

Thank you!