

## 第2章 模型评估与选择

---

### 2.1 经验误差与过拟合

---

学习器在训练集上的误差称为“训练误差”或“经验误差”，在新样本上的误差称为“泛化误差”。

过拟合：把训练样本自身的某些特点误以为是所有潜在样本都会具有的一般性质。例如：识别树叶时，把叶子的锯齿误以为是所有树叶都会具有的性质。过拟合会导致泛化性能降低。

欠拟合：指对训练样本的一般性质尚未学好。比如看见绿色就以为是树叶。

### 2.2 评估方法

---

#### 2.2.1 留出法

最常用的方法，也就是平时我们所说的 80% 的训练样本，20% 的测试样本。

#### 2.2.2 交叉验证法

k折交叉验证：将数据集D 划分为 k 个大小相似的互斥子集。然后每次用 k-1 个子集做训练集，剩下一个做测试集

#### 2.2.3 调参

把训练数据另外划分为训练集和验证集，基于验证集上的性能来进行模型选择和调参

### 2.3 性能度量

---

回归任务最常用的性能度量是 均方误差：

$$E = \frac{1}{m} \sum_{i=1}^m (f(x) - y_i)^2$$

#### 2.3.1 错误率和精度

错误率是分类错误的样本数占样本总数的比例。

如果该问题是二分类问题，查准率（precision）=（样本为正，预测为正）/（所有预测为正的）

查全率（recall）=（样本为正，预测为正）/（所有样本为正）

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

F1 度量的一般形式—— $F_\beta$  能让我们表达出对两种率的不同偏好，它定义为

$$F_\beta = \frac{(1 + \beta^2) \times P \times R}{(\beta^2 \times P) + R}$$

其中 $\beta > 0$  度量了查全率对查准率的相对重要性。 $\beta = 1$  是退化为  $F1$  ;  $\beta > 1$  时查全率有更大影响;  $\beta < 1$  时查准率有更大影响。