

期望风险、经验风险与结构风险

【参考资料】

李航 《统计学习方法》

1. 损失函数与风险函数

- **损失函数**：度量模型一次预测的好坏
- **风险函数**：度量平均意义下模型预测的好坏（表现为损失函数的期望或平均）

2. 期望风险

英文为expected risk，表示对所有样本（包含未知样本和已知的训练样本）的预测能力，是全局概念。

$$R_{\text{exp}}(f) = E_P[L(Y, f(X))] = \int_{\mathcal{X} \times \mathcal{Y}} L(y, f(x)) P(x, y) dx dy$$

因为是理论上模型 $f(X)$ 关于联合分布 $P(X, Y)$ 的平均意义下的损失，所以又被称为风险函数。

学习的目标就是选择期望风险最小的模型。由于联合概率分布 $P(X, Y)$ 是未知的， $R_{\text{exp}}(f)$ 不能直接计算。

3. 经验风险

英文名为empirical risk。是模型关于训练样本集的平均损失：

$$R_{\text{emp}}(f) = \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i))$$

其中 N 为样本数。

经验风险代表模型 $f(x)$ 对训练集中所有样本的预测能力。经验风险最小化（empirical risk minimization, ERM）是监督学习的两个基本策略之一：

$$\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i))$$

理想的模型（决策）函数应该是让所有的样本的损失函数最小（即期望风险最小化）。但是期望风险函数往往不可得，所以用局部最优代替全局最优。这就是经验风险最小化的理论基础。

当样本容量足够大时，经验风险最小化能保证有很好的学习效果，在现实中被广泛使用。比如，极大似然估计就是经验风险最小化的一个例子。

但是，当样本容量很小时，经验风险最小化学习的效果就未必很好，容易产生过拟合现象。

4. 结构风险

英文名为structural risk。结构风险在经验风险函数后面加一个表示模型复杂度的正则化项：

$$R_{\text{strm}}(f) = \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i)) + \lambda J(f)$$

结构风险在经验风险和模型复杂度之间进行了权衡。结构风险最小化 (structural risk minimization, SRM) 是监督学习的另一种策略, 等价于正则化, 可以有效防止过拟合:

$$\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i)) + \lambda J(f)$$

最大后验概率估计就是结构风险最小化的一个例子。