# 模型评估

# 1. 模型评估常用方法

## 分类模型常用评估方法:

指标	描述	
Accuracy	准确率	
Precision	精准度/查准率	
Recall	召回率/查全率	
P-R曲线	查准率为纵轴,查全率为横轴,作图	
F1	F1值	
Confusion Matrix	混淆矩阵	
ROC	ROC曲线	
AUC	ROC曲线下的面积	

## 回归模型常用评估方法:

指标	描述
Mean Square Error (MSE, RMSE)	平均方差
Absolute Error (MAE, RAE)	绝对误差
R-Squared	R平方值

# 2.误差 (Bias) 、偏差 (Error) 与方差 (Variance)

### **Error**:

误差 (error): 一般地, 我们把学习器的实际预测输出与样本的真是输出之间的差异称为"误差"

Error = Bias + Variance + Noise, Error反映的是整个模型的准确度

### Noise:

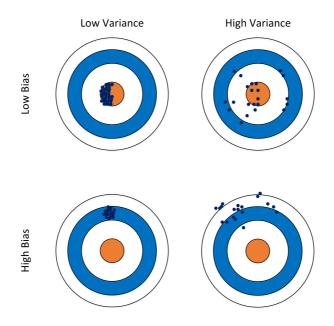
噪声 (noise) : 描述了在当前任务上任何学习算法所能达到的期望泛化误差的下界,即刻画了学习问题本身的难度。

#### Bias:

偏差(Bias): 衡量模型拟合训练数据的能力(训练数据不一定是整个 training dataset, 而是只用于训练它的那一部分数据, 例如: mini-batch), Bias反映的是模型在样本上的输出与真实值之间的误差,即模型本身的精准度。

Bias 越小,拟合能力越高(可能产生overfitting);反之,拟合能力越低(可能产生underfitting)。

偏差越大,越偏离真实数据。



#### Variance:

方差(variance): $S_N^2 = rac{1}{N} \sum_{i=1}^N (x_i - ar{x})^2$ 

描述的是预测值的变化范围,离散程度,也就是离其期望值的距离,反映的是模型每一次输出结果与模型输出期望之间的误差,即模型的稳定性。

Variance越小,模型的泛化的能力越高;反之,模型的泛化的能力越低。

# 3. K折交叉验证

- 1. 将含有N个样本的数据集,分成K份,每份含有N/K个样本。选择其中1份作为测试集,另外K-1份作为训练集,测试集就有K种情况。
- 2. 在每种情况中,用训练集训练模型,用测试集测试模型,计算模型的泛化误差。
- 3. 交叉验证重复K次,每份验证一次,平均K次的结果或者使用其它结合方式,最终得到一个单一估测,得到模型最终的泛化误差。
- 4. 将K种情况下,模型的泛化误差取均值,得到模型最终的泛化误差。

- 5. 一般 $2 \le K \le 10$ 。 k折交叉验证的优势在于,同时重复运用随机产生的子样本进行训练和验证,每次的结果验证一次,10折交叉验证是最常用的。
- 6. 训练集中样本数量要足够多,一般至少大于总样本数的50%。
- 7. 训练集和测试集必须从完整的数据集中均匀取样。均匀取样的目的是希望减少训练集、测试集与原数据集之间的偏差。当样本数量足够多时,通过随机取样,便可以实现均匀取样的效果

## 4. 混淆矩阵

## 第一种混淆矩阵:

真实情况T or F	预测为正例1, P	预测为负例0, N
本来label标记为1,预测结果真为	TP(预测为1,实	FN(预测为0,实际
T、假为F	际为1)	为1)
本来label标记为0,预测结果真为	FP(预测为1,实	TN(预测为0,实际
T、假为F	际为0)	也为0)

## 第二种混淆矩阵:

预测情况P or N	实际label为1,预测对了为T	实际label为0,预测对了为T
预测为正例1, P	TP(预测为1,实际为1)	FP(预测为1,实际为0)
预测为负例0, N	FN(预测为0,实际为1)	TN(预测为0,实际也为0)