

第四部分 机器学习技巧和经验 / Machine Learning Tips and Tricks

翻译&校正 | 韩信子@ShowMeAI

编辑 | 南乔@ShowMeAI

原文作者 | <https://stanford.edu/~shervine>

本节原文超链

[1]度量 / Metrics

1.1 分类评价指标 Classification metrics

在二分类问题中，下面这些主要度量标准对于评估模型的性能非常重要。

混淆矩阵 Confusion matrix

混淆矩阵可以用来评估模型的整体性能情况。定义如下：

		预测类别	
		+	-
实际类别	+	TP, True Positives 真阳性	FN, False Negatives Type II error 假阴性
	-	FP, False Positives Type I error 假阳性	TN, True Negatives 真阴性

主要度量标准 Main metrics

下表总结了评估分类模型的常用度量标准：

性能度量	公式	说明
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	模型总体性能
Precision	$\frac{TP}{TP+FP}$	预测为正样本的准确度
Recall Sensitivity	$\frac{TP}{TP+FN}$	真正样本的覆盖度
Specificity	$\frac{TN}{TN+FP}$	真负样本的覆盖度
F1 score	$\frac{2TP}{2TP+FP+FN}$	混合度量，对于不平衡类别非常有效

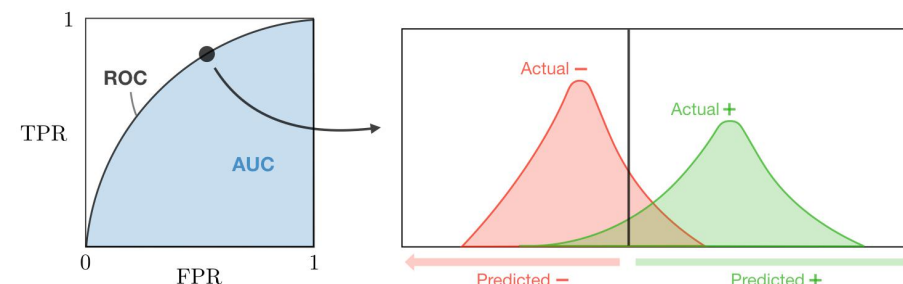
受试者工作曲线 ROC

受试者工作曲线（receiver operating curve，ROC），以真阳率（True Positive Rate, TPR）为纵轴，以假阳率（False Positive Rate, FPR）横轴，并经过调整阈值绘制而成。下表汇总了坐标轴的度量标准：

性能度量	公式	等价形式
True Positive Rate, TPR	$\frac{TP}{TP+FN}$	Recall, sensitivity
False Positive Rate, FPR	$\frac{FP}{TN+FP}$	1-specificity

AUC

受试者工作曲线的之下的部分（The area under the receiving operating curve，AUC / AUROC），即 ROC 曲线下的部分。如下图所示：



1.2 回归评价指标 Regression metrics

基本性能度量 Basic metrics

给定一个回归模型 f ，下面的度量标准通常用来评估模型的性能：

全部平方和	解释平方和	残差平方和
$SS_{\text{tot}} = \sum_{i=1}^m (y_i - \bar{y})^2$	$SS_{\text{reg}} = \sum_{i=1}^m (f(x_i) - \bar{y})^2$	$SS_{\text{res}} = \sum_{i=1}^m (y_i - f(x_i))^2$

确定性系数 Coefficient of determination

确定性系数（记作 R^2 或 r^2 ），用以衡量模型复现观测结果的能力，定义如下：

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

主要度量标准 Main metrics

以下度量标准常用于评估回归模型的性能（考虑到了变量的数量 n ）：

Mallow's Cp	AIC	BIC	Adjusted R^2
$\frac{SS_{res} + 2(n+1)\hat{\sigma}^2}{m}$	$2[(n+2) - \log(L)]$	$\log(m)(n+2) - 2\log(L)$	$1 - \frac{(1-R^2)(m-1)}{m-n-1}$

备注：L 代表似然， $\hat{\sigma}^2$ 代表方差。

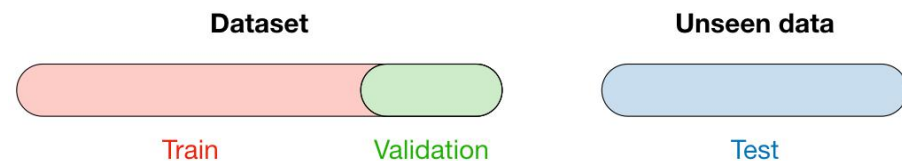
[2]模型选择 / Model Selection

词汇 Vocabulary

在选择模型时，将数据分为以下 3 个不同的部分：

训练集 Training set	验证集 Validation set	测试集 Testing set
用于模型训练 通常为数据集的 80%	用于模型评估 通常为数据集的 20% 又叫做留出集或者开发集	用于模型预测 使用未知数据

选定模型，在整个数据集上进行训练，并在测试集(未见过的数据)上进行测试。如图：



交叉验证 Cross-validation

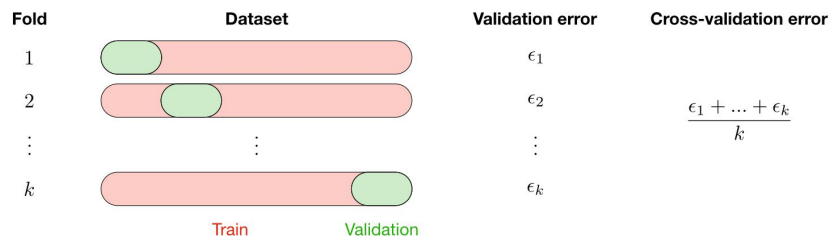
交叉验证（Cross-validation），是一种不必特别依赖于初始训练集的模型选择方法。

下表汇总了几种不同的方式：

k-fold	Leave-p-out
在 $k-1$ 个子集上训练 在剩余的一个子集中评估 通常 $k=5$ 或 10	在 $n-p$ 个子集上训练 在剩余的 p 个子集评估模型 $p=1$ 时又叫做留一法

最常用的模型选择方法是 k 折交叉验证，将训练集划分为 k 个子集，在 $k-1$ 个子集上训练模型，在剩余的一个子集上评估模型，用这种划分方式重复训练 k 次。

交叉验证损失是 k 次 k 折交叉验证的损失均值。



正则化 Regularization

正则化方法可以解决高方差问题，避免模型对于训练数据产生过拟合。

下表汇总了常用的正则化方法：

LASSO	Ridge	Elastic Net
将系数收缩为 0 有利于变量选择	使系数更小	在变量选择和小系数之间进行权衡
$\dots + \lambda \theta _1$ $\lambda \in \mathbb{R}$	$\dots + \lambda \theta _2^2$ $\lambda \in \mathbb{R}$	$\dots + \lambda [(1-\alpha) \theta _1 + \alpha \theta _2^2]$ $\lambda \in \mathbb{R}, \alpha \in [0,1]$

[3]诊断 / Diagnostics

偏差 Bias

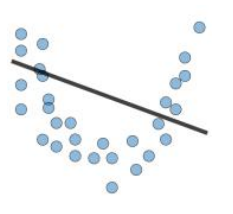

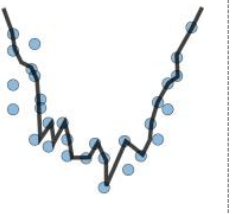
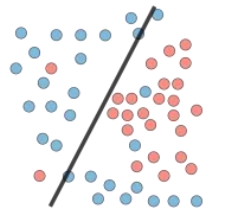
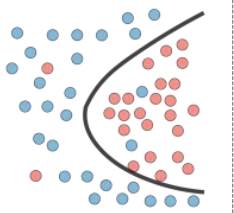
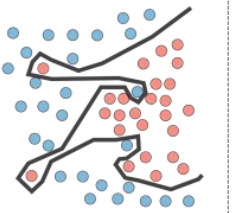

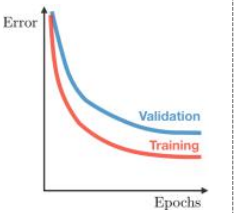
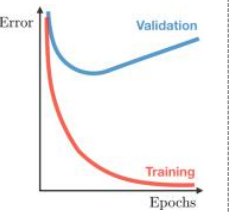
模型的偏差，表示模型预测值和真实值之间的差距。

方差 Variance

模型的方差，表示给定数据点的模型预测结果的波动程度。

■ 偏差/方差权衡 Bias/variance tradeoff

模型越简单，偏差越高；模型越复杂，方差越高。

	欠拟合 Underfitting	模型状态正常 Just right	过拟合 Overfitting
症状	高训练误差 训练误差接近于测试误差 高偏差	训练误差略低于测试误差	极低训练误差 训练误差远低于测试误差 高方差
回归图			
分类图			
深度学习插图			
可能的补救措施	模型复杂性 添加更多特征 训练更长时间	//	实施正则化 获得更多数据

■ 错误分析 Error analysis

分析当前模型和完美模型之间性能差异的根本原因。

■ 烧蚀分析 Ablative analysis

可以分析当前和基线模型（**baseline model**）之间性能差异的根本原因。

Awesome AI Courses Notes Cheat Sheets

Machine Learning CS229	Deep Learning CS230	Natural Language Processing CS224n	Computer Vision CS231n	Deep Reinforcement Learning CS285	Neural Networks for NLP CS11-747	DL for Self-Driving Cars 6.S094	...
Stanford	Stanford	Stanford	Stanford	UC Berkeley	CMU	MIT	...

是 **ShowMeAI** 资料库的分支系列，覆盖最具知名度的 TOP20+ 门 AI 课程，旨在为读者和学习者提供一整套高品质中文速查表，可以点击 [【这里】](#) 查看。

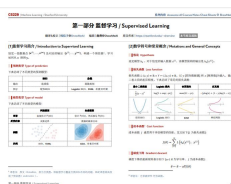
斯坦福大学（Stanford University）的 **Machine Learning（CS229）** 和 **Deep Learning（CS230）** 课程，是本系列的第一批产出。

本批两门课程的速查表由斯坦福大学计算机专业学生 **Shervine Amidi** 总结整理。原速查表为英文，可点击 [【这里】](#) 查看，**ShowMeAI** 对内容进行了翻译、校对与编辑排版，整理为当前的中文版本。

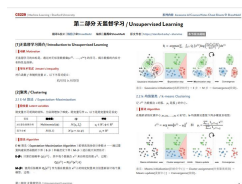
有任何建议和反馈，也欢迎通过下方渠道和我们联络 (*^_^*)

CS229 | Machine Learning @ Stanford University

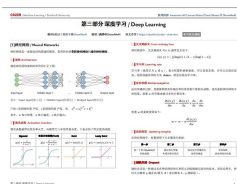
监督学习 Supervised Learning


[中文速查表链接](#)

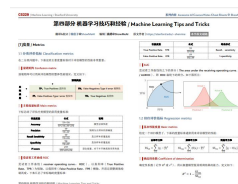
无监督学习 Unsupervised Learning


[中文速查表链接](#)

深度学习 Deep Learning

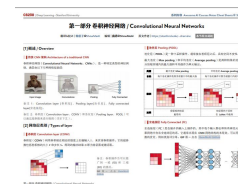

[中文速查表链接](#)

机器学习技巧和经验 Tips and Tricks

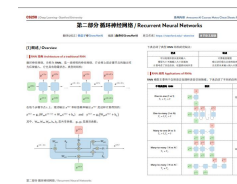

[中文速查表链接](#)

CS230 | Deep Learning @ Stanford University

卷积神经网络 CNN


[中文速查表链接](#)

循环神经网络 RNN


[中文速查表链接](#)

深度学习技巧与建议 Tips and Tricks


[中文速查表链接](#)

概率统计 Probabilities / Statistics


[中文速查表链接](#)

线性代数与微积分 Linear Algebra and Calculus


[中文速查表链接](#)

GitHub
ShowMeAI

<https://github.com/ShowMeAI-Hub/>



ShowMeAI 研究中心

扫码回复“**速查表**”
下载**最新**全套资料