第一章

1.2 泛化：在训练集上训练模型，模型应用于新样本的能力

1.4

①归纳偏好：在学习过程中对某种类型假设的偏好。（什么模型更好？/假设）

可看作学习算法在庞大的假设空间对假设进行选择的启发式或"价值观"

"奥卡姆剃刀"

归纳偏好是否与问题匹配

②没有免费午餐

前提：所有"问题"出现机会相同、或所有问题同等重要。但现

有时只关注于在解决问题，

考虑具体问题

第2章 模型评估与选择

2.1 经验误差 过拟合

经验误差：在训练集上

泛化误差：在新样本上

2.2 评估方法：

①留出法：分层采样

多次使用留出法，若干次随机划分得到不同的训练/测试集，最后平均。

1)测试集小：结果方差较大

2)训练集小：结果偏差较大

②K折交叉验证

10次10折交叉验证=100次留出

③自助法

可重复采样，有放回采样

m个样本数据集D

每次随机选一个，再放回

重复m次，得到包含m个样本的数据集

适用：数据集小，难以有效划分 利于集成学习

改变了分布 估计偏差

2.3 性能度量：

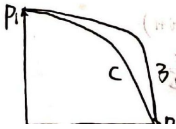
①均方误差 $E=\frac{1}{m}\sum_{i=1}^{m}(f(x_i)-y_i)^2$

②准确率 召回率

真实 \ 预测	正	反
正	TP 真正	FN 假反
反	FP 假正	TN 真反

$P=\frac{TP}{TP+FP}$ $R=\frac{TP}{TP+FN}$

③PR曲线：



若一个完全包住另一个，则前者优于后者

2.4 比较检验 ☆

1) 假设检验

"假设": 对学习器泛化错误率分布;

某种判断猜想

测试错误率 $\hat{\epsilon}$ $\xrightarrow{\text{推出}}$ 泛化错误率分布 ϵ

泛化错误率 ϵ 被测得测试错误率 $\hat{\epsilon}$

$$P(\hat{\epsilon};\epsilon)=\binom{m}{\hat{\epsilon}\times m}\epsilon^{\hat{\epsilon}\times m}(1-\epsilon)^{m-\hat{\epsilon}\times m}$$

$\frac{\partial P(\hat{\epsilon};\epsilon)}{\partial \epsilon}=0 \Rightarrow \epsilon=\hat{\epsilon}$ 时最大

2.5 偏差与方差

$\bar{f}(x)=E_D[f(x;D)]$ 期望预测

$var(x)=E_D[(f(x;D)-\bar{f}(x))^2]$ 不同测得率方差

$\varepsilon^2=E_D[(y_D-y)^2]$ 噪声

$bias^2(x)=(\bar{f}(x)-y)^2$ 偏差

期望泛化误差:

$$E(f;D)=E_D[(f(x;D)-\bar{f}(x))^2]+(\bar{f}(x)-y)^2+E_D[(y_D-y)^2]$$

$$=bias^2(x)+var(x)+\varepsilon^2$$

偏差 方差 噪声

偏差:
- 同样大小训练集的变动导致的学习性能变化
- 数据扰动带来的影响
- 过拟合
- 数据充分性

方差:
- 期望预测与真实结果偏离程度
- 刻画拟合能力
- 欠拟合
- 算法本身能力

噪声:
- 期望泛化误差的下界
- 问题本身的难度

$F1 = \frac{2 \times P \times R}{P+R}$ 调和平均 更重视较小值

相比于算术平均和几何均.

对P, R重视程度不同.

1) 商品推荐: 尽可能少打扰用户, 确保用户感兴趣的 P更重要

2) 逃犯信息检索: 不能漏掉, R更重要

$F_\beta = \frac{(1+\beta^2) \times P \times R}{(\beta^2 \times P) + R}$ β度量了R对P的相对重要性

$\beta > 1$: R更重要

$\beta < 1$: P更重要

③ROC, AUC

将预测值和分类阈值比较

根据这个值对测试样本排序.

以某个截断点分为两部分, 前为正例后为负例.

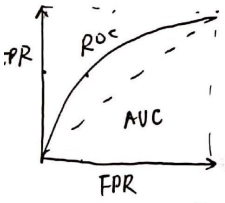
根据任务需求来设置截断点

{ 重视P: 靠前截断
重视R: 靠后截断

1) ROC 受试者工作特征

纵轴: 真正例率 $TPR = \frac{TP}{TP+FN}$

横轴: 假正例率 $FPR = \frac{FP}{TN+FP}$



根据预测结果排序, 按顺序逐个把样本作为正例进行预测

(0,1) 对应于所有正例在反例前面的理想模型.

真正例 $(x, y+\frac{1}{m^+})$

假正例 $(x+\frac{1}{m^-}, y)$

$AUC = \frac{1}{2} \sum_{i=1}^{m-1} (x_{i+1} - x_i) \cdot (y_i + y_{i+1})$

样本预测的排序质量, 排序误差

the steps you take don't need to be big. They just need to take you in the right direction

$l_{rank} = \frac{1}{m^+ m^-} \sum_{x^+ \in D^+} \sum_{x^- \in D^-} \left(\mathbb{I}(f(x^+) < f(x^-)) + \frac{1}{2} \mathbb{I}(f(x^+) = f(x^-)) \right)$

④代价敏感错误率, 代价曲线

患者 → 健康人

健康人 → 患者

权衡不同类型错误的代价. 非均等代价

二分类任务为例:

设定代价矩阵: $cost_{ij}$ 表示将第i类样本预测为第j类的代价 $cost_{ii} = 0$

真实 \ 预测	0	1
0	0	$cost_{01}$
1	$cost_{10}$	0

非均等代价下, "代价敏感"错误率

$$E(f; D; cost) = \frac{1}{m} \left(\sum_{x_i \in D^+} \mathbb{I}(f(x_i) \neq y_i) \times cost_{01} + \sum_{x_i \in D^-} \mathbb{I}(f(x_i) \neq y_i) \times cost_{10} \right)$$

非均等代价下: ROC曲线不行.

代价曲线:

横轴: 正例概率代价

$$P(+)cost = \frac{p \times cost_{01}}{p \times cost_{01} + (1-p) \times cost_{10}}$$

纵轴: [0,1] 归一化代价

$$cost_{norm} = \frac{FNR \times p \times cost_{01} + FPR \times (1-p) \times cost_{10}}{p \times cost_{01} + (1-p) \times cost_{10}}$$

p: 样例为正例的概率?