§2. 模型评估与选择

1. 经验误差与过拟合

error rate $E = \frac{a}{m}$　　m个样本a个样本分类错误

accuracy $= 1 - \frac{a}{m}$

overfitting 过拟合：学习器把训练样本学得太好，泛化能力下降

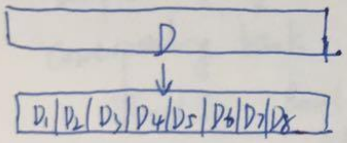
underfitting 欠拟合：训练样本一般性质尚未学好

2. 评估方法

1) hold-out 留出法　1000 < S 700个　注意训练/测试集的划分要尽可能保持数据分布的一致性
T 300个

方法：fidelity：将2/3～4/5样本用于训练，其余测试

2) 交叉验证法 cross validation　　Leave-one-out 留一法(特例)



训练集	测试集	
D1 D2 D3 D4 D5 D6 D7	D8	平均 → 评估
D1 D2 D3 D4 D5 D6 D8	D8	
⋮		

3) bootstrapping 自助法：每次取一个样本，将其拷贝放入D'，~~放回后~~放回原来的D，重复m次后，得到包含m样本的数据集D'

3. 性能度量 performance measure

1) mean squared error: $E(f;D) = \frac{1}{m}\sum_{i=1}^{m}(f(x_i) - y_i)^2$

$acc(f;D) = \frac{1}{m}\sum_{i=1}^{m}\mathbb{I}(f(x_i) = y_i) = 1 - E(f;D)$

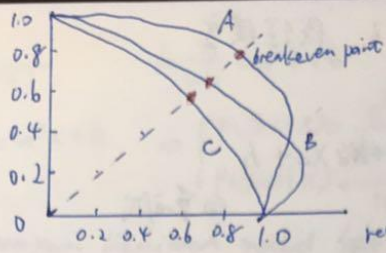
2) precision / recall rate

Confusion Matrix

Actual \ Prediction	+	−
+	TP	FN
−	FP	TN

$Precision = \frac{TP}{TP+FP}$　　$Recall = \frac{TP}{TP+FN}$

P-R plot



A is better than C

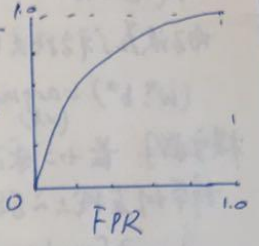
√ breakeven point

$F_1 = \frac{2PR}{P+R} = \frac{2 \cdot TP}{样例总数 + TP - TN}$

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

3) ROC AUC

$TPR = \frac{TP}{TP+FN}$ $\quad FPR = \frac{FP}{TN+FP}$



ROC: (0,1) AUC: area under curve

4) 代价敏感错误率与代价曲线

cost-sensitive error rate

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

4. 比较检验

1) 假设检验 2) 交叉验证t检验 3) McNemar

Contingency table

B \ A	+	−
+	e_{00}	e_{01}
−	e_{10}	e_{11}

$\tau_{\chi^2} = \frac{(|e_{01} - e_{10}| - 1)^2}{e_{01} + e_{10}}$

4) Friedman Nemenyi 后续检验

Friedman — 多算法比较 算法排序

Nemenyi 后续 — 进一步区分各算法

5) 偏差与方差 泛化误差 = 偏差 + 方差 + 噪声 $= bias^2(x) + var(x) + \varepsilon^2$

$var(x) = E_D[f(x;D) - \bar{f}(x)^2]$

$\varepsilon^2 = E_D[(y_D - y)^2]$

$bias^2(x) = (\bar{f}(x) - y)^2$