

机器学习

Lecture 6

性能度量

Definition 1 (均方误差).

$$E(f, \mathcal{D}) = \frac{1}{m} \sum_{i=1}^m (f(x_i) - y_i)^2$$

Definition 2 (错误率).

$$E(f, \mathcal{D}) = \frac{1}{m} \sum_{i=1}^m \mathbb{I}(f(x_i) \neq y_i)$$

Definition 3 (精度).

$$\text{acc}(f, \mathcal{D}) = \frac{1}{m} \sum_{i=1}^m \mathbb{I}(f(x_i) = y_i)$$

真实情况	预测结果	
	正例	反例
正例	TP (真正例)	FN (假反例)
反例	FP (假正例)	TN (真反例)

图 1: 混淆矩阵

Definition 4 (混淆矩阵).

Definition 5 (查准率). 预测正例在全部预测为正例中的比例

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

Definition 6 (查全率/召回率). 预测正例在全部正例中的比例

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

Definition 7 (特异度). 负例在全部负例中的比例

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Definition 8 (F_β score).

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

Defination 9 (ROC). 横轴为负类覆盖率，纵轴为正类覆盖率上的一条曲线

Defination 10 (AUC). *ROC* 曲线下的面积

Defination 11 (代价敏感错误率).

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