

2-6 二分类模型的衡量指标

王中雷

厦门大学王亚南经济研究院和经济学院, 2025

内容摘要

1. 混淆矩阵 (Confusion matrix)

2. ROC 曲线

回顾

1. 二分类模型

- 逻辑回归模型
- 支持向量机
- 随机森林
- (全连接) 神经网络

2. 对于大部分二分类模型而言,

- 我们首先估计一个“得分”
- 然后, 用一个阈值决定对应的类别

3. 在本节中, 我们考虑模型对**测试集**中每个样本都有一个估计好的类别

混淆矩阵

1. 对于测试集中的类别而言，我们有如下两种可能
 - 真实的正数 (True positive number)
 - 真实的负数 (True negative number)
2. 一个特定的模型将预测测试集中的标签，我们有如下两种可能
 - 预测的正数 (Predicted positive number)
 - 预测的负数 (Predicted negative number)

混淆矩阵

1. 对于一个预测模型而言，共有四种可能性

- TP （真正例个数）：正确预测为正类的个数
- FN （假负例个数）：错误地被预测为负例的个数
(真实标签为正例，但被错误地预测为负例)
- TN （真负例个数）：正确预测为负例的个数
- FP （假正例个数）：错误地被预测为正例的个数
(真实标签为负例，但被错误地预测为正例)

混淆矩阵

1. 混淆矩阵

	预测正例	预测负例
真正正例	TP	FN
真实负例	FP	TN

度量指标

	预测正例	预测负例
真实正例	TP	FN
真实负例	FP	TN

1. 准确率 (Accuracy)

$$\begin{aligned}\text{准确率} &= \frac{\text{正确预测个数}}{\text{全部预测个数}} \\ &= \frac{TP + TN}{TP + TN + FP + FN}\end{aligned}$$

2. 精确率 (Precision)

$$\begin{aligned}\text{精确率} &= \frac{\text{真正例个数}}{\text{真正例个数} + \text{假正类个数}} \\ &= \frac{TP}{TP + FP}\end{aligned}$$

度量指标

	预测正例	预测负例
真实正例	TP	FN
真实负例	FP	TN

3. 召回率 (Recall) 或者真阳率 (True Positive Rate, TPR)

$$\begin{aligned}\text{召回率 (真阳率)} &= \frac{\text{真正例个数}}{\text{真正例个数} + \text{假负例个数}} \\ &= \frac{TP}{TP + FN}\end{aligned}$$

4. 假阳率 (False Positive Rate, FPR)

$$\begin{aligned}\text{假阳率} &= \frac{\text{假正例个数}}{\text{假正例个数} + \text{真负例个数}} \\ &= \frac{FP}{FP + TN}\end{aligned}$$

度量指标

5. F1 分数

$$\text{F1 分数} = 2 \times \frac{\text{精确率} \times \text{召回率}}{\text{精确率} + \text{召回率}}$$

6. 对于某特定模型而言，

- 如果该模型仅预测类别，则只对应一组衡量指标
- 如果该模型先预测“得分”，则对应多组衡量指标
 - ▷ 不同阈值对应不同预测类别

ROC 曲线

1. 假设我们知道测试集上模型的预测类别

- 计算召回率 (敏感性, Sensitivity)
- 计算假阳率 (1-特异性 (Specificity))

2. ROC 曲线绘制召回率-假阳率

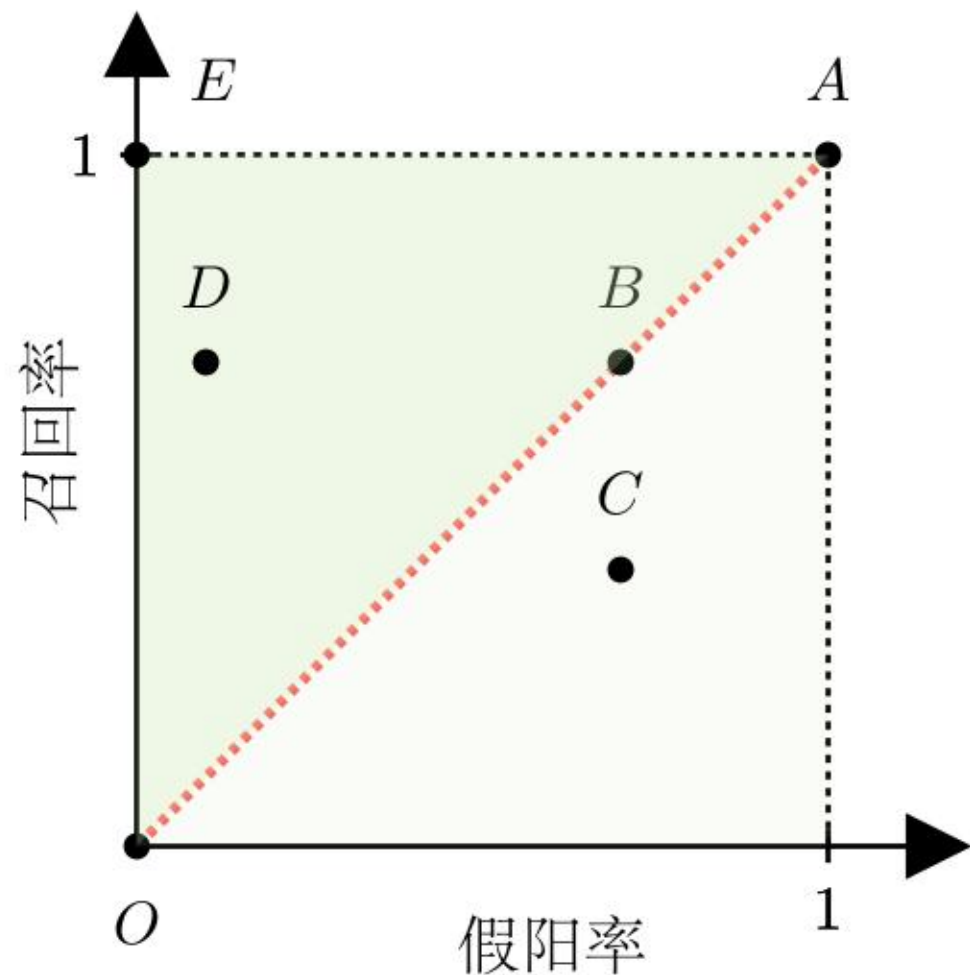
3. ROC 曲线对应的两种情况

- 对于只预测标签正负的模型, 其对应于一个点
- 对于预测“得分”的模型, 其对应一个曲线

ROC 曲线

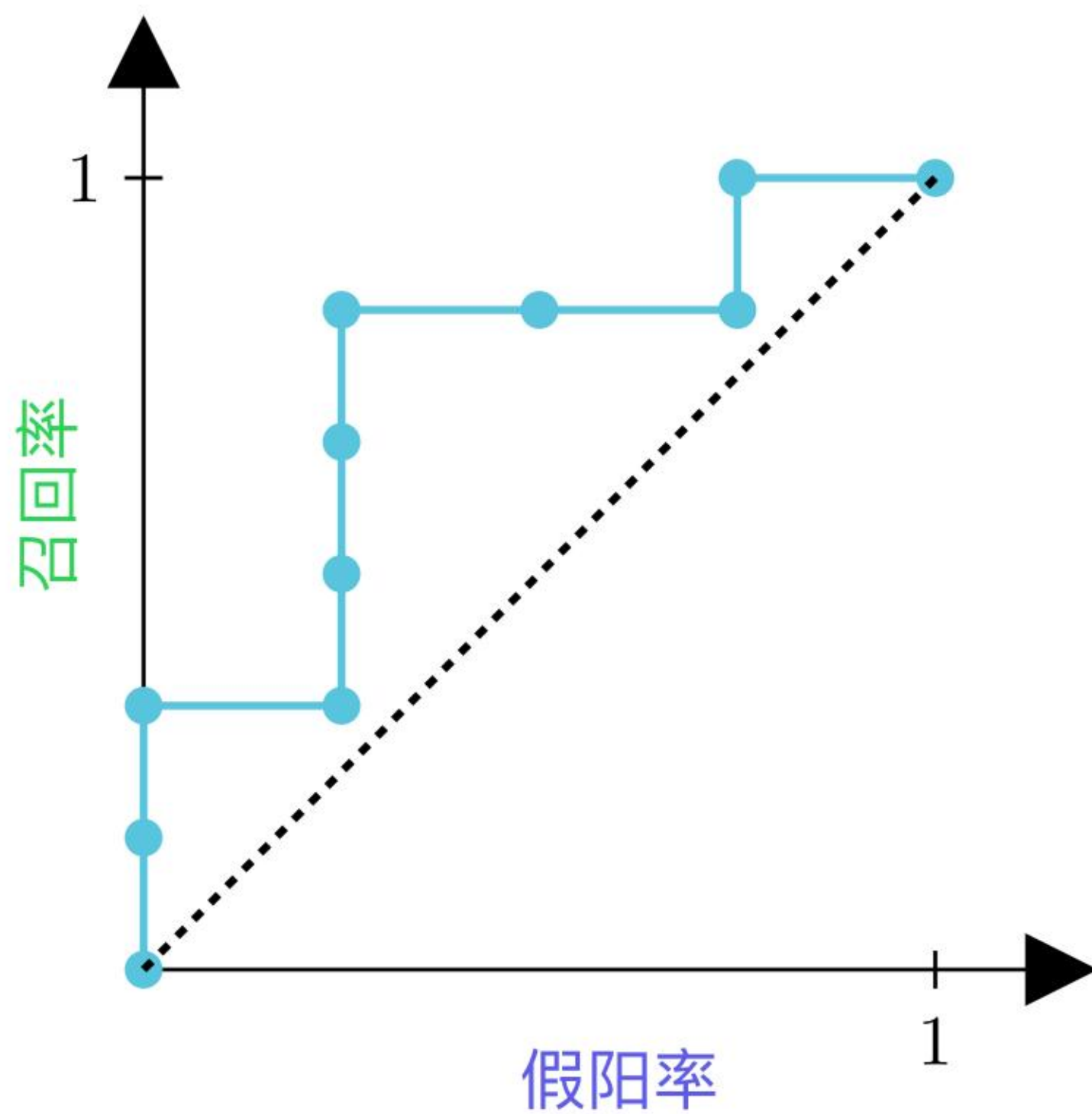
1. 不同的预测结果对应着不同的点

- 对角线上的点：成功率不同的随机猜测模型
 - ▷ $O(0,0)$ ：将所有测试集标签预测成负类
 - ▷ $B(0.7,0.7)$ ：以0.7为概率随机将测试集标签预测成正类
 - ▷ $A(1,1)$ ：将所有测试集标签预测成正类
- 对角线上面的点：好模型
 - ▷ $D(0.1,0.7)$ ：模型的假阳率低，且能够将 70% 的正类预测正确
 - ▷ $E(0,1)$ ：完美模型
 - ▷ 越好的模型对应的点越偏左上角！
- 对角线下方的点：实际上，没有那么糟糕！
 - ▷ $C(0.7,0.4)$ ：尽管召回率低且假阳率高，原因在于模型把指标搞反了！



绘制一条ROC 曲线

1. 考虑一个预测“得分”的二分类模型，
 - 不同的阈值对应着 ROC 曲线上的一个点
 - 将这些点连起来，就得到了一条 ROC 曲线
2. 我们用一个简单的例子展示如何绘制一条 ROC 曲线



样本编号	真实标签	预测标签	预测得分
1	1		0.9
2	1		0.8
3	0		0.7
4	1		0.6
5	1		0.55
6	1		0.54
7	0		0.53
8	0		0.52
9	1		0.51
10	0		0.505

AUC

1. AUC 是 “area under the ROC curve” 的缩写

- AUC 计算的是 ROC 曲线与 x 轴围成的面积

2. 一个好的二分类模型往往对应一个比较大的 AUC 值

- 如果一个二分类模型能够利用 “得分” 将正类样本排在负类样本之前，其 AUC 值往往较高
- 完美模型的 AUC 等于 1，其能够将所有正类排在负类之前.

3. 如果一个模型的 AUC 很低，那这个模型没有那么差，我们颠倒其预测类别即可得到一个好模型

