

模型评估指标的代码实现

scikit-learn

小胖

查准率与查全率

严谨定义

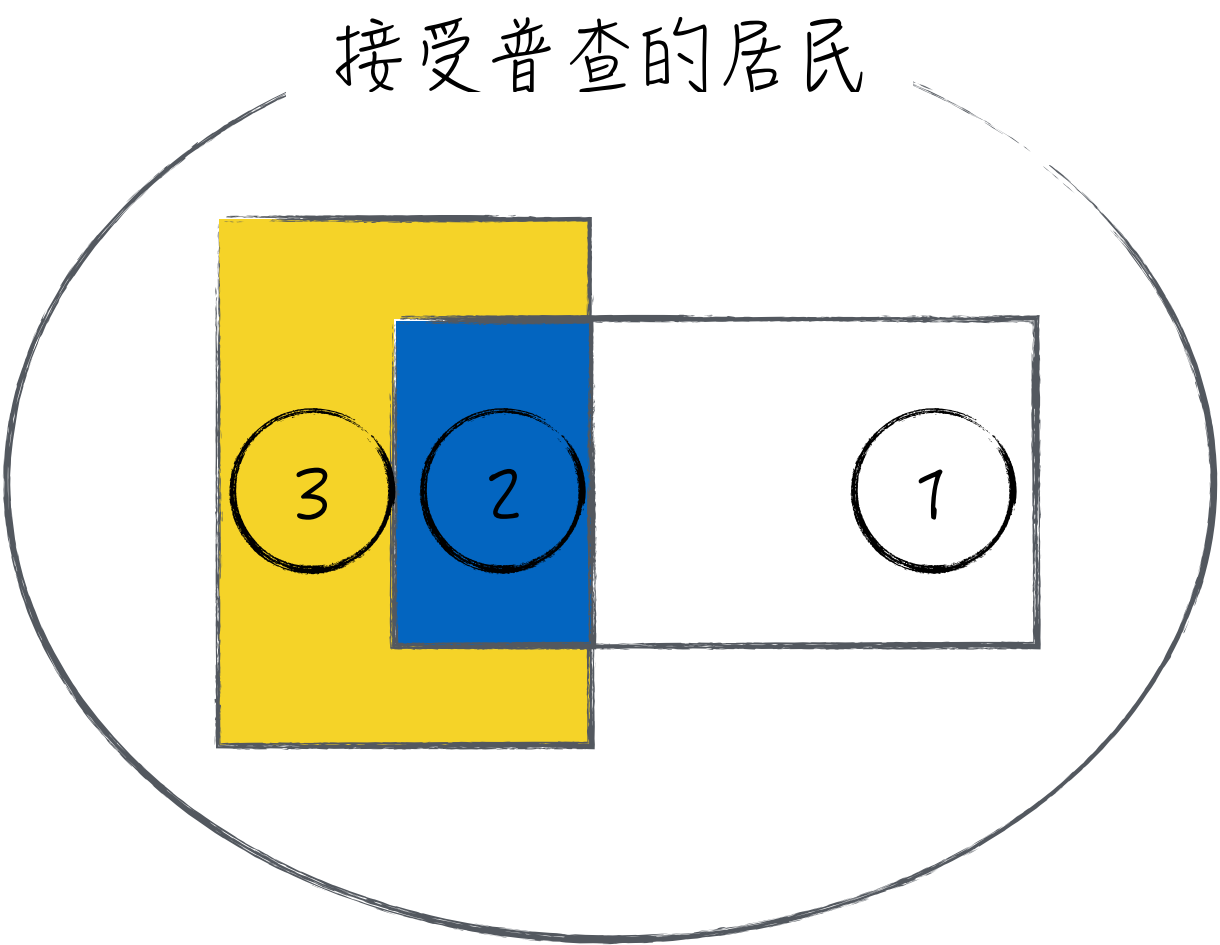
		真实值	
		1	0
预测值	1	真阳性 (true positive) TP 图中2的部分	伪阳性 (false positive) FP 图中3的部分
	0	伪阴性 (false negative) FN 图中1的部分	真阴性 (true negative) TN

$Precision = \frac{TP}{TP + FP},$

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

$Precision = P(y_i = 1 | \hat{y}_i = 1)$

$Recall = P(\hat{y}_i = 1 | y_i = 1)$



1

+

2

= 年收入大于5万的居民

3

+

2

= (模型预测) 年收入大于5万的居民

查准率(precision) =

2

÷

(

3

+

2

)

查全率(recall) =

2

÷

(

1

+

2

)

查准率与查全率

严谨定义

对于一个二分类问题，我们往往对其中一个类别更感兴趣：

- 将这个类别定义为 $y = 1$
- 查准查全率定义都是针对 $y = 1$ 而言的

事实上对于 $y = 0$ 这个类别，也能计算针对这个类别的查准查全率

- 将 y 的取值颠倒一下，重新建模和计算指标，就能得到 $y = 0$ 的查准查全率

		真实值	
		1	0
预测值	1	真阳性 (true positive) TP 图中2的部分	伪阳性 (false positive) FP 图中3的部分
	0	伪阴性 (false negative) FN 图中1的部分	真阴性 (true negative) TN

针对 $y = 1$

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

针对 $y = 0$

$$Precision = \frac{TN}{TN + FN}, \quad Recall = \frac{TN}{TN + FP}$$

THANK YOU

—