性能度量

一.前言

```
错误率和精度虽常用，但并不能满足所有任务需求，错误率衡量了有多少比例被判别错误，挑出的中有多少比例
是正样本，或者所有正样本中有多少比例被挑了出来，显然，错误率和精度不够了，我们需要其他的度量方式
```

二.查全率(recall)、查准率(precision)、F1

1.基础概念：

```
- 真正类(true positive) TP :将真实分类为真的预测为正类（正确预测）
- 假正类(false positive) FP:将真实分类为假的预测为正类（错误预测）
- 正反类(true negative) TN:将真实分类为真的预测为负类（错误预测）
- 假反类(false negative) FN :将真实分类为假的预测为负类（正确预测）
```

2.混淆矩阵：

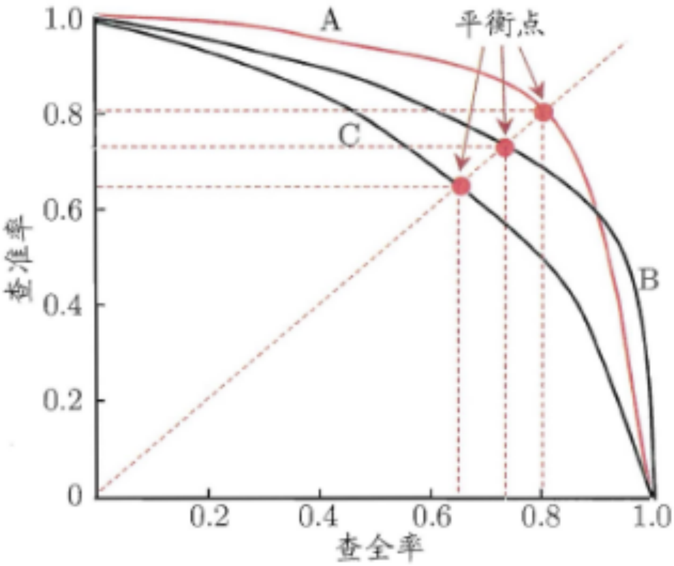
<table>
<tr><th colspan="2" rowspan="2">混淆矩阵</th><th colspan="2">预测结果</th></tr>
<tr><td>正例</td><td>反例</td></tr>
<tr><td rowspan="2">真实情况</td><td>正例</td><td>TP</td><td>FN</td></tr>
<tr><td>反例</td><td>FP</td><td>TN</td></tr>
</table>

3.查准率、查全率

- 查准率：$P = \frac{TP}{TP+FP}$
- 查全率：$R = \frac{TP}{TP+FN}$

> 查准率和查全率是一对矛盾的度量.一般来说，查准率高时，查全率往往 偏低，而查全率高时，查准率往往偏低.

4.P-R图



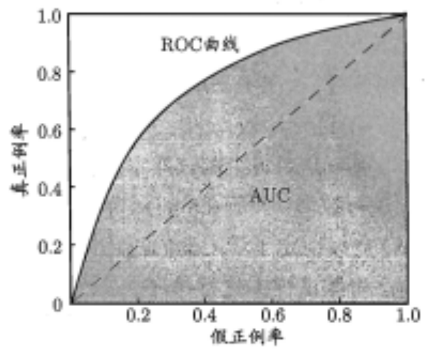
A 、B 、C 三个学习器P-R曲线哪个更好呢？

- "平衡点" (Break-EventPoint，简称BEP),它是"查准率=查全率"时的取值, 例如图中学习器 C 的 BEP 是 0.64，而基于 BEP 的比较，可认为学习器 A 优于B
- F1度量: $F1=\frac{2\times P\times R}{P+R}=\frac{2\times TP}{\text{样例总数}+TP-TN}$

三.ROC与AUC

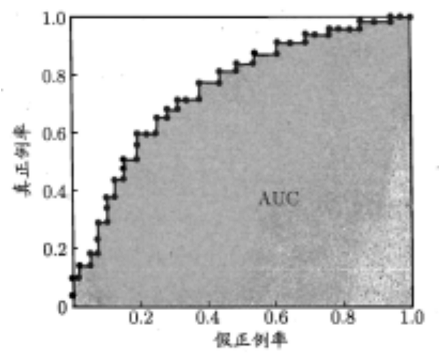
很多学习器是为测试样本产生一个实值或者概率预测，然后设置一个阈值进行分类的，若大于阈值则正类，否则负类（逻辑回归的二分类问题中通常阈值0.5等），这个阈值的好坏直接影响学习器的泛化能力！
在不同的应用中，我们可以先对测试样本进行排序，“最可能”是正例的排在最先面，“最不可能”是正例的排在后面，分类问题就相当于在这个排序中以某个截断点将样本分为两个部分，前一部分是正例，后一部分是反例，然后，采用不同的截断点，例如若更重视“查准率” ，则选择靠前的截断点，若更重视“查全率”可选择靠后的位置截断，因此，排序的好坏，直接影响学习器在不同任务下的泛化能力的好坏，ROC就是从这个角度出发研究泛化性能的工具 。
真正例率：$\mathrm{TPR}=\frac{TP}{TP+FN}$

假正例率：$\mathrm{FPR}=\frac{FP}{TN+FP}$



(a) ROC 曲线与 AUC

现实任务中通常是利用有限个测试样例来绘制ROC图此时仅能获得有限个(真正例率，假正例率)坐标对，无法产生上图中的光滑 ROC 曲线：



(b) 基于有限样例绘制的 ROC 曲线与 AUC

只能绘制出近似ROC曲线.绘图过程很简单:给定 m^{+} 个正例和 m^{-} 个反例根据学习器预测结果对样例进行排序，然后把分类阈值设为最大，即把所有样例均预测为反例，此时真正例率和假正例率均为0，在坐标(0，0)处标记一个点.然后，将分类阐值依次设为每个样例的预测值，即依次将每个样例划分为正例.设前一个标记点坐标为(X，y)，当前若为真正例，则对应标记点的坐标为 (X ，$y+\frac{1}{m^{2}}$);当前若为假正例，则对应标记点的坐标为($x+\frac{1}{m^{-}}$ ，y)，然后用线段连接相邻点即得.

进行学习器的比较时,与P-R图相似，若一个学习器的ROC曲线被另一个学习器的曲线完全"包住"，则可断言后者的性能优于前;若两个学习器的ROC曲线发生交叉，则难以一般性地断言两者孰优孰劣. 此时如果一定要进行比较,则较为合理的判据是比较ROC曲线下的面积，即AUC(Area Under ROC Curve)，如图上图所示.从定义可知，AUC可通过对ROC曲线下各部分的面积求和而得.假定ROC曲线是由坐标为{(XI,yI),(X2,Y2),...,(xm,Ym)} 的点按序连接而形成(x_1 =0, x_m = 1),则 AUC 可估算为:
$\mathrm{AUC}=\frac{1}{2} \sum_{i=1}^{m-1}\left(x_{i+1}-x_i\right) \cdot\left(y_i+y_{i+1}\right)$
可以看出，AUC是样本预测的排序质量，所以与排序顺序有紧密联系 。排序损失（loss）定义为：
$\ell_{\text {rank }}=\frac{1}{m^{+} m^{-}} \sum_{r+6n+r^{-} \in D^{-}}\left(\mathbb{I}\left(f\left(\boldsymbol{x}^{+}\right)<f\left(\boldsymbol{x}^{-}\right)\right)+\frac{1}{2} \mathrm{I}\left(f\left(\boldsymbol{x}^{+}\right)=f\left(\boldsymbol{x}^{-}\right)\right)\right)$

m^+ :正例个数
m^- ：反例个数
D^+ ：正例集合
D^- ：反例集合

AUC = 1 - ℓ_{rank}