准确率、精确率以及召回率

* TP（True Positive）：正确的正例，一个实例是正类并且也被判定成正类
* FN（False Negative）：错误的反例，漏报，本为正类但判定为假类
* FP（False Positive）：错误的正例，误报，本为假类但判定为正类
* TN（True Negative）：正确的反例，一个实例是假类并且也被判定成假类

准确率：

精确率：

召回率：

F-Score:

F1是精确率和召回率的调和平均。F1-score越大说明模型质量更高。也可以绘制出P-R曲线图，观察它们的分布情况。AP就是PR曲线与X轴围成的图形面积，AP值为1时模型性能最好。

图表, 折线图

描述已自动生成

ROC和AUC

1. 伪阳性率（[FPR](https://link.zhihu.com/?target=https%3A//baike.baidu.com/item/FPR/6343296)）：
2. 真阳性率（[TPR](https://link.zhihu.com/?target=https%3A//baike.baidu.com/item/TPR/5548598)）：
3. 伪阴性率（[FNR](https://link.zhihu.com/?target=https%3A//baike.baidu.com/item/FNR/5609400)）：
4. 真阴性率（[TNR](https://link.zhihu.com/?target=https%3A//baike.baidu.com/item/TNR/4663071)）：

ROC曲线就是根据TPR和FPR的坐标点绘制出来的。在一个二分类模型中，对于所得到的连续结果，假设已确定一个阀值，比如说 0.6，大于这个值的实例划归为正类，小于这个值则划到负类中。如果减小阀值，减到0.5，固然能识别出更多的正类，也就是提高了识别出的正例占所有正例 的比类，即TPR,但同时也将更多的负实例当作了正实例，即提高了FPR。

图表, 折线图

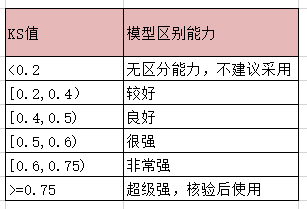
描述已自动生成

ROC曲线上的每一个点对应于一个threshold，对于一个分类器，每个threshold下会有一个TPR和FPR。比如Threshold最大时，，对应于原点；Threshold最小时，，对应于右上角的点(1,1)；随着阈值theta增加，TP和FP都减小，TPR和FPR也减小，ROC点向左下移动；而AUC就是ROC曲线与下方坐标轴围成的面积。

KS：KS曲线的纵轴是表示TPR和FPR的值，就是这两个值可以同时在一个纵轴上体现，横轴就是阈值，然后在两条曲线分隔最开的地方，对应的就是最好的阈值，也是该模型最好的AUC值。

图表, 折线图

描述已自动生成



|  |  |  |
| --- | --- | --- |
| 监督学习  (Supervised learning) | 使用有标签的数据训练模型 | 图表, 散点图  描述已自动生成 |
| 无监督学习  (Unsupervised learning) | 使用无标签的数据训练模型 |
| 强化学习  (Reinforcement learning) | 使用reward机制训练模型 | 图示  描述已自动生成 |