

ROC 曲线的具体画法

分类结果混淆矩阵

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)

一个二分类问题，两个概念：

真正例率（TPR）：正例中预测为正例的比例

假正例率（FPR）：反例中预测为正例的比例

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

$$FPR = \frac{FP}{TN + FP}$$

假设样本数为 6，现在有一个**分类器 1**，它对样本的分类结果如下表（按预测值从大到小排序）

标签	预测值
1	0.9
1	0.8
1	0.7
0	0.3
0	0.2
0	0.1

ROC 曲线的横轴为假正例率，纵轴为真正例率，范围都是[0, 1]，现在我们开始画图——根据从大到小遍历预测值，把当前的预测值当做阈值，计算 FPR 和 TPR。

step1: 选择阈值最大，即为 1，正例中和反例中都没有预测值大于等于 1 的，所以 FPR=TPR=0。

step2: 根据上表，选择阈值为 0.9，正例中有 1 个样本的预测值大于等于 0.9，反例中有 0 个，所以，TPR=1/3，FPR=0。

step3: 根据上表，选择阈值为 0.8，正例中有 2 个样本的预测值大于等于 0.8，反例中有 0 个，所以，TPR=2/3，FPR=0。

step4: 根据上表，选择阈值为 0.7，正例中有 3 个样本的预测值大于等于 0.7，反例中有 0 个，所以，TPR=1，FPR=0。

step5: 根据上表，选择阈值为 0.3，正例中有 3 个样本的预测值大于等于 0.3，

反例中有 1 个，所以， $TPR=1$ ， $FPR=1/3$ 。

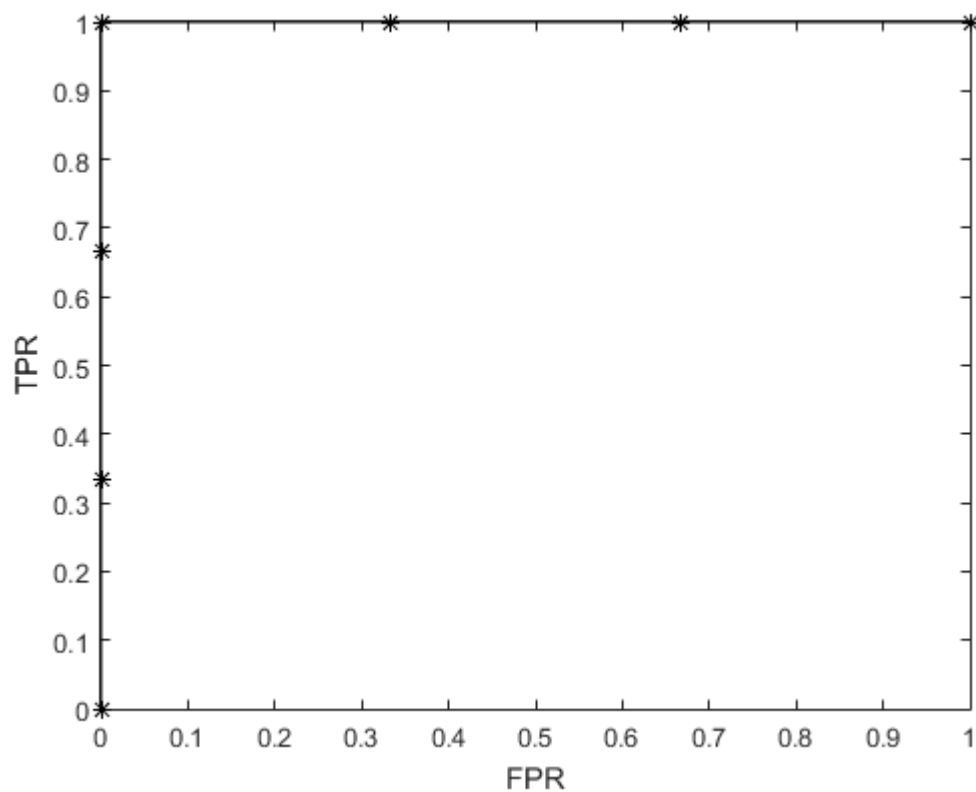
step6: 根据上表，选择阈值为 0.2，正例中有 3 个样本的预测值大于等于 0.2，反例中有 2 个，所以， $TPR=1$ ， $FPR=2/3$ 。

step7: 根据上表，选择阈值为 0.1，正例中有 3 个样本的预测值大于等于 0.1，反例中有 3 个，所以， $TPR=1$ ， $FPR=1$ 。

综上，我们得到下表

FPR	TPR
0	1/3
0	2/3
0	1
1/3	1
2/3	1
1	1

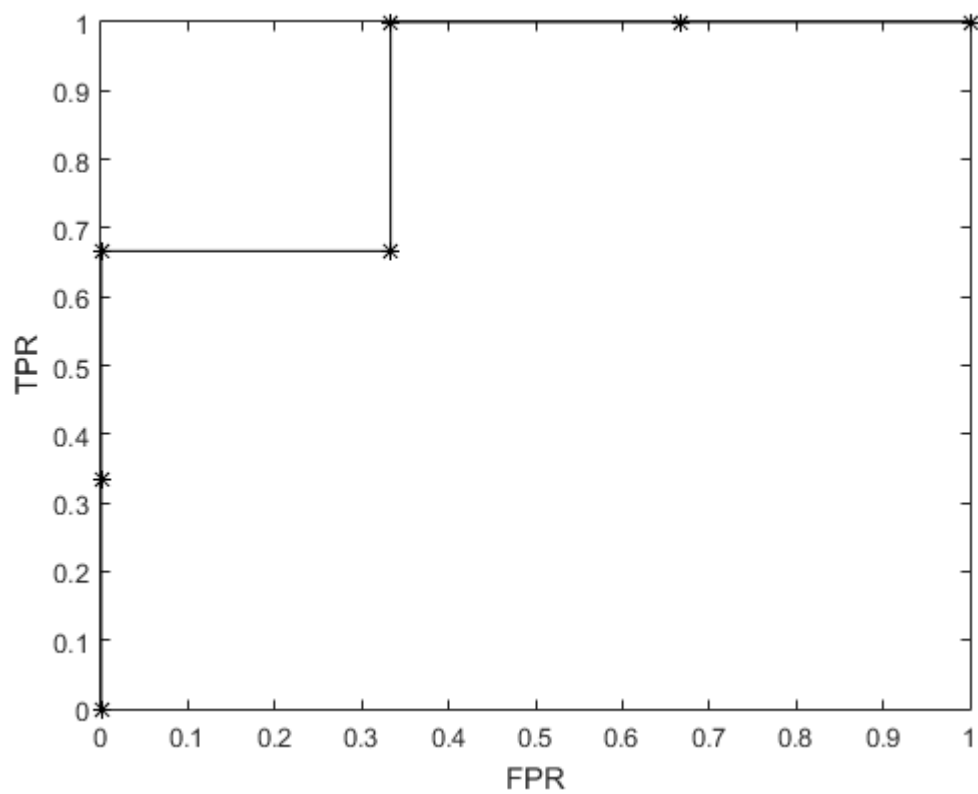
描点连线，ROC 图如下：



有一个**分类器 2**，对同样一组样本，分类结果如下

标签	预测值
1	0.9
1	0.8
0	0.75
1	0.7
0	0.2
0	0.1

根据上面描述的方法，画出 ROC 曲线如下



很明显，分类器 1 更好。