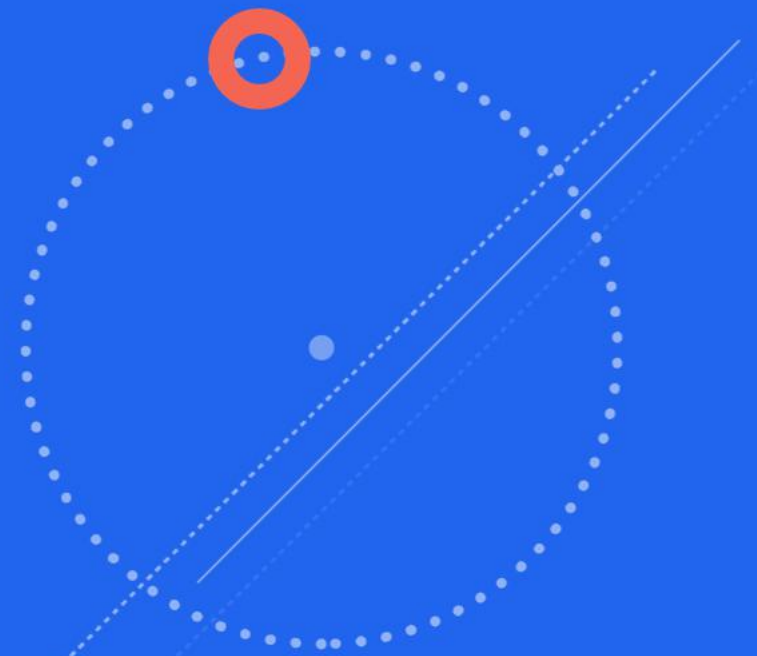


ROC曲线

Receiver Operating Characteristic



两个先导概念

False Positive Rate, FPR, 假阳

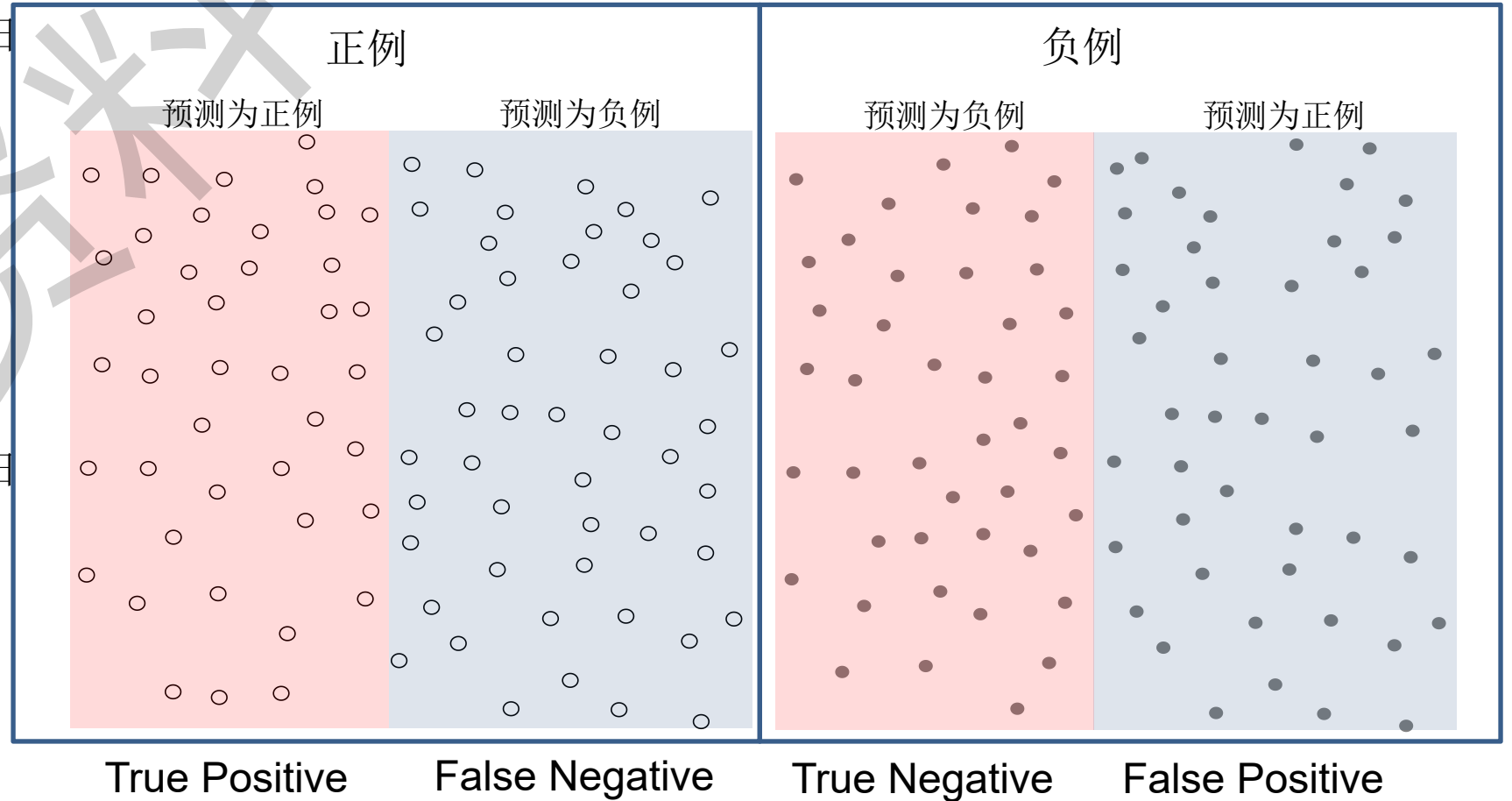
$$FPR = \frac{FP}{FP + TN}$$

虚惊一场

True Positive Rate, TPR, 真阳

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

不幸言中



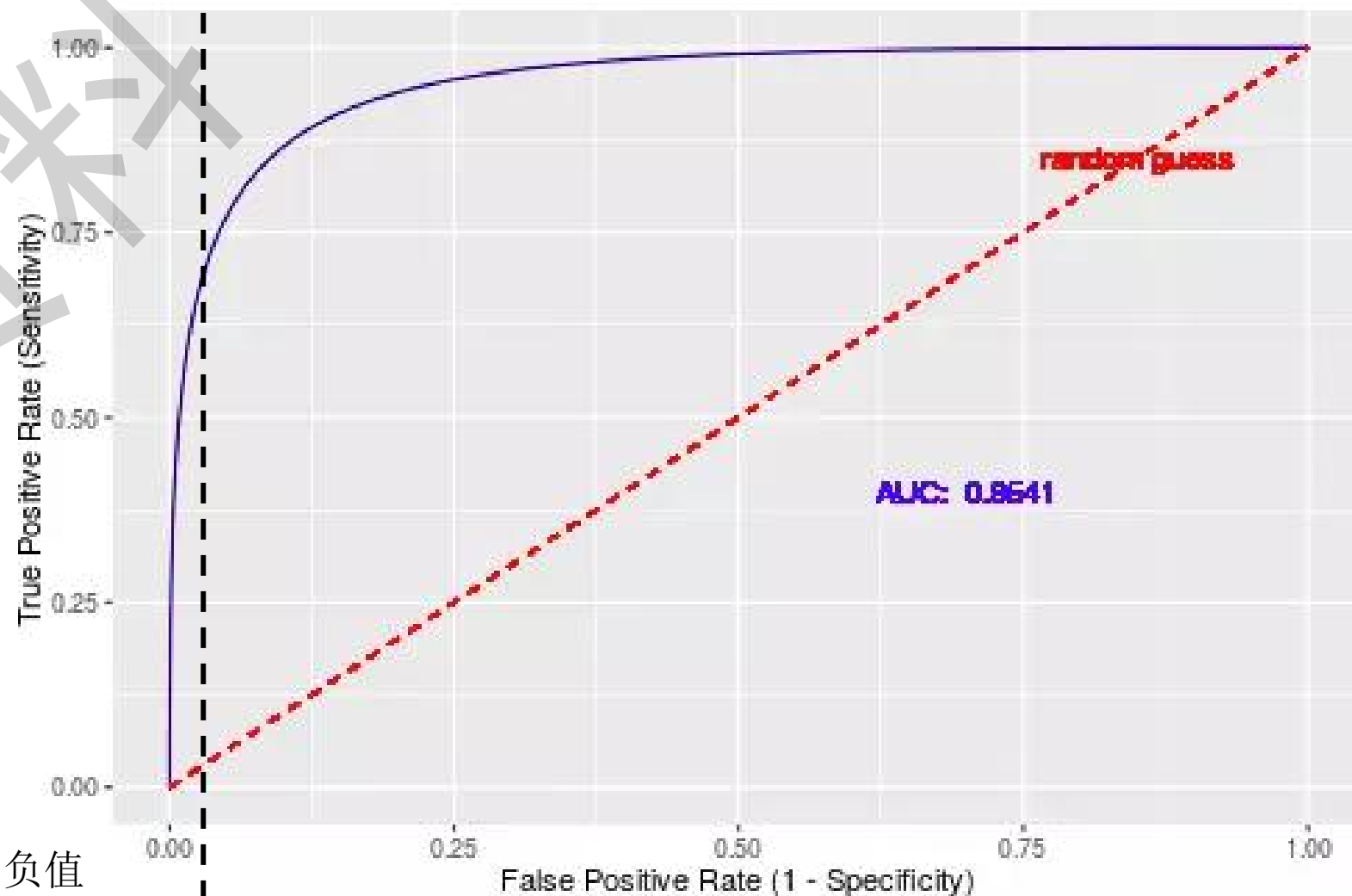
- **ROC(Receiver Operating Characteristic):** 是一条曲线, 由FPR和TPR的点连成。横轴是 FPR, 纵轴是 TPR

ROC curve

无脑预测正值

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

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



AUC 后面讲到

无脑预测负值

- ROC 来源于二战时雷达兵对雷达信号的判断。
- 雷达兵需要解析雷达信号。
- 当雷达兵收到一个雷达信号的时候，就要做出判断，这个信号是由对方的轰炸机造成的还是由于大鸟飞过造成的。
- 比较谨慎的雷达兵会更倾向于将信号分类到轰炸机，而较为乐观的雷达兵可能更倾向于大鸟。
- 为了更好的度量雷达效果，所以，引入了 ROC 曲线

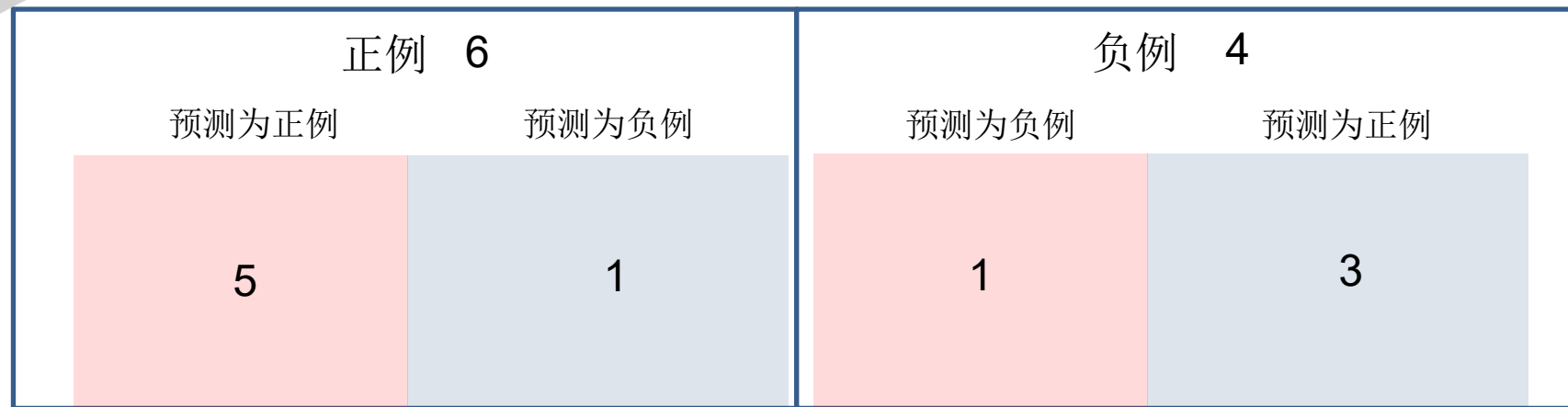
示例

- 假设存在 10个 雷达信号，其中 6个是轰炸机，4个鸟。（真实分类情况）
- 分析员给出的分析结果是：8个轰炸机 2个鸟。其中，错把 3 只鸟分类为了轰炸机，把1个轰炸机当成了鸟（机器吐出的预测分类情况）

飞机视为正例：TP = 5 FN = 1 TN = 1 FP = 3

$$FPR = \frac{FP}{FP + TN} = \frac{3}{4} = 0.75$$

$$TPR = \frac{TP}{TP + FN} = \frac{5}{6} = 0.83$$



ROC 曲线上的点

$(FPR, TPR) = (0.75, 0.83)$

True Positive

False Negative

True Negative

False Positive

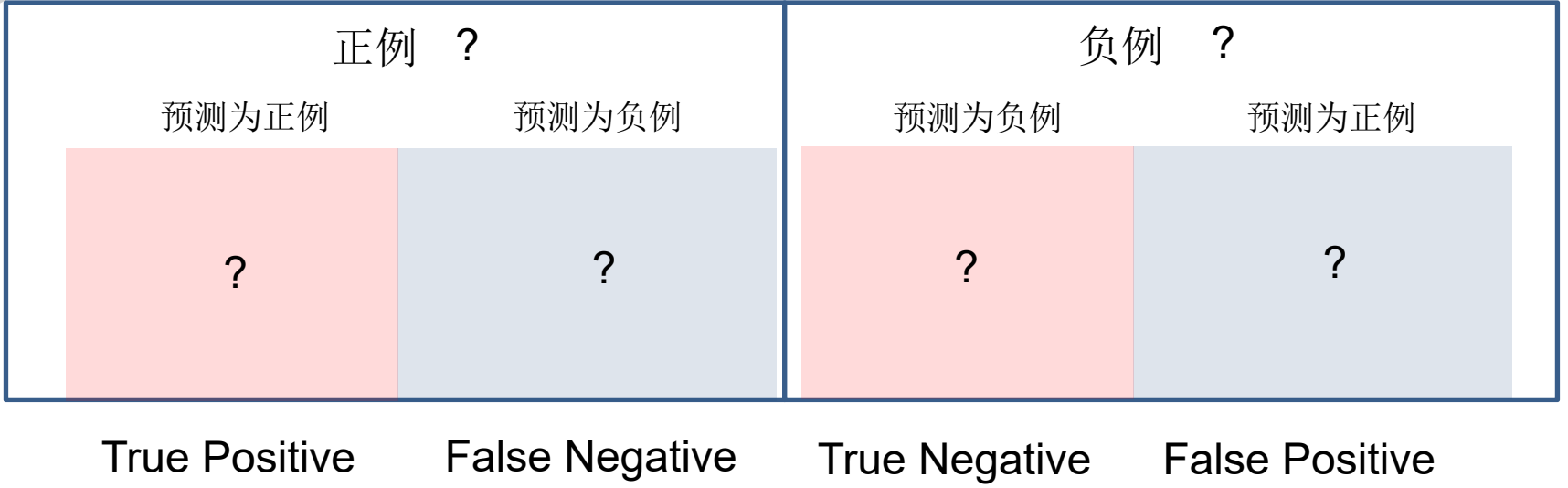
练习题1—题目

- 假设存在 100个 雷达信号，其中 75 个是轰炸机，25 个鸟。（真实分类情况）
- 分析员给出的分析结果是：76 个轰炸机 24 个鸟。其中，错把 2 只鸟分类为了轰炸机，把 3 个轰炸机当成了鸟（机器吐出的预测分类情况）

飞机视为正例：TP = ? FN = ? TN = ? FP = ?

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

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



ROC 曲线上的点

$(FPR, TPR) = (?, ?)$

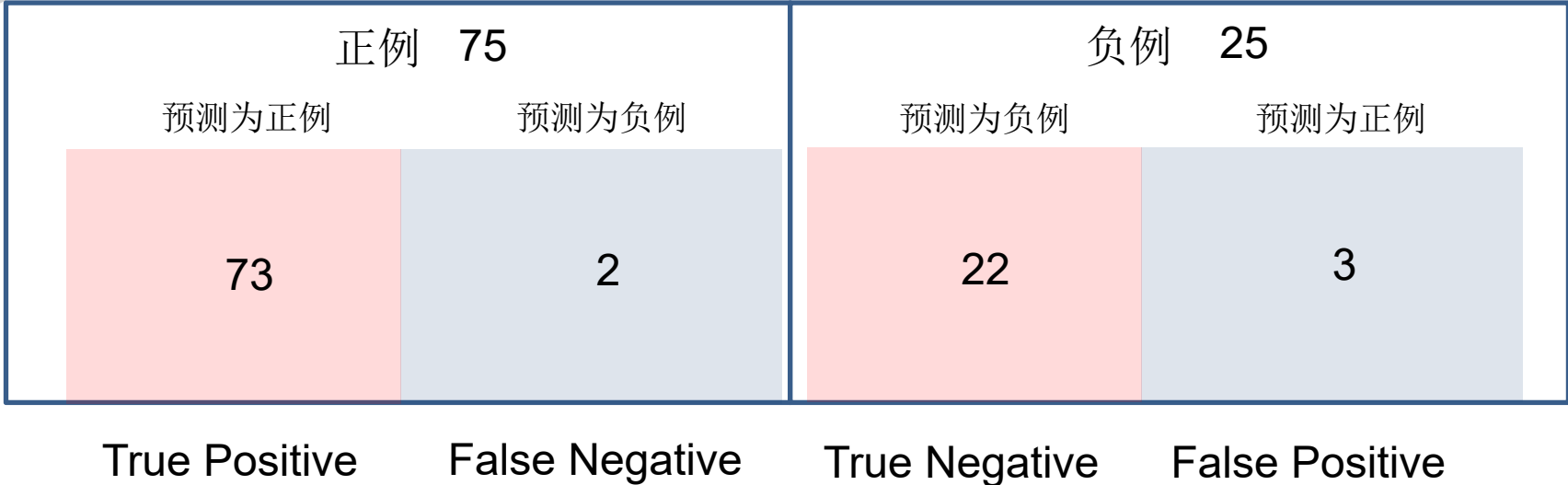
练习题1—参考答案

- 假设存在 100个 雷达信号，其中 75 个是轰炸机，25 个鸟。（真实分类情况）
- 分析员给出的分析结果是：76 个轰炸机 24 个鸟。其中，错把 2 只鸟分类为了轰炸机，把 3 个轰炸机当成了鸟（机器吐出的预测分类情况）

飞机视为正例：TP = 72 FN = 2 TN = 22 FP = 3

$$FPR = \frac{FP}{FP + TN} = \frac{3}{25} = 0.12$$

$$TPR = \frac{TP}{TP + FN} = \frac{73}{75} = 0.97$$



ROC 曲线上的点

$(FPR, TPR) = (0.12, 0.97)$

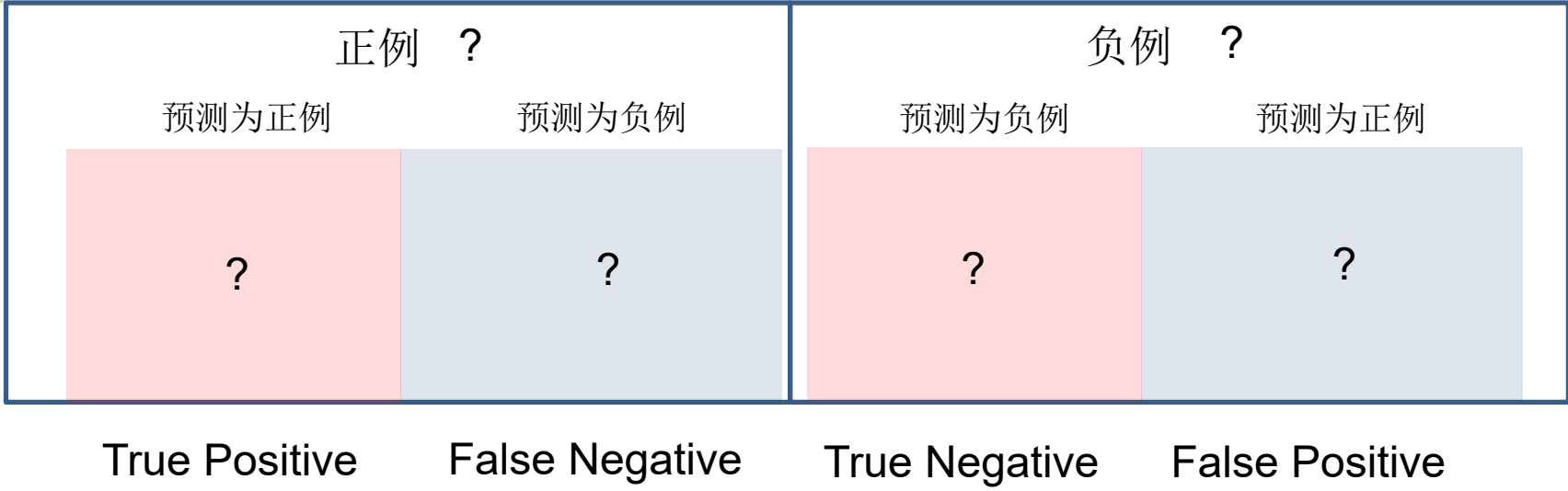
练习题2—题目—课后练习

- 假设存在 100个 雷达信号，其中 65 个是轰炸机，35 个鸟。（真实分类情况）
- 分析员给出的分析结果是：80 个轰炸机 20 个鸟。其中，错把 20只鸟分类为了轰炸机，把 5 个轰炸机当成了鸟（机器吐出的预测分类情况）

飞机视为正例：TP = ? FN = ? TN = ? FP = ?

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

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



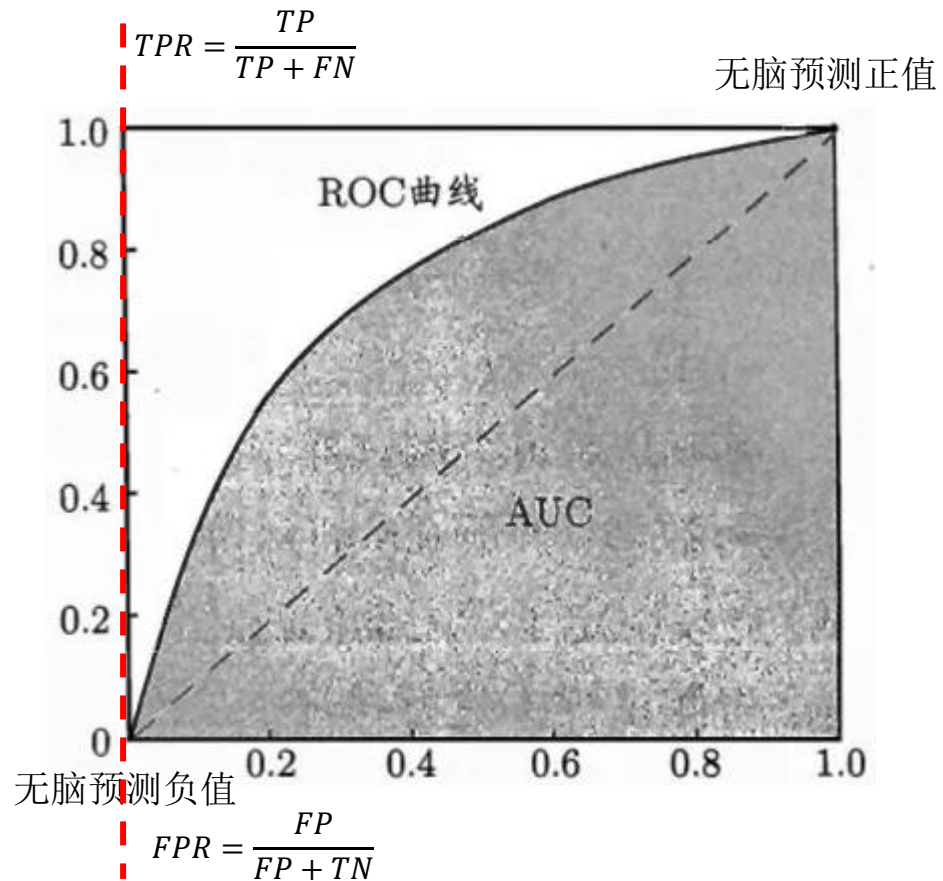
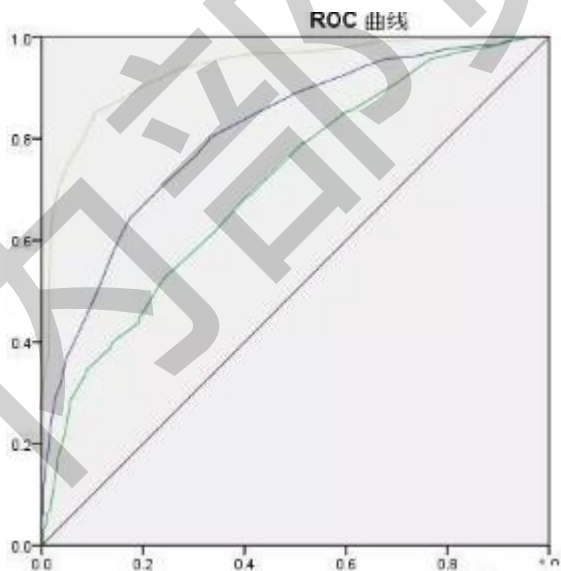
ROC 曲线上的点

(FPR, TPR) = (? , ?)

- 我们根据学习器的预测结果，把阈值从0变到最大
 - 即刚开始是把每个样本作为正例进行预测，随着阈值的增大，学习器预测正样例数越来越少，直到最后没有一个样本是正样例。
 - 在这一过程中，每次计算出 FPR、TPR 的值，分别以它们为横、纵坐标作图，就得到了“ROC曲线”

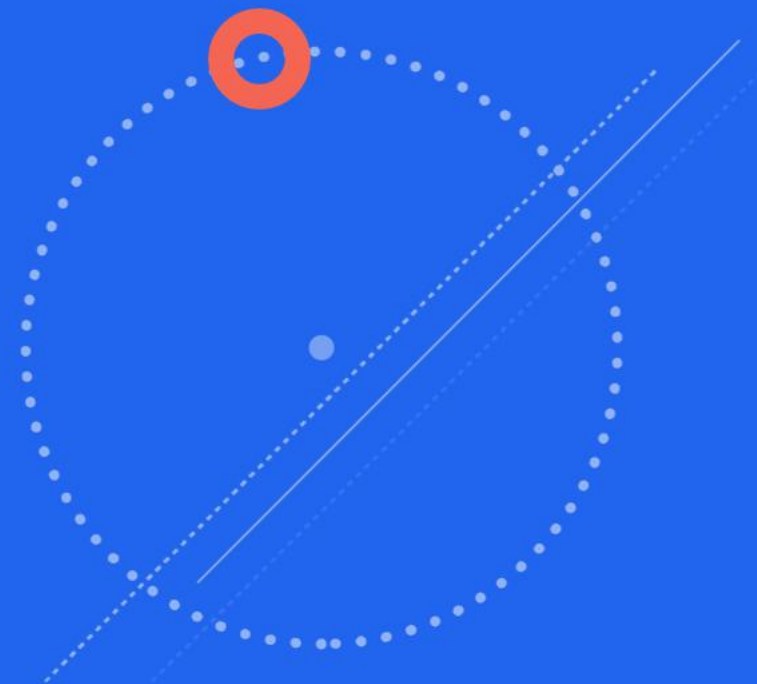
ROC 曲线的意义

- ROC 曲线能很容易的查出任意阈值对学习器的影响
- 有助于选择最佳的阈值
 - ROC 曲线越靠近左上角，模型越好
- 可以对不同的学习器比较性能
 - 将各个学习器的ROC曲线绘制到同一坐标中，直观地鉴别优劣，靠近左上角的ROC曲线所代表的学习器更好



AUC

Area under roc Curve
曲线下面积



- 假设两个雷达分析员，就是两个算法的话，那么，在已经绘制出了 ROC 曲线后哪个算法更优呢？

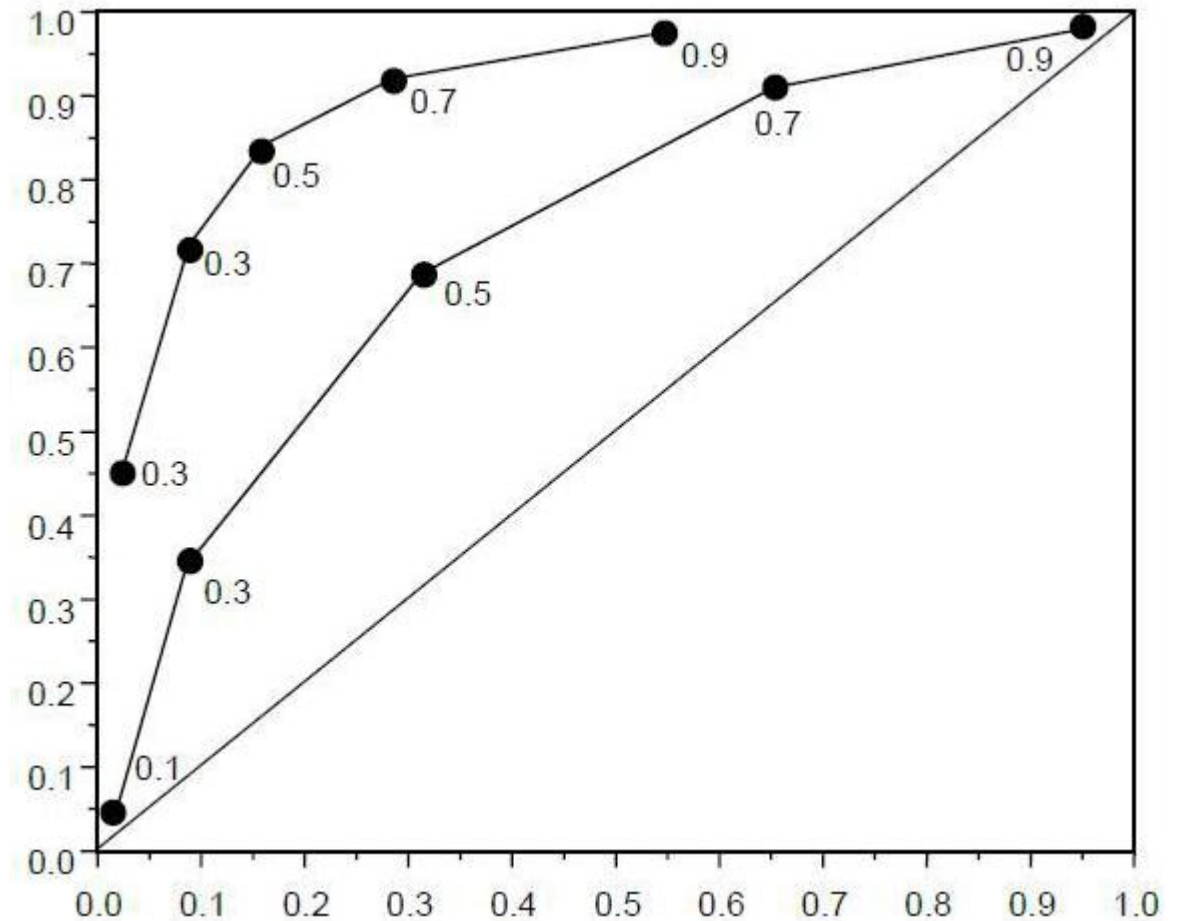
如果两条ROC曲线没有相交，
我们可以根据哪条曲线最靠近左上角
来判断哪条曲线更优

但是如果有交叉呢？

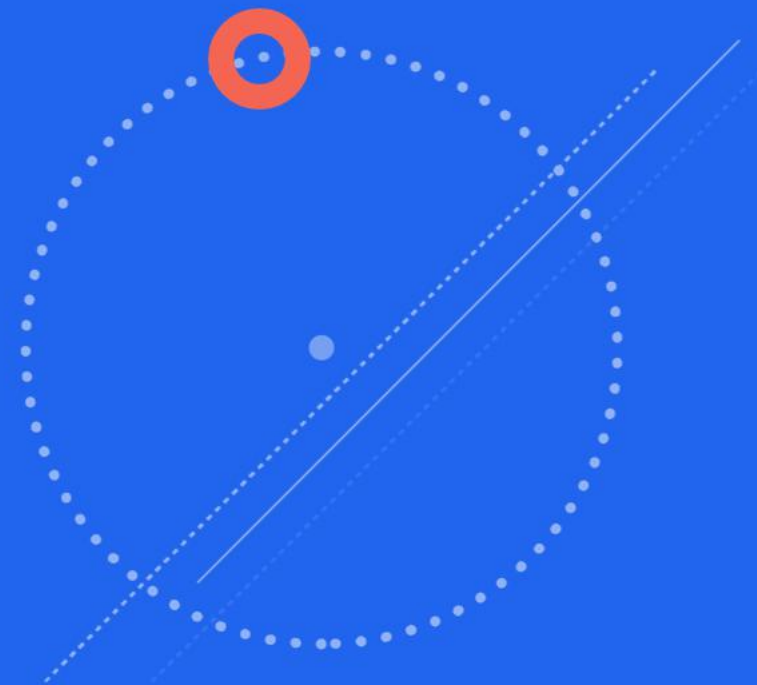
AUC 曲线

度量的方法：AUC 曲线下面积
哪个曲线覆盖面积更大，哪个曲线就更优秀

计算AUC值：沿着ROC横轴做积分就可以了



谢谢



内部资料