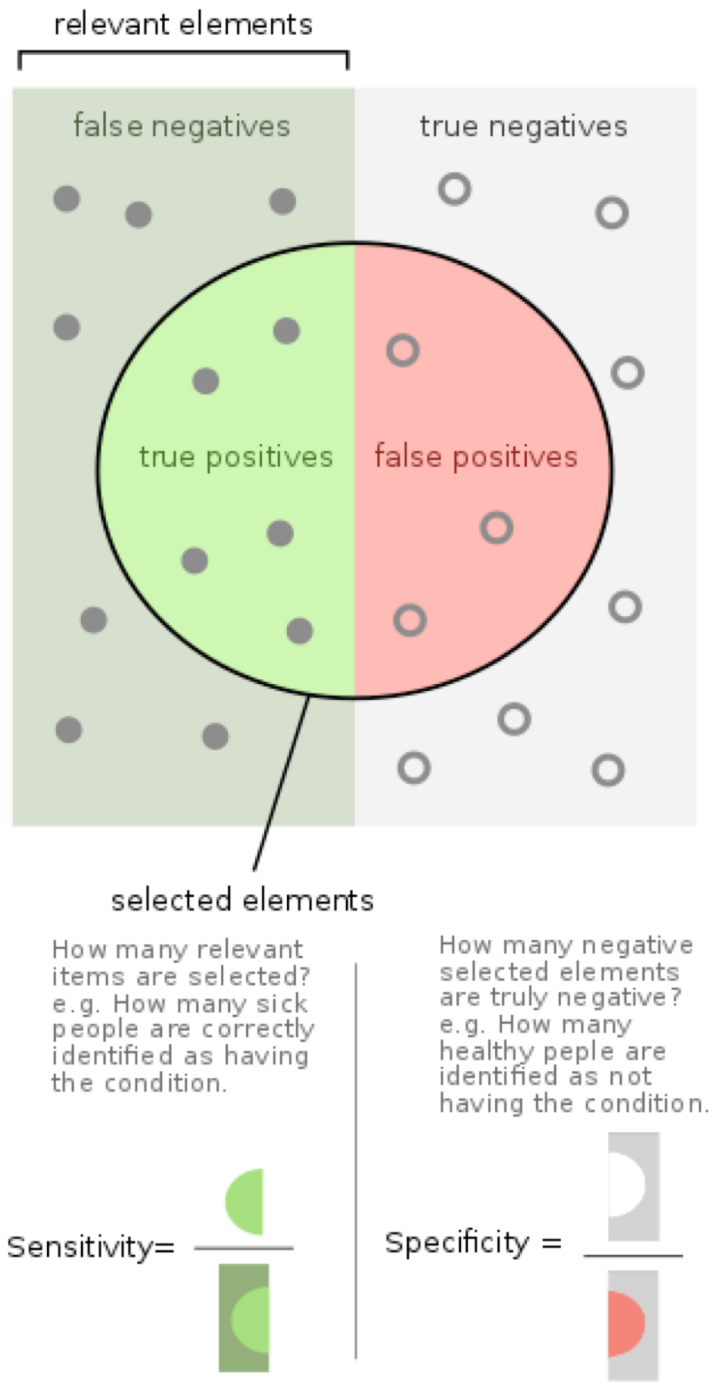
AUC: sensitivity and specificity

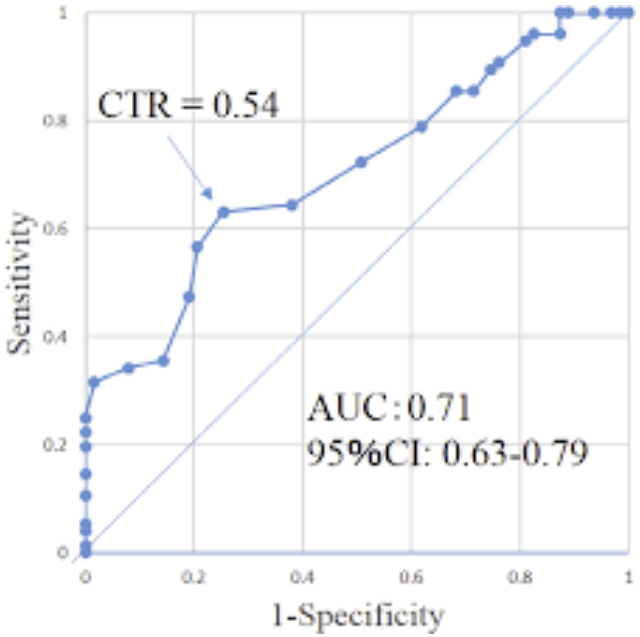
- Sensitivity = True Positive Rate
- Specificity = False Positive Rate



AUC: 解释AUC

- 从正样本中抽出一个样本，x_1
- 从负样本中抽出一个样本，x_2
- 用训练好的分类起对这两个样本进行预测，分别得到，$P(x_1)$ 和 $P(x_2)$
- AUC=正样本大于负样本概率的概率，即

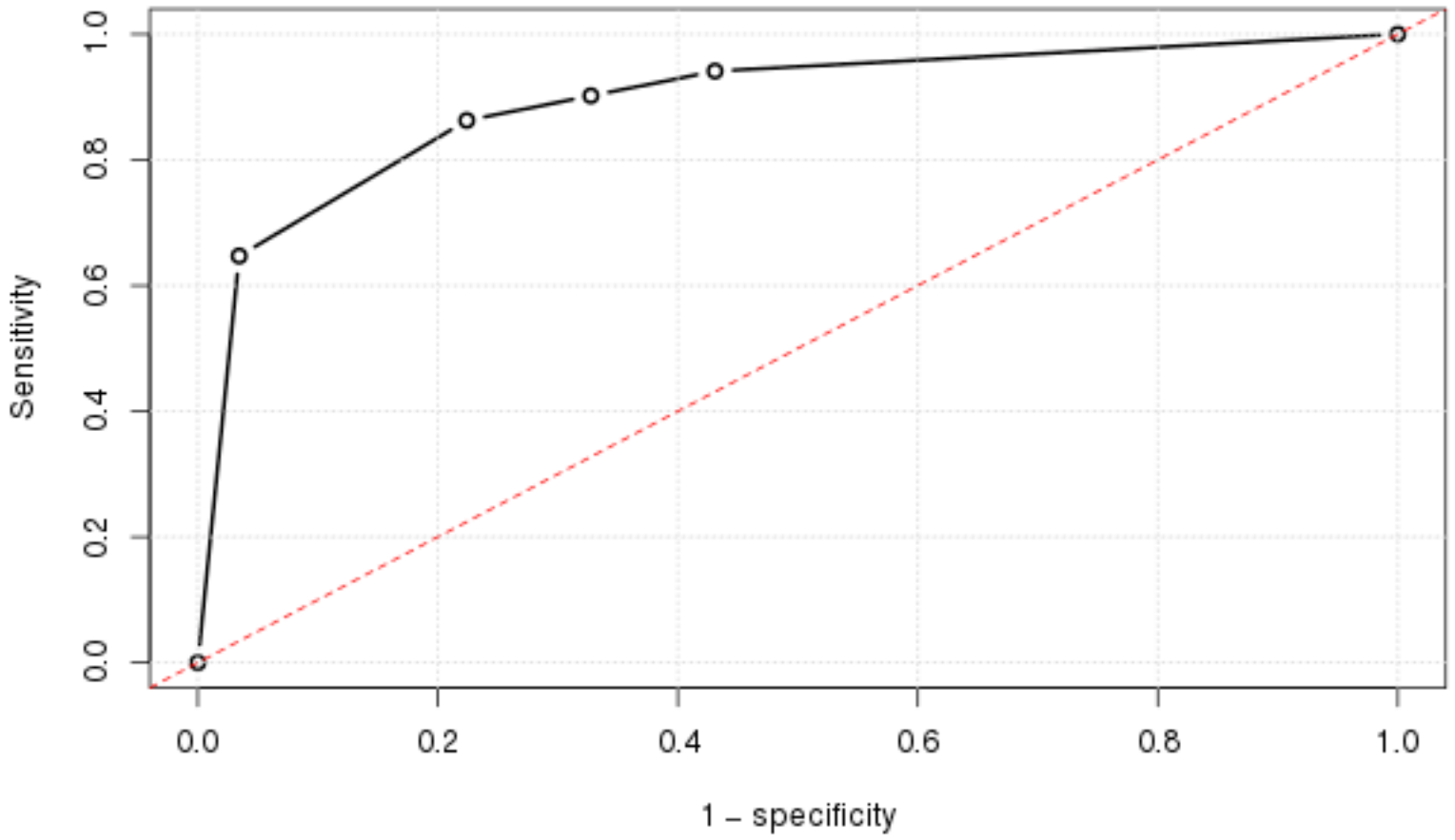
$$AUC = P(P(x_1) > P(x_2))$$



AUC：计算AUC，定义法

- 根据阈值计算坐标点
- 分段计算梯形或者矩形的面积，$f(x_i)$

$$AUC = \sum_{i=1}^{n} f(x_i)$$



AUC：计算AUC，索引法

- 根据AUC的概率解释
- Wilcoxon-Mann-Witney Test

$$\frac{\sum I(P_{正样本}, P_{负样本})}{M * N}$$

$$I(P_{正样本}, P_{负样本}) = \begin{cases} 1, P_{正样本} > P_{正样本} \\ 0.5, P_{正样本} = P_{负样本} \\ 0, P_{正样本} < P_{负样本} \end{cases}$$

AUC：计算AUC，索引法

Sample	Label	ctr(方法一)	ctr(方法二)
A	0	0.25	0.25
B	0	0.35	0.35
C	1	0.30	0.35
D	1	0.40	0.40

$$I(P_{正样本}, P_{负样本}) = \begin{cases} 1, P_{正样本} > P_{正样本} \\ 0.5, P_{正样本} = P_{负样本} \\ 0, P_{正样本} < P_{负样本} \end{cases}$$

$$\begin{cases} I_1(D,A) = 1 \\ I_1(D,B) = 1 \\ I_1(C,A) = 1 \\ I_1(C,B) = 0 \end{cases} \qquad \begin{cases} I_2(D,A) = 1 \\ I_2(D,B) = 1 \\ I_2(C,A) = 1 \\ I_2(C,B) = 0.5 \end{cases}$$

AUC：计算AUC，索引法

Sample	Label	ctr(方法一)	ctr(方法二)
A	0	0.25	0.25
B	0	0.35	0.35
C	1	0.30	0.35
D	1	0.40	0.40

$$AUC_1 = \frac{I_1(D,A) + I_1(D,B) + I_1(C,A) + I_1(C,B)}{4} = \frac{1+1+1+0}{4} = 0.75$$

$$AUC_2 = \frac{I_2(D,A) + I_2(D,B) + I_2(C,A) + I_2(C,B)}{4} = \frac{1+1+1+0.5}{4} = 0.875$$

AUC：计算AUC，排序法

$$AUC = \frac{\sum_{i \in P} r_i - \frac{M(M+1)}{2}}{M \times N}$$

其中，P是正样本集合，M为正样本个数，N为负样本个数，r 代表了ctr排序后的位置。

AUC：计算AUC，排序法

Sample	Label	ctr	rank
A	0	0.25	1
B	0	0.35	3
C	1	0.30	2
D	1	0.40	4

$$AUC = \frac{\sum_{i \in P} r_i - \frac{M(M+1)}{2}}{M \times N} = \frac{4 + 2 - \frac{2 \times 3}{2}}{2 \times 2} = 0.75$$

AUC: GAUC

- User weighted AUC

$$GAUC = \frac{\sum_{i \in m} I_m \times AUC_i}{\sum_{i \in m} I_m}$$

- 用户个数 m，用户的展现量I_m，用户对应的AUC_i