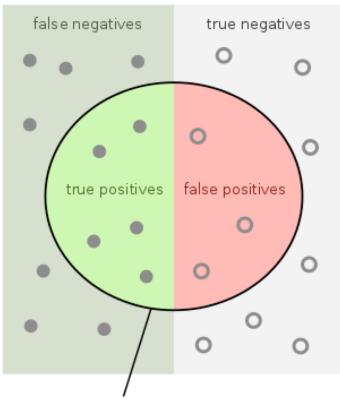
### **AUC:** sensitivity and specificity

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

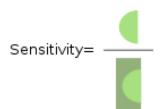
#### relevant elements



#### selected elements

How many relevant items are selected? e.g. How many sick people are correctly identified as having the condition.

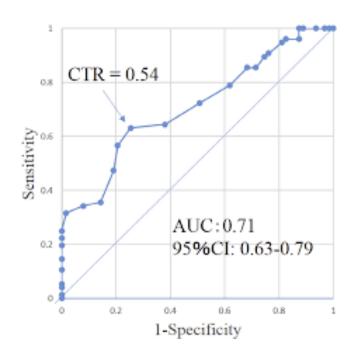
How many negative selected elements are truly negative? e.g. How many healthy peple are identified as not having the condition.



#### AUC: 解释AUC

- 从正样本中抽出一个样本, x<sub>1</sub>
- 从负样本中抽出一个样本, x<sub>2</sub>
- 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

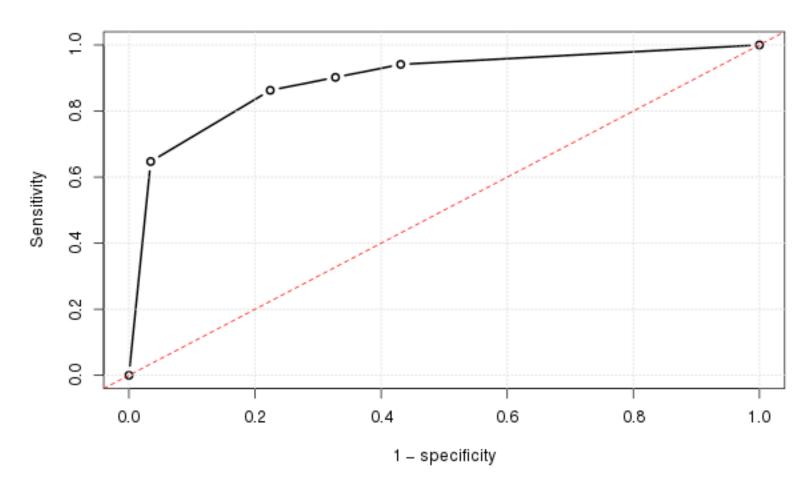
$$AUC = P(P(x_1) > P(x_2))$$



# 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

$$\frac{\sum I(P_{\mathbb{E}^{ ext{#}}}, P_{\oplus \mathbb{F}^{ ext{#}}})}{M*N}$$

$$I(P_{ ilde{ ilde{L}},P_{ ilde{ ilde{L}},P_{ ilde{ ilde{L}}}) = egin{cases} 1,P_{ ilde{ ilde{L}},P_{ ilde{ ilde{L}},P_{ ilde{L}},P_{ ilde{L}},P_{ ilde{L},P_{ ilde{L}},P_{ ilde{L}},P_{ ilde{L},P_{ ilde{L}},P_{ ilde{L}},P_{ ilde{L},P_{ ilde{L}},P_{ ilde{L}},P_{ ilde{L}},P_{ ilde{L}},P_{ ilde{L},P_{ ilde{L}},P_{ ilde{L}},P_{ ilde{L}},P_{ ilde{L}},P_{ ilde{L},P_{ ilde{L}},P_{ ilde{L}$$

### AUC: 计算AUC, 索引法

Sample	Label	ctr( <b>方法一</b> )	ctr( <b>方法二</b> )
А	0	0.25	0.25
В	0	0.35	0.35
С	1	0.30	0.35
D	1	0.40	0.40

$$I(P_{\text{E}|\text{$^{\perp}$}}, P_{\text{$\oplus$}|\text{$\downarrow$}}) = \begin{cases} 1, P_{\text{E}|\text{$\downarrow$}} > P_{\text{E}|\text{$\downarrow$}} \\ 0.5, P_{\text{E}|\text{$\downarrow$}} = P_{\text{$\oplus$}|\text{$\downarrow$}} \\ 0, P_{\text{E}|\text{$\downarrow$}} < P_{\text{$\oplus$}|\text{$\downarrow$}} \end{cases} \qquad \begin{cases} I_1(D, A) = 1 \\ I_1(D, B) = 1 \\ I_1(C, A) = 1 \\ I_1(C, A) = 1 \\ I_1(C, B) = 0 \end{cases} \qquad \begin{cases} I_2(D, A) = 1 \\ I_2(D, B) = 1 \\ I_2(C, A) = 1 \\ I_2(C, B) = 0.5 \end{cases}$$

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

### AUC: 计算AUC, 索引法

Sample	Label	ctr( <b>方法一</b> )	ctr( <b>方法二</b> )
А	0	0.25	0.25
В	0	0.35	0.35
С	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
А	0	0.25	1
В	0	0.35	3
С	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 $AUC_i$ 

