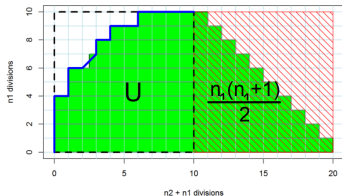


Introduction to Machine Learning

Evaluation: AUC & Mann-Whitney-U Test

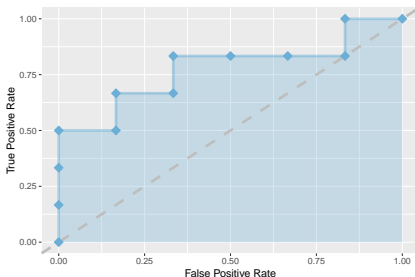


Learning goals

- Understand the rank-based nature of AUC
- See the connection between AUC and Mann-Whitney-U statistic

AUC AS A RANK-BASED METRIC

- The AUC metric is intimately related to the **Mann-Whitney-U test**, also known as **Wilcoxon rank-sum test**.
- This connection is best understood viewing the AUC from a slightly different angle: it is, in effect, a **rank-based** metric.
- Recall that, constructing the ROC curve, we count TP and FP.



- The AUC abstracts from the actual classification scores and considers only their rank.

AUC AS A RANK-BASED METRIC

- We can interpret the AUC as the probability of our classifier ranking a random positive observation higher than a random negative one.
- A perfect classifier will rank all positive above all negative observations, achieving $AUC = 1$.

| Truth | Score |
|-------|-------|
| 1 | 0.9 |
| 1 | 0.76 |
| 1 | 0.7 |
| 0 | 0.5 |
| 1 | 0.45 |
| 0 | 0.3 |
| 0 | 0.1 |

Choose a random positive



| | |
|---|------|
| 1 | 0.76 |
|---|------|

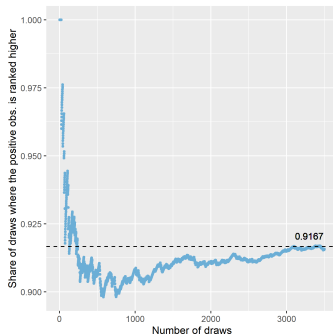
Choose a random negative



| | |
|---|-----|
| 0 | 0.3 |
|---|-----|

AUC = 0.9167

- Classifier ranks the positive higher than the negative
- This happens with a mean probability of 0.9167



MANN-WHITNEY-U TEST

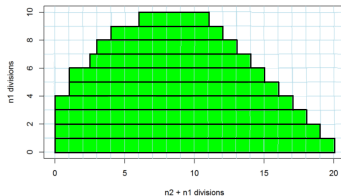
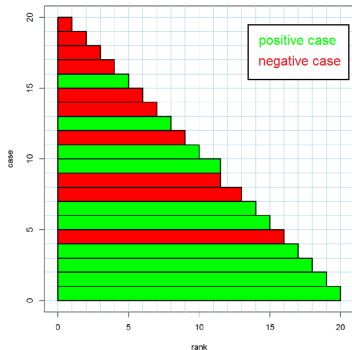
- The Mann-Whitney-U test is a **non-parametric hypothesis test** on the difference in location between two samples X_1 , X_2 of sizes n_1 and n_2 , respectively.
- Under the null, X_1 and X_2 follow the same (unknown) distribution \mathbb{P} , and for any pair of observations $x_{1,1} \in X_1$, $x_{2,1} \in X_2$ drawn at random from \mathbb{P} , the following statement holds: $\mathbb{P}(x_{1,1} \in X_1) > \mathbb{P}(x_{2,1} \in X_2) = \mathbb{P}(x_{1,1} \in X_1) < \mathbb{P}(x_{2,1} \in X_2) = 0.5$.
- The test statistic estimates the probability of a random sample from X_1 ranking higher than one from X_2 (R_1 denoting the sum of ranks of the $x_{1,i}$):

$$U = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \mathbb{I}[x_{1,i} > x_{2,j}] = R_1 - \frac{n_1(n_1 + 1)}{2}$$

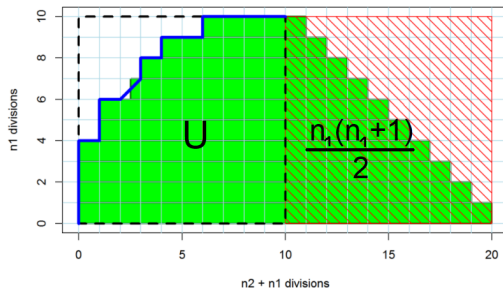
- For large samples, U is approximately normally distributed.

AUC & MANN-WHITNEY-U TEST

- We can directly interpret the AUC in the light of the U statistic.
- In order to see this, plot the ranks of all the scores as a stack of horizontal bars, and color them by label.
- Next, keep only the green ones, and slide them horizontally to get a nice even staircase on the right edge:



AUC & MANN-WHITNEY-U TEST

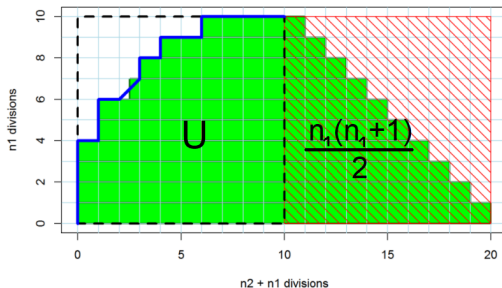


- Denoting the respective numbers of cases as n_+ and n_- , we have:

$$U = R_+ - \frac{n_+(n_+ + 1)}{2}.$$

- The area of the green bars on the right is equal to $\frac{n_+(n_+ + 1)}{2}$.

AUC & MANN-WHITNEY-U TEST



- U : area of the green bars on the left.
- $n_+ \cdot n_-$: area of the dashed rectangle.

⇒ AUC is U normalized to the unit square:

$$\text{AUC} = \frac{U}{n_+ \cdot n_-}.$$