

6 Receiver Operating Characteristic

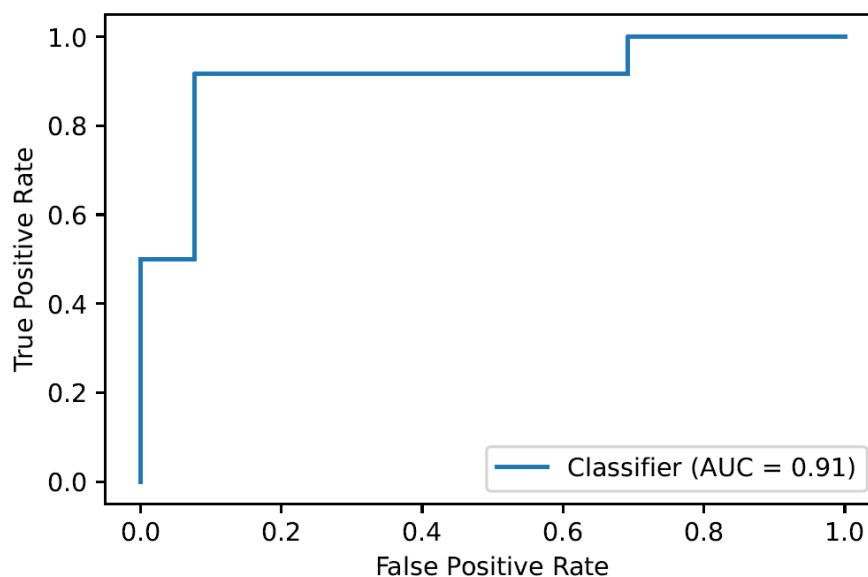
In a binary classification problem, a model generates a probability score for each instance to predict its class. This probability score is then compared to a classification threshold to assign the instance to either the positive or negative class. The classification threshold is a value between 0 and 1 that determines the boundary between the two classes. By default, the threshold is typically set to 0.5, which means that instances with a probability score above 0.5 are classified as the positive class, and those with a score below 0.5 are classified as the negative class. A ROC (Receiver Operating Characteristic) curve is a graphical representation of this relationship, plotting the true positive rate (TPR) against the false positive rate (FPR) at various classification thresholds. The True Positive Rate (TPR) is the proportion of actual positive cases that are correctly identified by the model:

$$\text{TPR} = \text{True Positives} / (\text{True Positives} + \text{False Negatives}).$$

The False Positive Rate (FPR) is the proportion of actual negative cases that are incorrectly identified as positive by the model:

$$\text{FPR} = \text{False Positives} / (\text{False Positives} + \text{True Negatives}).$$

An example of a ROC curve is depicted below:



The area under the ROC curve (AUC-ROC) is a popular metric used to evaluate the overall performance of a binary classifier, with higher AUC-ROC values indicating better classification performance. An ideal binary classifier would correspond to a unitary AUCROC, as, for each value of the False Positive Rate, the True Positive Rate would be always 1 (the maximum possible). Fortunately, we can change the GridSearchCV's optimizing goal: Instead of using the accuracy score (the default option), we can use the `roc_auc` score, in the expectation that, by thus proceeding, we are optimizing the recall score, by changing the level of the complexity of the tree (by trying different values for the parameter `ccp_alpha`).