1 Metrics

1.1 Receiver Operator Characteristic (ROC)

Consider a binary classification problem where we have a classifier that outputs a continuous score s(X) for an instance X. We then set a threshold τ to decide the predicted class. If $s(X) > \tau$, we predict the positive class; otherwise, we predict the negative class.

$$\begin{aligned} TPR(\tau) &= \mathbb{P}(s(X) > \tau | Y = 1) \\ FPR(\tau) &= \mathbb{P}(s(X) > \tau | Y = 0) \end{aligned}$$

Here, Y is the true class of an instance. TPR is the probability that the classifier ranks a randomly chosen positive instance higher than a randomly chosen negative instance. Similarly, FPR is the probability that the classifier ranks a randomly chosen negative instance higher than a randomly chosen positive instance.

The ROC curve plots $TPR(\tau)$ against $FPR(\tau)$ for all possible thresholds τ, producing a curve that ranges from $(0, 0)$ to $(1, 1)$. We can interpret a ROC plot as plotting the path of a function

$$f : \mathbb{R} \to [0, 1]^2, \quad \tau \mapsto (TPR(\tau), FPR(\tau))$$

The Area Under the ROC Curve (AUC) then provides a single scalar value that represents the expected performance of the classifier. An AUC of 1 indicates a perfect classifier, while an AUC of 0.5 indicates a classifier that performs no better than random chance.