

1 Definitions

Consider a space X that is divided into two classes. The true classification of $x \in X$ is determined by a true classification function y such that

$$y(x) = \begin{cases} 0 & x \in \text{Class Negative} \\ 1 & x \in \text{Class Positive} \end{cases}$$

We are frequently interested in learning a classifier function

$$h : X \rightarrow [0, 1]$$

such that a prediction of $y(\mathbf{x})$ can be made using the function

$$\hat{y}(\mathbf{x}, s) = \begin{cases} 0 & h(\mathbf{x}) \leq s \\ 1 & h(\mathbf{x}) > s \end{cases}$$

The discriminating power of a classification rule for a particular value of s can be measured in terms of the "True positive rate"

$$t(s) = P(\hat{y}(\mathbf{x}, s) = 1 | y(\mathbf{x}) = 1) \quad (1)$$

and the "False positive rate"

$$\begin{aligned} f(s) &= P(\hat{y}(\mathbf{x}, s) = 1 | y(\mathbf{x}) = 0) \\ &= P(h(\mathbf{x}) > s | y(\mathbf{x}) = 0) \end{aligned} \quad (2)$$

The performance of a classification function $h(\mathbf{x})$ is often visualized using the ROC curve, which plots true positive rate t as a function of false positive rate f . We will denote the ROC curve as a function by $\tau(f)$. A summary measure of a classifier's overall discrimination ability, independent of a particular threshold choice, is the Area under the curve, or AUC:

$$\text{AUC} = \int_0^1 \tau(u) du$$

Sometimes it can be more useful to look at a specific region of the ROC Curve rather than at the whole curve. For example, one could focus on the region of the curve with low false positive rate, which is often of prime interest for population screening tests.

$$\text{pAUC}(f_0) = \int_0^{f_0} \tau(u) du$$

2 Probabilistic Interpretation of AUC

The AUC is equal to the probability that a randomly chosen example from the positive class will have a higher value of h than a randomly chosen example from the negative class. We can see this as follows:

$$\begin{aligned} \text{AUC} &= \int_0^1 \tau u du \\ &= \int_0^1 P(h(\mathbf{x}) > f^{-1}(u) | y(\mathbf{x}) = 1) du \end{aligned}$$

Transform into an integral over the threshold s with the variable substitution $u = f(s)$.

$$\text{AUC} = \int_1^0 P(h(\mathbf{x}) > f^{-1}(f(s)) | y(\mathbf{x}) = 1) \frac{\partial f}{\partial s} ds$$

Because $f(s) = 1 - H(s)$ where H is the cumulative distribution function of $h(\mathbf{x})$, it follows that

$$\frac{\partial f}{\partial s} = -P(h(\mathbf{x}') = s | y(\mathbf{x}') = 0)$$

so that

$$\begin{aligned} \text{AUC} &= \int_0^1 P(h(\mathbf{x}) > s | y(\mathbf{x}) = 1) P(h(\mathbf{x}') = s | y(\mathbf{x}') = 0) ds \\ &= P(h(\mathbf{x}) > h(\mathbf{x}') | y(\mathbf{x}) = 1 \ \& \ y(\mathbf{x}') = 0) \end{aligned}$$

That is, as stated at the beginning of this section, AUC is equal to the probability that a randomly chosen example from the positive class will have a higher value of h than a randomly chosen example from the negative class.

3 Probabilistic Interpretation of pAUC

Given a fixed false positive rate threshold f_0 , the partial AUC is the area under the ROC curve in the region where false positive rates are less than f_0 .

$$\text{pAUC}(f_0) = \int_0^{f_0} \tau(u) du \quad (3)$$

This quantity is of interest when a classification test will drive a treatment for which high false positive rates are unacceptable, hence discrimination performance at lower thresholds where false positive rates would be higher is irrelevant and therefore should be ignored. A frequent example in the literature are medical screening tests, where false positives will drive expensive and potentially harmful medical treatments. Fraud identification is another example, as false positive rate above a low threshold can drive away legitimate customers unable to complete legitimate transactions.

We will see that the partial AUC is equal to f_0 times the probability that a randomly chosen example from the positive class will have a higher value of h than a randomly chosen example from the negative class *that would be a false positive* when using the decision threshold s_0 that yields a false positive rate f_0 .

Continue by substituting definition (??) into Equation (??), noting $s = f^{-1}(u)$

$$\begin{aligned} \text{pAUC}(f_0) &= \int_0^{f_0} P(h(\mathbf{x}) > f^{-1}(u) | y(\mathbf{x}) = 1) du \\ &= \int_{f^{-1}(0)}^{f^{-1}(f_0)} P(h(\mathbf{x}) > s | y(\mathbf{x}) = 1) \frac{\partial f}{\partial s} ds \\ &= \int_1^{s_0} -P(h(\mathbf{x}) > s | y(\mathbf{x}) = 1) P(h(\mathbf{x}') = s | y(\mathbf{x}') = 0) ds \\ &= \int_{s_0}^1 P(h(\mathbf{x}) > h(\mathbf{x}') \ \& \ h(\mathbf{x}') = s | y(\mathbf{x}) = 1 \ \& \ y(\mathbf{x}') = 0) ds \end{aligned}$$

Integrating over s gives

$$\begin{aligned}
 \text{pAUC}(f_0) &= P(h(\mathbf{x}) > h(\mathbf{x}') \ \& \ h(\mathbf{x}') > s_0 | y(\mathbf{x}) = 1 \ \& \ y(\mathbf{x}') = 0) \\
 &= P(h(\mathbf{x}') > s_0 | y(\mathbf{x}) = 1 \ \& \ y(\mathbf{x}') = 0) P(h(\mathbf{x}) > h(\mathbf{x}') | y(\mathbf{x}) = 1 \ \& \ y(\mathbf{x}') = 0 \ \& \ h(\mathbf{x}') > s_0) \\
 &= P(h(\mathbf{x}') > s_0 | y(\mathbf{x}') = 0) P(h(\mathbf{x}) > h(\mathbf{x}') | y(\mathbf{x}) = 1 \ \& \ y(\mathbf{x}') = 0 \ \& \ h(\mathbf{x}') > s_0) \\
 &= f_0 P(h(\mathbf{x}) > h(\mathbf{x}') | y(\mathbf{x}) = 1 \ \& \ y(\mathbf{x}') = 0 \ \& \ h(\mathbf{x}') > s_0)
 \end{aligned}$$

using the definition of conditional probability.

As stated above, we have shown that the partial AUC is equal to f_0 times the probability that a randomly chosen example from the positive class will have a higher value of h than a randomly chosen example from the negative class *that would be a false positive* when using the decision threshold s_0 that yields a false positive rate f_0 .