1 Criteria of Learning Performance of A Classier

Suppose that we are given a data set $D = \{(x_1, a_1), (x_2, a_2), \ldots, (x_m, a_m)\}$ such that $x_i \in \mathbb{R}^K$ and $a_i \in \{0, 1\}$, $i = 1, 2, \ldots, m$, where we call a_i the *class of* x_i. A feature vector x_i with $a_i = 1$ (resp., 0) is called *positive* (resp., negative). We call a function $\eta : \mathbb{R}^K \to \{0, 1\}$ that estimates the class of a feature vector a *classifier*.

There are various criteria to evaluate the learning performance of a classifier. We summarize some of them.

1.1 Accuracy Based Criteria

Let us partition the data set D into $D = D_1 \cup D_0$, where D_1 (resp., D_0) is the set of all positive (resp., negative) feature vectors in D. For a classifier η, a feature vector $x_i \in D$ is called

- *true positive* if $a_i = 1$ and $\eta(x_i) = 1$;
- *true negative* if $a_i = 0$ and $\eta(x_i) = 0$;
- *false positive* if $a_i = 0$ and $\eta(x_i) = 1$; and
- *false negative* if $a_i = 1$ and $\eta(x_i) = 0$.

We denote by $\mathrm{TP}(\eta; D)/\mathrm{TN}(\eta; D)/\mathrm{FP}(\eta; D)/\mathrm{FN}(\eta; D)$ the sets of true positive/true negative/false positive/false negative feature vectors, respectively. We define $\mathrm{TPR}(\eta; D)$, $\mathrm{TNR}(\eta; D)$, $\mathrm{FPR}(\eta; D)$ and $\mathrm{FNR}(\eta; D)$ as follows;

$$\mathrm{TPR}(\eta; D) \triangleq \frac{\mathrm{TP}(\eta; D)}{|D_1|}; \quad \mathrm{TNR}(\eta; D) \triangleq \frac{\mathrm{TN}(\eta; D)}{|D_0|};$$
$$\mathrm{FPR}(\eta; D) \triangleq \frac{\mathrm{FP}(\eta; D)}{|D_0|}; \quad \mathrm{FNR}(\eta; D) \triangleq \frac{\mathrm{FN}(\eta; D)}{|D_1|}.$$

It holds that $\mathrm{TPR}(\eta; D) + \mathrm{FNR}(\eta; D) = \mathrm{TNR}(\eta; D) + \mathrm{FPR}(\eta; D) = 1$. The *accuracy* $\mathrm{ACC}(\eta; D)$ is defined to be:

$$\mathrm{ACC}(\eta; D) \triangleq \frac{\mathrm{TP}(\eta; D) + \mathrm{TN}(\eta; D)}{|D|}.$$

The *balanced accuracy* $\text{B-ACC}(\eta; D)$ is defined to be:

$$\text{B-ACC}(\eta; D) \triangleq \frac{1}{2}(\mathrm{TPR}(\eta; D) + \mathrm{TNR}(\eta; D)).$$

1.2 ROC Curve and AUC

Let $f : \mathbb{R}^K \to \mathbb{R}$ be a function and $\theta \in \mathbb{R}$ be a real number. We construct a classifier $\eta_{f,\theta} : \mathbb{R}^K \to \{0, 1\}$ as follows; for $x \in \mathbb{R}^K$,

$$\eta_{f,\theta}(x) \triangleq \begin{cases} 1 & \text{if } f(x) \geq \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Suppose that $x_1, x_2, \ldots, x_m$ are sorted so that $f(x_1) \leq f(x_2) \leq \cdots \leq f(x_m)$ holds. For $i = 1, \ldots, m-1$, let us define

$$\theta_i \triangleq \frac{f(x_i) + f(x_{i+1})}{2},$$

where we let $\theta_0 \triangleq f(x_1) - \varepsilon$ and $\theta_m \triangleq f(x_m) + \varepsilon$ for a small constant $\varepsilon \in \mathbb{R}_+ \setminus \{0\}$.

We have $m + 1$ classifiers $\eta_{f,\theta_0}, \eta_{f,\theta_1}, \ldots, \eta_{f,\theta_m}$. Let us denote a 2D point $p_i \triangleq (\mathrm{FPR}(\eta_{f,\theta_i}; D), \mathrm{TPR}(\eta_{f,\theta_i}; D))$, $i = 0, 1, \ldots, m$. Observe that $p_0 = (1, 1)$ holds since $\eta_{f,\theta_0}(x) = 1$ holds for all $x \in D$, and that $p_m = (0, 0)$ holds since $\eta_{f,\theta_m}(x) = 0$ holds for all $x \in D$. Also we have;

$$\begin{aligned} \mathrm{FPR}(\eta_{f,\theta_m}; D) = 0 \leq \mathrm{FPR}(\eta_{f,\theta_{m-1}}; D) \leq \cdots \leq \mathrm{FPR}(\eta_{f,\theta_0}; D) = 1; \\ \mathrm{TPR}(\eta_{f,\theta_m}; D) = 0 \leq \mathrm{TPR}(\eta_{f,\theta_{m-1}}; D) \leq \cdots \leq \mathrm{TPR}(\eta_{f,\theta_0}; D) = 1. \end{aligned}$$

The *Receiver Operating Characteristic curve* (*ROC curve*) *of* f is a set of m line segments $(p_m = (0, 0), p_{m-1}), (p_{m-1}, p_{m-2}), \ldots, (p_1, p_0 = (1, 1))$. The *Area Under Curve* (*AUC*) *of* f, which we denote by $\mathrm{AUC}(f; D)$, is defined to be the area between the ROC curve and the x-axis. Hence we have $0 \leq \mathrm{AUC}(f; D) \leq 1$.

Theorem:
When $f : R^K - > \{0, 1\}$, $AUC(f; D) = BACC(\eta; D)$