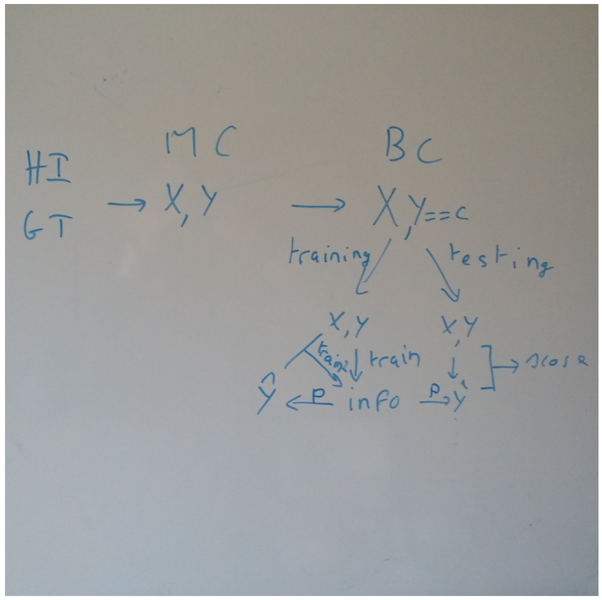

Figure 1: Linear Classifier for a Binary Classification Problem

Content4
ROC curves and linear classifiers for binary classifications

1 Keywords

- Area Under Curve

$$\text{AUC} \approx \sum_k \frac{\left(\text{TPR}_{\lambda_k} + \text{TPR}_{\lambda_{k+1}}\right)}{2} \left(\text{FPR}_{\lambda_{k+1}} - \text{FPR}_{\lambda_k}\right) \tag{1}$$

- Deriving a linear classifier from the mean square error.

$$\mathcal{L}\left((y_n)_n, (\sum_f a_f x_{n,f} + b)_n\right) = \frac{1}{N}(XA + b - Y)^T(XA + b - Y) \tag{2}$$

Let $X^e = [X\ \mathbf{1}]$. Finding A and b is finding $A^e = \begin{bmatrix} A \\ b \end{bmatrix}$ as $X^e A^e = AX + b$.

$$\mathcal{L}\left((y_n)_n, (\sum_f a_f x_{n,f} + b)_n\right) = \frac{1}{N}(X^e A^e - Y)^T(X^e A^e - Y) \tag{3}$$

Let $\widehat{A^e} = X^{e+} Y$

$$\mathcal{L}\left((y_n)_n, (\sum_f a_f x_{n,f} + b)_n\right) \geq \frac{1}{N}(X^e \widehat{A^e} - Y)^T(X^e \widehat{A^e} - Y) \tag{4}$$

Let $\begin{bmatrix} \widehat{A} \\ \widehat{b} \end{bmatrix} = \widehat{A^e}$

$$\begin{bmatrix} \widehat{A} \\ \widehat{b} \end{bmatrix} = [X\ \mathbf{1}]^+ Y \tag{5}$$

Most often $\begin{bmatrix} \widehat{A} \\ \widehat{b} \end{bmatrix} = ([X\ \mathbf{1}]^T[X\ \mathbf{1}])^{-1}[X\ \mathbf{1}]^T Y$

- Linear Classifier of a binary classification problem

$$h(\mathbf{x}) = \mathbf{1}\left(\sum_f a_f x_f + b > 0\right) \tag{6}$$

where x_f are the component of x as a row-vector.

- Intercept

$$b \text{ in } \mathbf{1}(\sum_f a_f x_f + b > 0) \tag{7}$$

- Loss function Given binary classifying problem and a training set (X, Y), the parameter values of the linear classifier are defined as

$$a_f, b = \underset{(a_f)_f, b}{\arg} \min \mathcal{L}\left((y_n)_n, (\sum_f a_f x_{n,f} + b)_n\right) \tag{8}$$

- Loss function derived from overall accuracy

$$\mathcal{L}\left((y_n)_n, (\sum_f a_f x_{n,f} + b)_n\right) = \frac{1}{N}\sum_n \left(y_n - \mathbf{1}(\sum_f a_f x_{n,f} + b)\right)^2 \tag{9}$$

- Mean square error

$$\mathcal{L}\left((y_n)_n, (\sum_f a_f x_{n,f} + b)_n\right) = \frac{1}{N}\sum_n \left(y_n - \sum_f a_f x_{n,f} - b\right)^2 \tag{10}$$

- ROC curve:

$$(\text{FPR}_\lambda, \text{TPR}_\lambda)_\lambda \tag{11}$$

- Pseudo-inverse (`pinv` in Matlab)

$$(AZ - B)^T(AZ - B) \leq (AX - B)^T(AX - B) \text{ when } Z = A^+B \tag{12}$$

If A^TA is invertible

$$A^+ = (A^TA)^{-1}A^T \tag{13}$$

If AA^T is invertible

$$A^+ = A^T(AA^T)^{-1} \tag{14}$$

If $A = U\Sigma V^T$ (singular value decomposition) then

$$A^+ = V\Sigma^+U^T \tag{15}$$

where Σ^+ is also a diagonal matrix having their non-zero components inversed.

- Weight

$$a_f \text{ in } \mathbf{1}(\sum_f a_f x_f + b > 0) \tag{16}$$