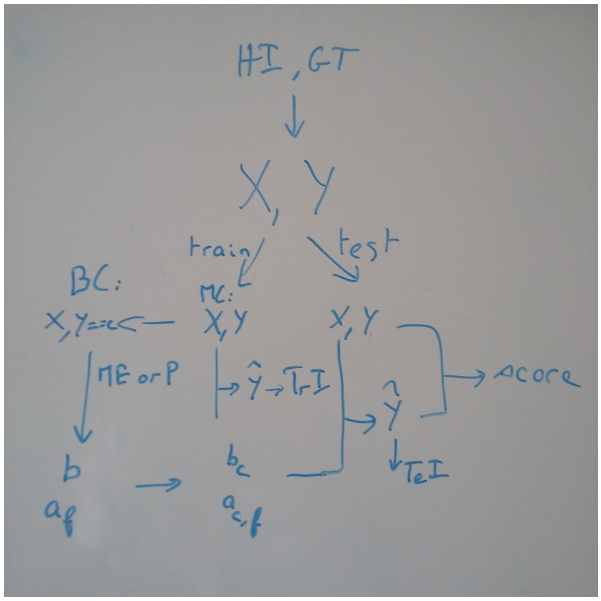

Figure 1: Linear Classifier

Content5
Linear Classifiers and Loss Functions

1 Keywords

- One-Versus-All (OVA)

$$h(\mathbf{x}) = \underset{c\in\{1,\ldots,C\}}{\arg\max} \sum_f a_{c,f} x_f + b_c \tag{1}$$

where $h_c(\mathbf{x}) = \mathbf{1}(\sum_f a_{c,f} x_f + b_c > 0)$

- Perceptron method for a binary classifying problem
Deriving a linear classifier from a OA loss function.

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

where $\widehat{Y} = [\mathbf{1}(XA + b > 0)]$ Let $X^e = [X\ \mathbf{1}]$. Finding A and b is finding $A^e = \begin{bmatrix} A \\ b \end{bmatrix}$ as $X^e A^e = XA + b$.

$$\begin{cases} A_0^e &= 0 \\ A_{t+1}^e &= A_t^e - rX^{eT}(\widehat{Y}_t - Y) \end{cases} \tag{3}$$

where $\widehat{Y}_t = [\mathbf{1}(X^e A_t^e > 0)]$ and $r > 0$ is the learning factor.