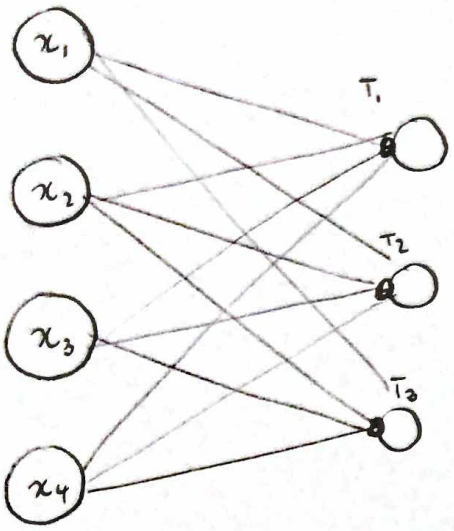
LOSS FUNCTION / ARC-FACE

SOFT-MAX



$$T_1 = W_1^T \underline{x} + b_1 \qquad y_1 = \frac{e^{T_1}}{e^{T_1} + e^{T_2} + e^{T_3}}$$

$$T_2 = W_2^T x + b_2 \qquad \vdots$$

$$T_3 = W_3^T x + b_3 \qquad y_3 = \frac{e^{T_3}}{e^{T_1} + e^{T_2} + e^{T_3}}$$

$$\underline{X} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

CLASSIFICATION:

$$y(\underline{x}) = \text{argmax}(y_i) \qquad j = 1 \dots 3$$

LOSS-FUNCTION (para N muestras)

$$L_1 = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{W_{y_i}^T \underline{x}_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T \underline{x}_i + b_j}}$$

función de pérdida clásica

donde $y_i = y(\underline{x}_i)$

LOSS-FUNCTION PARA ARCFACE

→ se asume $b_j = 0$

$W_j^T \underline{x}_i = \|W_j\| \|x_i\| \cos\theta_j$ ← producto punto

→ se asume $\|W_j\| = 1$ (l_2 normalization) $W_j := \frac{W_j}{\|W_j\|}$

→ se asume $\|x_i\| = s$ (l_2 normalization) $\underline{x}_i := s \cdot \frac{\underline{x}_i}{\|\underline{x}_i\|}$

Entonces:

$$L_2 = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{s\cos\theta_{y_i}}}{\sum_{j=1}^{n} e^{s\cos\theta_j}} = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{s\cos\theta_{y_i}}}{e^{s\cos\theta_{y_i}} + \sum_{j \neq y_i} e^{s\cos\theta_j}}$$

Para aumentar simultaneamente la "compactness" intra clase y

la "discrepancy" inter clase

se usa un margen "m" de penalidad:

$$L_3 = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))} + \sum_{j \neq y_i} e^{s\cos\theta_j}}$$