Discriminant Training

Maximum A Posteriori

- posterior probability of a model

$$P(M_j|X) = \frac{P(X|M_j)P(M_j)}{P(X)}$$

- It can be rewritten as

$$\begin{aligned} P(M_j|X) &= \frac{P(X|M_j)P(M_j)}{\sum_k P(X|M_k)P(M_k)} \\ &= \frac{1}{1 + \sum_{k \neq j} \frac{P(X|M_k)P(M_k)}{P(X|M_j)P(M_j)}} \end{aligned}$$

- Increasing $P(X|M_j)$ does not guarantee the increase of $P(M_j|X)$.

Maximum Mutual Information

- mutual information

$$I(M, X|\Theta) = E\left[\log \frac{P(M, X|\Theta)}{P(M|\Theta)P(X|\Theta)}\right]$$

- The information of a model M_j provided by acoustics X

$$\begin{aligned} I(M_j, X|\Theta) &= \log \frac{P(M_j, X|\Theta)}{P(M_j|\Theta)P(X|\Theta)} \\ &= \log \frac{P(X|M_j, \Theta)}{\sum_k P(X|M_k, \Theta)P(M_k|\Theta)} \end{aligned}$$

Corrective Training

- Corrects the parameters for the utterances in which the correct models have a lower likelihood than the best models.
- If

  $$P(X|M_r, \Theta) \geq P(X|M_c, \Theta) + \Delta$$

  then

  $$\Theta \rightarrow \Theta'$$

  such that

  $$P(X|M_c, \Theta') \geq P(X|M_c, \Theta) \text{ and } P(X|M_r, \Theta') \geq P(X|M_r, \Theta)$$

Generalized Probabilistic Descent

- discriminant functions

$$g_j(X;\Theta) = \log P(X|M_j,\Theta)$$

- classification rule

$$j^* = \arg\max_j \; g_j(X;\Theta)$$

- mapping discriminant functions to a loss function

$$\text{define } d_j(X;\Theta) = \log\left\{\frac{1}{K-1}\sum_{k\neq j} e^{\eta g_k(X;\Theta)}\right\} - g_j(X;\Theta)$$

$$\text{and minimize } E(\Theta) = \sum_j \sum_{X\in M_j} F(d_j(X;\Theta)),$$

where $F(x)$ is a function ~ 0 when x is negative and ~ 1 when x is positive.