

Minimum Entropy Regularization

①

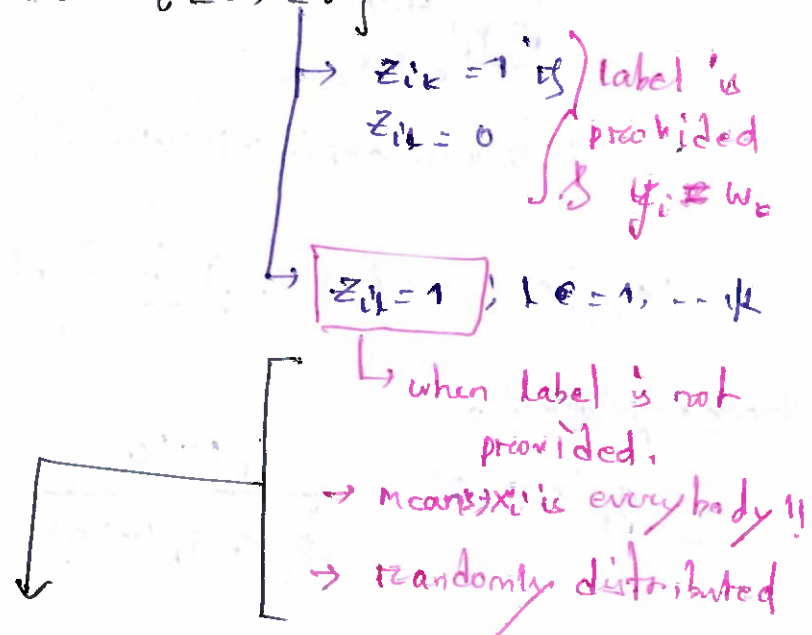
② minimum Entropy regularization

original Data $L_n = \{x_i, y_i\}$ $x_i \in X$ Data input
 $y_i \in \{w_1, \dots, w_k\}$ \rightarrow target output

Some of the label is missing !!

Criterion Derivation:

new learning set $L_n = \{x_i, z_i\}$



for an unlabeled case $P(z | x_i, w_k) = P(z | x_i, w_l)$
 $\forall (w_k, w_l)$

$$\therefore \text{Now, } P(w_k | x, z) = \frac{z_k P(w_k | x)}{\sum_{l=1}^k z_l P(w_l | x)}$$

conditional log likelihood

$$L(\theta, L_n) = \sum_{i=1}^n \log \left(\sum_{k=1}^K z_{ik} f_k(x_i; \theta) \right) + h(z_i)$$

model for $P(w_k | x)$
 \uparrow # parameters (concave)
 independent of $P(x, y)$

⑦

when unlabeled data is informative:

Conditional entropy unknown about $y|x, z$ is known

$$H(y|x, z) = - \mathbb{E}_{x,y,z} \log [P(y|x, z)]$$

maximum entropy in θ :

$$\mathbb{E}_{\theta, \psi} [H(y|x, z)] = c$$

$\underbrace{\theta, \psi}_{\text{model params.}} \rightarrow \text{log-trick}$

$$p(\theta, \psi) \propto \exp(-\lambda H(y|x, z)) // \text{prior on } \theta$$

$$H_{\text{emp}}(y|x, z; \mathcal{L}_n) = -\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K P(w_k|x_i, z_i) \log P(w_k|x_i, z_i)$$

Entropy Regularization:

$$g_k(x, z; \theta) = \frac{z_k f_k(x; \theta)}{\sum_{k=1}^K z_k f_k(x; \theta)}$$

→ labeled case $g_k(x, z; \theta) = z_k$

→ model for $P(w_k|x, z)$

→ unlabeled case $g_k(x, z; \theta) = f_k(x; \theta)$

So the maximizer $\sim P(\theta, \psi) \log P(w_k|x, z) // \text{conditional.}$

$$c(\theta, \lambda; \mathcal{L}_n) = L(\theta; \mathcal{L}_n) - \lambda H_{\text{emp}}(y|x, z; \mathcal{L}_n)$$

$$= \underbrace{\sum_{i=1}^n \log \left(\sum_{k=1}^K z_{ik} f_k(x_i) \right)}_{\text{labeled data}} + \lambda \sum_{i=1}^n \sum_{k=1}^K \underbrace{g_k(x_i, z_i) \log g_k(x_i, z_i)}_{\text{unlabeled data}}$$

labeled data

unlabeled data