② minimum Entropy regularization

original Data $\mathcal{L}_n = \{ x_i, y_i \}$ — y_i → target output; $x_i \in \mathcal{X}$ Data input

$y_i \in \{ w_1, \dots, w_K \}$

Some of the label is missing !!

<u>Criterion Derivation:</u>

new learning Set $\mathcal{L}_n = \{ x_i, z_i \}$

→ $z_{ik} = 1$ if label is provided & $y_i = w_k$
$z_{il} = 0$

→ $\boxed{z_{il} = 1}$; $l = 1, \dots, K$

→ when label is not provided.

→ means x_i is everybody !!

→ randomly distributed

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

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

conditional log likelihood

model for $P(w_k \mid x)$
↑ θ parameters (concave)

$$L(\theta, \mathcal{L}_n) = \sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} z_{ik} f_k(x_i ; \theta) \right) + \underbrace{h(z_i)}_{\text{independent of } P(x,y)}$$

when unlabeled data is informative:

Conditional Entropy unknown about $y \mid x, \& z$ is known

$$H(y \mid x, z) = -E_{xyz} \log [P(y \mid x, z)]$$

maximum Entropy in Avg:

$$\underbrace{E_{\theta, \psi}}_{\text{model params.}} [H(y \mid x, z)] = c$$

Lagrangian

$$\Rightarrow \quad P(\theta; \psi) \propto \exp(-\lambda H(y \mid x, z)) \quad // \text{prior on } \theta$$

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

Entropy Regularization:

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

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

model for $P(\omega_k \mid x, z)$

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

So the maximizer $\sim P(\theta, \psi)$ ~~log~~ $P(\omega_k \mid x, z)$ // conditional.

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

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