

$$x_i \begin{cases} \rightarrow u_{0i} \triangleq w_0^T x_i + d_{0i} \\ \rightarrow u_{1i} \triangleq w_1^T x_i + d_{1i} \end{cases}$$

DATE _____
PAGE _____

$$y_i = \mathbb{I}(u_{1i} > u_{0i})$$

$$\text{define } z_i = u_{1i} - u_{0i} + \epsilon_i \quad (\epsilon_i \sim \mathcal{N}(0, 1))$$

$$z_i \triangleq w^T x_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, 1)$$

$$y_i = 1 = \mathbb{I}(z_i \geq 0)$$

$$\begin{aligned} P(y_i = 1 | x_i, w) &= \int \mathbb{I}(z_i \geq 0) \mathcal{N}(z_i | w^T x_i, 1) dz_i \\ &= P(w^T x_i + \epsilon \geq 0) = P(\epsilon \geq -w^T x_i) \\ &= 1 - \Phi(-w^T x_i) = \underline{\Phi(w^T x_i)} \end{aligned}$$

this model can be trained using EM. $P(y_i = 1 | z_i) = \mathbb{I}(z_i \geq 0)$
 $z_i \sim \mathcal{N}(w^T x_i, 1)$ is latent var.

$$\text{complete log likelihood} \quad \ell(z, w | v_0) = \ln(P(y|z)) + \ln(z | x, w, \mathbb{I}) + \ln(w | 0, v_0)$$

$$= \sum_i \ln(P(y_i | z_i)) - \frac{1}{2} (z - xw)^T (z - xw) - \frac{1}{2} w^T V_0^{-1} w$$

①. the posterior of E-step is a truncated gaussian + const.

$$P(z_i | y_i, x_i, w) = \begin{cases} \mathcal{N}(z_i | w^T x_i, 1) \mathbb{I}(z_i \geq 0) & \text{if } y_i = 1 \\ \mathcal{N}(z_i | w^T x_i, 1) \mathbb{I}(z_i < 0) & \text{if } y_i = 0 \end{cases}$$

note $l(z, w | v_0)$ is linearly dependent on z
 so $E[z_i | y_i, w, x_i] = \begin{cases} \mu_i + \frac{\phi(\mu_i)}{1 - \phi(-\mu_i)} & y_i = 1 \\ \mu_i - \frac{\phi(\mu_i)}{\phi(-\mu_i)} & y_i = 0 \end{cases}$

In the M-step estimate w using ridge regression, where $\mu = E[z]$ is the output we wish to predict.

$$\hat{w} = (V_0^{-1} + X^T X)^{-1} X^T \mu$$

$$(X^T w - \sum y_i) = (X^T w - \sum y_i) = 0$$

For $\mu = E[z]$, we have $\mu = \frac{1}{n} \sum_{i=1}^n z_i$

$$E[z_i] = \mu = \frac{1}{n} \sum_{i=1}^n z_i$$

$$(w^T x - \mu)^T (w^T x - \mu) = \sum_{i=1}^n (w^T x_i - \mu)^2$$

Now, we can write the ridge regression problem as

$$\min_w \sum_{i=1}^n (w^T x_i - \mu)^2 + \lambda \|w\|^2$$

$$\min_w \sum_{i=1}^n (w^T x_i - \mu)^2 + \lambda \|w\|^2$$