hw3.

1.

a.

$$\Rightarrow P(\beta|\mathbb{X},\mathbb{Y}) = \frac{P(\mathbb{Y}|\mathbb{X},\beta)P(\mathbb{X})P(\beta)}{P(\mathbb{X}|\mathbb{Y}) * P(\mathbb{Y})}$$

$$\propto P(\mathbb{Y}|\mathbb{X},\beta) * P(\beta)$$

$$\Rightarrow f_x(x) = 1/(\sqrt{2\pi}^n \sqrt{\det(\Sigma)}) * \exp(-\tfrac{1}{2}(\mathbb{X}-\mu)^T \Sigma^{-1} (\mathbb{X}_i-\mu_i)$$

$\Rightarrow \det(\Sigma) = \sigma_\beta$ for β & σ for ϵ

$\Rightarrow \Sigma^{-1} = \frac{1}{\det(\Sigma)}$ & $z^T \cdot z = \|z\|_2^2$ $\longleftarrow$ diagonal matrix

$$\Rightarrow P(\beta|\mathbb{X},\mathbb{Y}) = (2\pi)^{-d/2} \sigma_\beta^{-d} \exp(-\tfrac{1}{2} * \tfrac{1}{\sigma_\beta^2} \|\beta\|_2^2)$$
$$*$$
$$(2\pi)^{-n/2} \sigma^{-n} \exp(-\tfrac{1}{2} * \tfrac{1}{\sigma} \|X\beta - Y\|_2^2).$$

$$\Rightarrow P(\beta|\mathbb{X},\mathbb{Y})$$
$$\propto_\beta (1/(2\pi)^{(n+d)/2}) * \sigma_\beta^{-d} \sigma^{-n} * \exp(\tfrac{1}{-2\sigma^2}\|X\beta-Y\|_2^2 \tfrac{1}{-2\sigma_\beta^2}\|\beta\|_2^2).$$

b.

$$\Rightarrow \underset{\beta}{\text{argmax}}\, P(\beta|\mathbb{X},\mathbb{Y}) = \underset{\beta}{\text{argmax}} \log[P(\beta|\mathbb{X},\mathbb{Y}))$$

$$= -\log((2\pi)^{(n+d)/2} \sigma_\beta^n \sigma^d - \tfrac{1}{2\sigma^2}\|\mathbb{X}\beta - \mathbb{Y}\|_2^2 - \tfrac{1}{2\sigma_\beta^2}\|\beta\|_2^2$$

$\Rightarrow$ we can rule out the term that don't contain β

$$\Rightarrow \underset{\beta}{\text{argmax}}\, f(x)$$

$$= \underset{\beta}{\operatorname{argmax}} -\frac{1}{2\sigma^2}\|X\beta - Y\|_2^2 - \frac{1}{2\sigma_\beta^2}\|\beta\|_2^2$$

$$= \underset{\beta}{\operatorname{argmin}} \left[\frac{1}{2\sigma^2}\|X\beta - Y\|_2^2 + \frac{1}{2\sigma_\beta^2}\|\beta\|_2^2\right] * 2\sigma^2 \longleftarrow$$ multiplying by a constant won't change.

$$= \underset{\beta}{\operatorname{argmin}} \left[\|X\beta - Y\|_2^2 + \frac{\sigma^2}{\sigma_\beta^2}\|\beta\|_2^2\right]$$

$\Rightarrow$ thus, $== \underset{\beta}{\operatorname{argmin}} \|X\beta - Y\|_2^2 + \lambda\|\beta\|_2^2$ for $\lambda = \frac{\sigma^2}{\sigma_\beta^2}$

$\Rightarrow$ thus,

$$\hat{\beta}_{MAP} := \underset{\beta}{\operatorname{argmax}} P(\beta | X, y) == \underset{\beta}{\operatorname{argmin}} \|X\beta - y\|_2^2 + \lambda\|\beta\|_2^2 \text{ for } \lambda = \frac{\sigma^2}{\sigma_\beta^2}.$$

c.

$\Rightarrow$ L2 regulation $= (Y - X\beta) + \lambda \sum_{i=1}^{d} (\beta_i)^2$

$\Rightarrow \hat{\beta}_{MAP} = \underset{\beta}{\operatorname{argmin}}$ (L2 regulation)

$\Rightarrow$ as $\lambda \uparrow \rightarrow$ the penalty will go up

so in order to reach the minimize β need to $\downarrow$

$\Rightarrow$ thus, the smaller σ_β^2 will decrease β

while bigger σ_β^2 will increase β.

2. $\rightarrow X =$ poisson dist.

$\rightarrow$ latent layer = sunny, rainy, snowing. $(\pi_1 ; \pi_2 ; \pi_3)$.

$\rightarrow z_i \sim$ multinominal $(1, (\pi_1, \pi_2, \pi_3))$.

$X_i \sim$ poisson (λ_{z_i}).

→ hyperparameter

$$\begin{tikzcd}
 & & \pi, \lambda \arrow[dll] \arrow[dl] \arrow[d] \arrow[dr] & \\
z_1 \arrow[d] & z_2 \arrow[d] & z_3 \arrow[d] & z_4 \arrow[d] \\
x_1 & x_2 & x_3 & x_4
\end{tikzcd}$$

a.

$$\Rightarrow \mathbb{P}(x_i \mid z_i = j, \lambda) = e^{-\lambda_j} * \frac{(\lambda_j)^{x_i}}{(x_i)!}$$

b.

$$\Rightarrow \mathbb{P}(x_i, z_i = j \mid \pi, \lambda) = \mathbb{P}(x_i \mid z_i = j, \pi, \lambda) * \mathbb{P}(z_i = j \mid \pi, \lambda).$$

$$= e^{-\lambda_j} * \frac{(\lambda_j)^{x_i}}{(x_i)!} * \pi_j$$

c.

$$\Rightarrow \mathbb{P}(x \mid \lambda, \pi) = \prod_{i=1}^{n} P(x_i \mid \lambda, \pi).$$

$$= \prod_{i=1}^{n} \left[\sum_{z_i} P(x_i, z_i \mid \lambda, \pi) \right]$$

$$= \prod_{i=1}^{n} \left[\sum_{j=1}^{3} \left(e^{-\lambda_j} * \frac{(\lambda_j)^{x_i}}{(x_i)!} * \pi_j \right) \right]$$

d.

(i).

$$\Rightarrow \mathbb{P}(z_i = j \mid x_i, \lambda^{(t)}, \pi^{(t)})$$

$$= \left[\mathbb{P}(X_i \mid z_i = j, \lambda^{(t)}, \pi^{(t)})\, \mathbb{P}(z_i = j \mid \lambda^{(t)}, \pi^{(t)})\right] / \mathbb{P}(X_i \mid \lambda^{(t)}, \pi^{(t)})$$

$$\Rightarrow q_i^{(t)}(z_i) = \left[e^{-\lambda_j^{(t)}} * \frac{(\lambda_j^{(t)})^{x_i}}{(x_i)!} * \pi_j^{(t)}\right] \Big/ \sum_{j=1}^{3} \left(e^{-\lambda_j} * \frac{(\lambda_j)^{x_i}}{(x_i)!} * \pi_j\right)$$

(ii).

$\Rightarrow$ set λ & π as hyperparameter θ

$$\Rightarrow q^{(t)}(\underset{\sim}{z}) = \prod_{i=1}^{n} q^{(t)}(z_i).$$

$$\Rightarrow \mathbb{P}(z \mid X, \lambda^{(t)}, \pi^{(t)}) \propto \mathbb{P}(z_1, z_2 \ldots X_n \mid X_1 \ldots, X_n, \theta)$$

$$\propto \mathbb{P}(z_1, z_2 \ldots, z_n \mid \theta)\, \mathbb{P}(X_1, X_2 \ldots X_n \mid z_1, z_2 \ldots z_n, \theta)$$

$$= \prod_{i=1}^{n} \mathbb{P}(z_i \mid \theta) \prod_{i=1}^{n} \mathbb{P}(X_i \mid z_i, \theta)$$

$$= \prod_{i=1}^{n} q_i^{(t)}(z_i).$$

e.

(i).

$$\Rightarrow \sum_{i=1}^{n} \mathbb{E}_{z_i \sim q_i(t)} \left[\log\left(\pi_{z_i} \frac{\lambda_{z_i}^{x_i}}{x_i!} \exp(-\lambda_{z_i})\right]\right.$$

$$= \mathbb{E}_{z_i \sim q_i(t)} \left[\mathbb{1}(z_i = j)\right] * q_i^{(t)}(z_i = j).$$

$$= \mathbb{E}_{z_i \sim q_i(t)} \left[\mathbb{1}(z_i = j)\right] * \left[\cancel{\log(\pi_{z_i})} + x_i(\log(\lambda_{z_i})) - \cancel{\log(x_i!)} - \lambda_{z_i}\right]$$

$\Rightarrow$ since we need to set $\dfrac{\partial f(\pi, \lambda)}{\theta \lambda_j} = 0$

$\Rightarrow$ derivative with related to λ_j.

$$= \mathbb{E}_{z_i \sim q_i(t)}\left[\mathbb{1}(z_i = j)\right] * \left[\frac{x_i}{\lambda_{z_i}} - 1\right]$$

$\Rightarrow$ then we substitute backward $\mathbb{E}\left[\mathbb{1}(z_i = j)\right] = q_i^{(t)}(z_i = j)$

$$\Rightarrow \sum_{i=1}^{n} q_i^{(t)}(z_i = j)\left[\frac{x_i}{\lambda_{z_i}} - 1\right] = 0.$$

$$\Rightarrow \sum_{i=1}^{n} q_i^{(t)}(z_i = j)\frac{x_i}{\lambda_{z_i}} - \sum_{i=1}^{n} q_i^{(t)}(z_i = j) = 0.$$

$\Rightarrow$ since we already set $z_i = j$ & $\lambda_{z_i} = \lambda_j$

$$\Rightarrow \text{derivative} = \sum_{i=1}^{n} q_i^{(t)}(z_i = j)\frac{x_i}{\lambda_j} - \sum_{i=1}^{n} q_i^{(t)}(z_i = j) = 0.$$

$$\Rightarrow \frac{1}{\lambda_j} \sum_{i=1}^{n} q_i^{(t)}(z_i = j)\, x_i - \sum_{i=1}^{n} q_i^{(t)}(z_i = j) = 0$$

$$\Rightarrow \lambda_j = \frac{\sum_{i=1}^{n} q_i^{(t)}(z_i = j)\, x_i}{\sum_{i=1}^{n} q_i^{(t)}(z_i = j)} \leftarrow \text{argmax}$$

$\Rightarrow$ thus,

$$\lambda_j^{(t+1)} = \frac{\sum_{i=1}^{n} q_i^{(t)}(z_i = j)\, x_i}{\sum_{i=1}^{n} q_i^{(t)}(z_i = j)}$$

(ii)

$$\Rightarrow \frac{\partial f(\pi, \lambda)}{\partial \pi_j} - n = 0$$

$$\Rightarrow \mathbb{E}_{z_i \sim q_i(t)}\left[\mathbb{1}(z_i = j)\right] * \left[\log(\pi_{z_i}) + x_i(\cancel{\log(\lambda_{z_i})}) - \cancel{\log(x_i!)} - \cancel{\lambda_{z_i}}\right]$$

$$\Rightarrow \frac{\text{derivative}}{\pi_j} = \mathbb{E}_{z_i \sim q_i(t)}\left[\mathbb{1}(z_i = j)\right] * \left(\frac{1}{\pi_j}\right)$$

$$\Rightarrow \frac{\partial f(\pi, \lambda)}{\partial \pi_j} - n = 0$$

$$\Rightarrow \frac{1}{\lambda_j} \sum_{i=1}^{n} q_i^{(t)}(z_i = j) - n = 0$$

$$\Rightarrow \lambda_j = \frac{\sum_{i=1}^{n} q_i^{(t)}(z_i = j)}{n}.$$