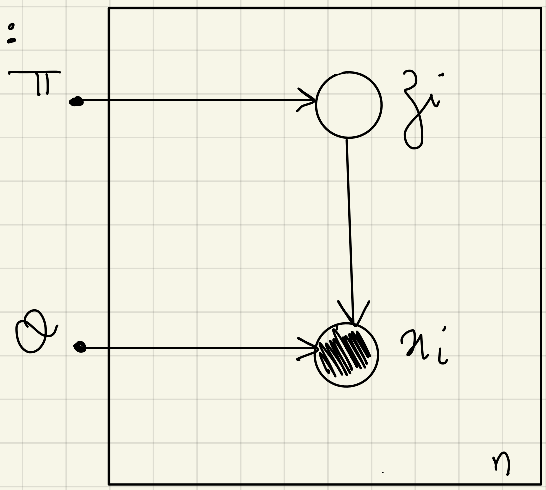
GMM : ① $Z_i \overset{iid}{\sim} \mathcal{M}(1, \pi), \forall i \in \{1, \dots, n\}$ hidden latent

② $X_i \mid Z_{ik} = 1 \sim \mathcal{N}(\mu_k, \Sigma_k)$

graphical model :



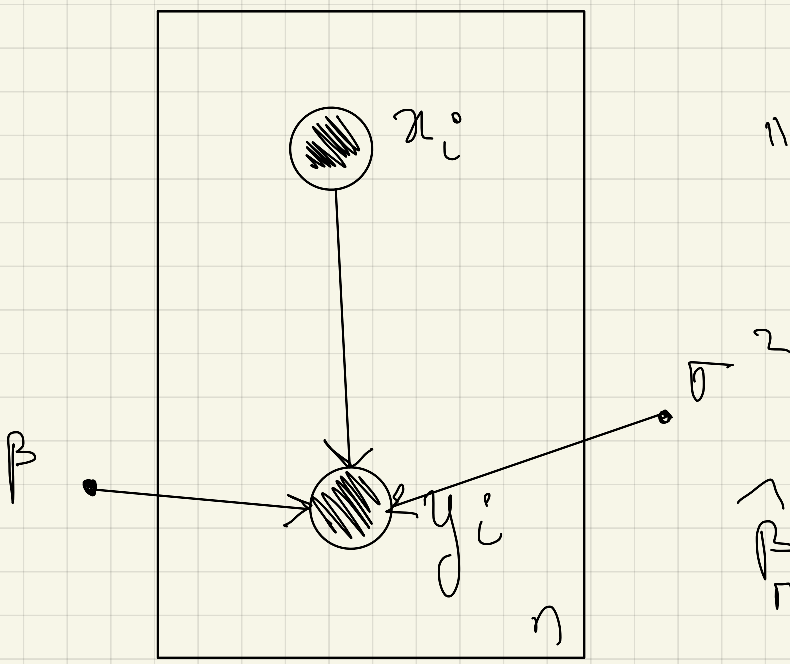
"one-to-one relationship"

$$\pi = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_K \end{pmatrix}$$

$$\theta = (\mu_k, \Sigma_k)_k$$

log-likelihood : $\log p((x_i)_i \mid \pi, \theta) = \log \left( \sum_{(z_i)_i} p((x_i, z_i)_i \mid \pi, \theta) \right)$

Regression (linear) $y_i \mid x_i \sim \mathcal{N}(x_i^T \beta, \sigma^2)$
$\forall i \in \{1, \dots, n\}$



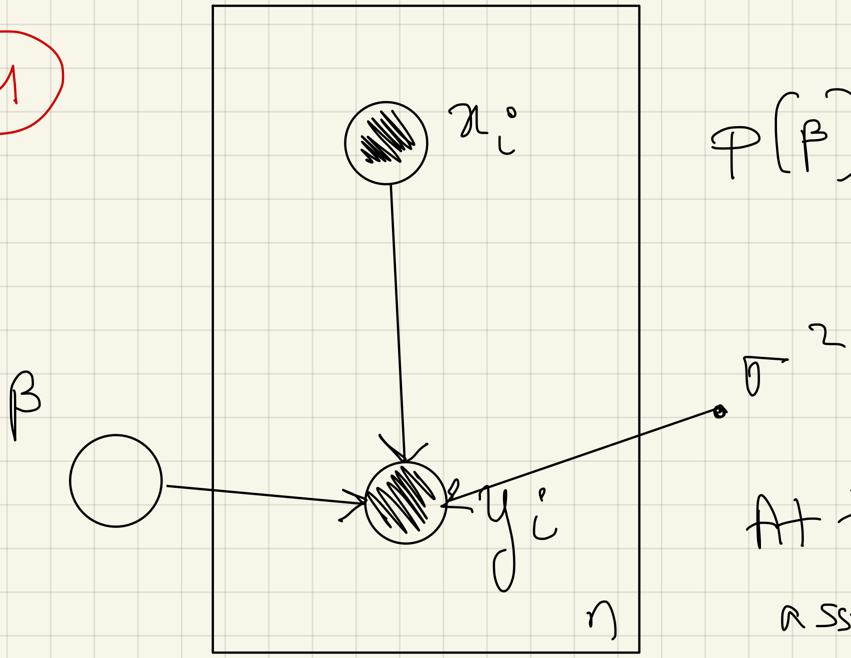
"frequentist framework"

$$\hat{\beta}_{MLE} = (X^T X)^{-1} X^T Y$$

## Bayesian linear regression :

$$y_i \mid x_i \sim \mathcal{N}(x_i^T\beta, \sigma^2) \quad \forall i \in \{1, \dots, n\}$$

V1



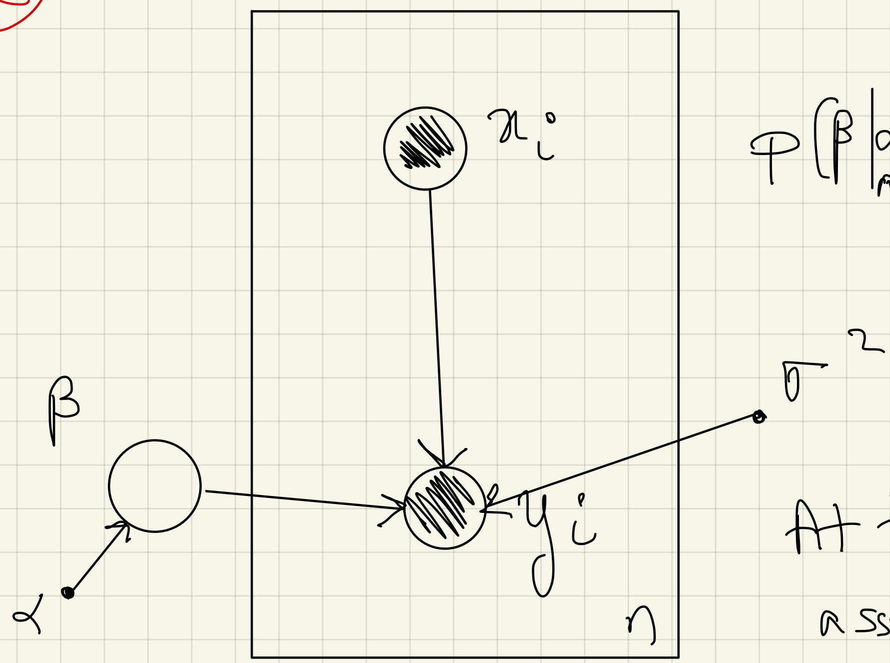

$$p(\beta) = \mathcal{N}\left(\beta; 0_p, \frac{I_p}{\alpha}\right)$$

At that point, $\alpha$ is assumed to be fixed.

---

$$y_i \mid x_i \sim \mathcal{N}(x_i^T\beta, \sigma^2) \quad \forall i \in \{1, \dots, n\}$$

V2




$$p(\beta \mid \alpha) = \mathcal{N}\left(\beta; 0_p, \frac{I_p}{\alpha}\right)$$

$\alpha$ is now a parameter

At that point, $\alpha$ is assumed to be fixed.

log-likelihood : $\log p(Y \mid X, \alpha, \sigma^2)$

(type-2 log-likelihood)

$$= \log\left\{ \int p(Y \mid X, \beta, \sigma^2)\, p(\beta \mid \alpha)\, d\beta \right\}$$

**Problem:** $$\boxed{\hat{\alpha}, \hat{\sigma}^2 = \underset{\alpha, \sigma^2}{\operatorname{argmax}} \log p(Y \mid X, \alpha, \sigma^2)}$$

**Question:** EM ?

$$p(\beta \mid X, Y, \alpha, \sigma^2) = \ ?$$

**Reminder:** $Z \sim \mathcal{N}(\mu, \Sigma)$

$$\log p(z \mid \mu, \Sigma) = -\frac{1}{2}(z-\mu)^T \Sigma^{-1} (z-\mu) + ct_1$$

$$= -\frac{1}{2} z^T \Sigma^{-1} z + \frac{1}{2} z^T \Sigma^{-1} \mu + \frac{1}{2} \mu^T \Sigma^{-1} z - \frac{1}{2} \mu^T \Sigma^{-1} \mu + ct_1$$

$$= -\frac{1}{2} z^T \Sigma^{-1} z + z^T \Sigma^{-1} \mu + ct_2$$

$$\Sigma \qquad \Sigma\, \Sigma^{-1} \mu = \mu$$

$$\log p(\beta \mid X, Y, \alpha, \sigma^2) = \log\left[\frac{p(Y \mid X, \beta, \sigma^2)\, p(\beta \mid \alpha)}{p(Y \mid X, \alpha, \sigma^2)}\right]$$

$$= \log p(Y \mid X, \beta, \sigma^2) + \log p(\beta \mid \alpha) + ct_1 .$$

$$= -\frac{1}{2}(Y - X\beta)^T(\sigma^2 I_n)^{-1}(Y - X\beta) - \frac{1}{2}\beta^T\left(\frac{I_p}{\alpha}\right)^{-1}\beta + ct_2$$

$$= -\frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta) - \frac{\alpha}{2}\beta^T\beta + ct_2$$

$$= -\frac{1}{2\sigma^2}\left(Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta\right) - \frac{\alpha}{2}\beta^T\beta + ct_2$$

$$= -\frac{1}{2}\beta^T\frac{X^TX}{\sigma^2}\beta + \beta^T\frac{X^TY}{\sigma^2} - \frac{\alpha}{2}\beta^T\beta + ct_3$$

$$= -\frac{1}{2}\beta^T\left(\frac{X^TX}{\sigma^2} + \alpha I_p\right)\beta + \beta^T\frac{X^TY}{\sigma^2} + ct_3$$

$S_n^{-1} =$

$\Downarrow$

$$S_n = \left(\frac{X^TX}{\sigma^2} + \alpha I_p\right)^{-1}$$

$$\mu = S_n \frac{X^TY}{\sigma^2}$$

$$\Rightarrow p(\beta \mid X, Y, \alpha, \sigma^2) = \mathcal{N}(\beta ; m_n, S_n)$$

with $S_n = \left( \frac{X^T X}{\sigma^2} + \alpha I_p \right)^{-1}$

and $m_n = S_n \frac{X^T Y}{\sigma^2} = \left( \frac{X^T X}{\sigma^2} + \alpha I_p \right)^{-1} \frac{X^T Y}{\sigma^2}$

$$= \left( X^T X + \alpha \sigma^2 I_p \right)^{-1} X^T Y$$

$$= \hat{\beta}_{MAP}$$

$\Rightarrow$ this expression is exact $\Rightarrow$ can use EM to maximize $\log p(Y \mid X, \alpha, \sigma^2)$

David McKay "evidence procedure".

<u>E step:</u> compute $S_n$ and $m_n$ ($\alpha, \sigma^2$ fixed)

<u>M step:</u> $\hat{\alpha} = \frac{p}{\mathrm{Tr}(S_n + m_n m_n^T)}$ ($m_n, S_n$ fixed)

$$\hat{\sigma}^2 = \frac{1}{n} \left\{ \| Y - X m_n \|^2 + \mathrm{Tr}(X^T X S_n) \right\}$$

<u>for the M step:</u> $\Rightarrow \hat{\alpha} = ?$ $\Rightarrow \hat{\sigma}^2 = ?$

$$\log p(Y, \beta \mid X, \alpha, \sigma^2)$$

$$= \log p(Y \mid X, \beta, \sigma^2) + \log p(\beta \mid \alpha)$$

$$= \log \left\{ \frac{1}{\sqrt{(2\pi)^n |\sigma^2 I_n|}} \exp\left( -\frac{1}{2} (Y - X\beta)^T (\sigma^2 I_n)^{-1} (Y - X\beta) \right) \right\}$$

$$+ \log \left\{ \frac{1}{\sqrt{(2\pi)^p \left| \frac{I_p}{\alpha} \right|}} \exp\left( -\frac{1}{2} \beta^T \left( \frac{I_p}{\alpha} \right)^{-1} \beta \right) \right\}$$

$$= -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} Y^T Y + \frac{1}{\sigma^2} Y^T X \beta - \frac{1}{2\sigma^2} \beta^T X^T X \beta$$

$$+ \frac{p}{2} \log(\alpha) - \frac{\alpha}{2} \beta^T \beta + ct_1$$

$$\underset{\beta}{E}[\ldots] = ?$$

from the posterior distribution.

$$E_{\beta}\left[\log p(y, \beta \mid X, \alpha, \sigma^2)\right]$$

$$= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2} y^T y + \frac{1}{\sigma^2} y^T X m_n - \frac{1}{2\sigma^2} \mathrm{Tr}\left(X^T X (S_n + m_n m_n^T)\right)$$

$$+ \frac{p}{2}\log(\alpha) - \frac{\alpha}{2}\mathrm{Tr}(S_n + m_n m_n^T) + ct_1$$

$$E_{\beta}\left[\underline{\beta^T X^T X \beta}\right] = E_{\beta}\left[\mathrm{Tr}\left(X^T X \beta \beta^T\right)\right]$$

$$= \mathrm{Tr}\left(X^T X \, E_{\beta}\left[\beta \beta^T\right]\right)$$

$$= \mathrm{Tr}\left(X^T X \left(S_n + m_n m_n^T\right)\right)$$

$$E_{\beta}\left[\beta^T \beta\right] = E_{\beta}\left[\sum_{j=1}^{p} \beta_j^2\right] = \sum_{j=1}^{p} E_{\beta_j}\left[\beta_j^2\right]$$

$$= \sum_{j=1}^{p}\left[(S_n)_{jj} + (m_n)_j^2\right] = \mathrm{Tr}(S_n + m_n m_n^T)$$

$$\hat{\alpha}, \hat{\sigma}^2 = \underset{\alpha, \sigma^2}{\arg\max} \; E_\beta \left[ \log p(y, \beta \mid X, \alpha, \sigma^2) \right]$$

$$= \underset{\alpha, \sigma^2}{\arg\max} \left[ -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} y^T y + \frac{1}{\sigma^2} y^T X m_n - \frac{1}{2\sigma^2} \mathrm{TR}\left( X^T X (S_n + m_n m_n^T) \right) + \frac{p}{2} \log(\alpha) - \frac{\alpha}{2} \mathrm{TR}(S_n + m_n m_n^T) + c A_1 \right]$$

$$\frac{\partial(\ldots)}{\partial \alpha} = \frac{p}{\cancel{2}} \times \frac{1}{\alpha} - \frac{1}{\cancel{2}} \mathrm{TR}(S_n + m_n m_n^T) = 0$$

$$\Rightarrow \hat{\alpha} = \frac{p}{\mathrm{TR}(S_n + m_n m_n^T)}$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n} \left\{ \| y - X m_n \|^2 + \mathrm{TR}(X^T X S_n) \right\}$$