$\boxed{1. \text{ Model Selection Criteria}}$

- Let M_j : Model j
  l_j : log-likelihood of M_j
  $\hat{\theta}_j$: MLEs for M_j
  d_j : # free parameters in M_j
  n : # data points.

We want to minimize:

AIC criterion : $AIC(M_j) = -l(\hat{\theta}_j) + d_j$.

BIC criterion : $BIC(M_j) = -l_j(\hat{\theta}_j) + \frac{1}{2} d_j \log(n)$.

- Also leave-one-out procedure, Monte-Carlo cross-validation, cross-validation with a K-fold approach, etc.

• Example of free parameters for the Gaussian mixture model:

Number of components : K

$$X|_{Y=y_j} \sim N\left(\begin{bmatrix} \mu_1^j \\ \mu_2^j \end{bmatrix} ; \begin{bmatrix} \sigma_{11}^j & \sigma_{12}^j \\ \sigma_{12}^j & \sigma_{22}^j \end{bmatrix} \right)$$

$$\mathbb{P}[Y = y_j] = \pi_j \quad ; \quad \sum_{j=1}^{K} \pi_j = 1$$

Free parameters for the mean: $2K$

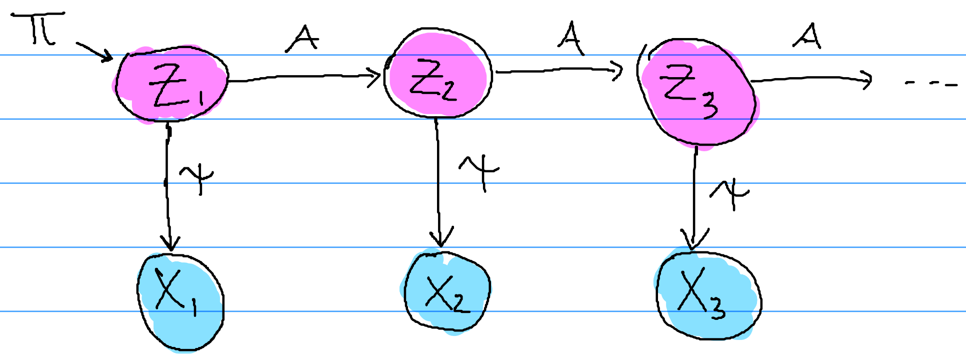
Free parameters for the Σ : $3K$

Free parameters for π : $K-1$

$K=4$ $\quad \pi_1, \pi_2, \pi_3 \Rightarrow \pi_4 = 1 - \sum_{j=1}^{K-1} \pi_j = 1 - (\pi_1 + \pi_2 + \pi_3)$.

2. Hidden Markov Models

- Latent variables and observable variables



- Notation:

$$\pi_i = \mathbb{P}[Z_1 = i]$$
$$A_{ij} = \mathbb{P}[Z_{t+1} = j \mid Z_t = i]$$
$$\psi_i(x) = \mathbb{P}[X_t = x \mid Z_t = i] .$$

$$Z_t \in \{1, \dots, K\} \quad \text{and} \quad X_t \in \mathcal{X} .$$

- The log-likelihood:

$$\ell(\Theta \mid z) = \sum_{i=n}^{K} \log \overbrace{\pi_i \cdot \mathbb{1}(z_1 = i)}^{\text{initial } Z_1}$$
$$+ \sum_{t=1}^{n} \sum_{i=1}^{K} \log \overbrace{\psi_i(x_t) \cdot \mathbb{1}(z_t = i)}^{\text{emission probability}}$$
$$+ \sum_{t=1}^{n} \sum_{i=1}^{K} \log \overbrace{A_{ij} \cdot \mathbb{1}(z_t = i, z_{t+1} = j)}^{\text{transition probability}}$$