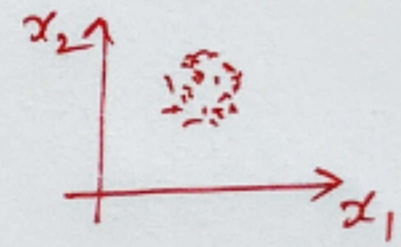
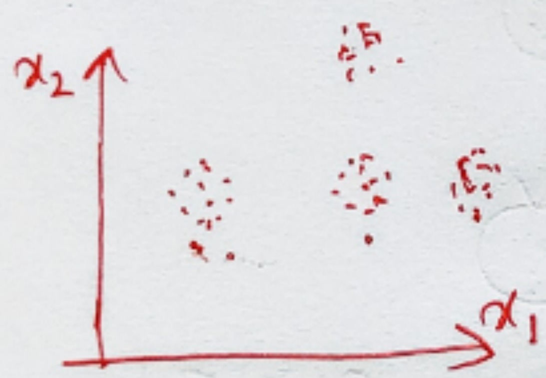
$$P(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad x \in \mathbb{R}.$$

$x \in \mathbb{R}^d$

$$P(x \mid \mu_d, \Sigma_{d\times d}) = \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)} = \mathcal{N}(x \mid \mu, \Sigma)$$



Mixture Gaussian Distributions



z_1, z_2, z_3, z_4

$z_1 = 1, \; z_2 = 1$

$$P(x \mid \mu, \Sigma) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}(x \mid \mu_k, \Sigma_k)$$

Let us introduce a K-dimensional binary random vector

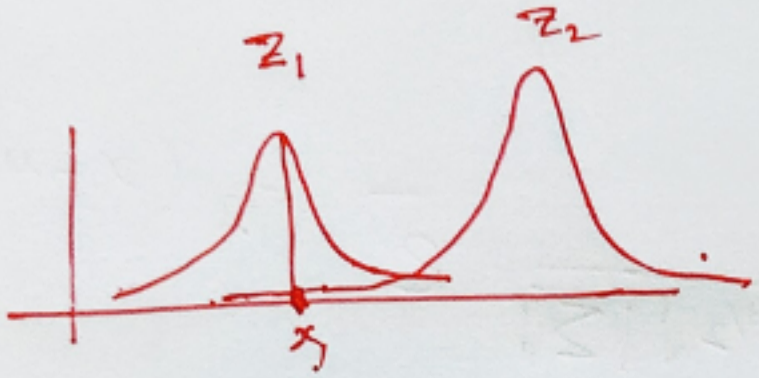
$z \in \{0,1\}^K$

$z = [z_1 \; z_2 \; \cdots \; z_K]$ $\qquad z = [1,0,0,0]$

$z_k \in \{0,1\}$. $\sum_{i=1}^{K} z_i = 1$ $\qquad z = [0,0,1,0]$

Bf

$S = \{x_1, x_2, \ldots, x_N\} \quad x_i \in \mathbb{R}^d,$



$P(x_i \mid z_1 = 1)$ and $P(x_i \mid z_2 = 1)$

$P(x \mid z)$

$$\underline{P(x, z)} = \underbrace{P(z)} \cdot \underbrace{P(x \mid z)}$$ Bayes Theorem.

$P(z_k = 1) = \pi_k$

$P(z_k = 0) = 1 - \pi_k$

$$\underline{P(z)} = P(z_1, z_2, \ldots, z_K)$$
$$= P(z_1)\, P(z_2) \cdots P(z_K)$$
$$= \prod_{k=1}^{K} (\pi_k)^{z_k}$$

$$\sum_{k=1}^{K} \pi_k = 1.$$

$0 \le \pi_k \le 1.$

$$p(x \mid z_k = 1) = \mathcal{N}(x \mid \mu_k, \Sigma_k)$$

$$p(x \mid z_1 = 1) = \mathcal{N}(x \mid \mu_1, \Sigma_1)$$

$$p(x \mid z_2 = 1) = \mathcal{N}(x \mid \mu_2, \Sigma_2)$$

$$p(x \mid z) = \prod_{k=1}^{K} \mathcal{N}(x \mid \mu_k, \Sigma_k)^{z_k} \qquad z = [z_1 \; \cdots \; z_K]$$

$$p(x, z) = p(z) \cdot p(x \mid z) \qquad z = [z_1 \; \cdots \; z_k]$$

K

$$p(x) = \sum_{z} p(z) \, p(x \mid z)$$

$$= \sum_{k=1}^{K} \underline{p(z_k = 1)} \, p(x \mid z_k = 1)$$

$$p(x) = \underline{\sum_{k=1}^{K} \pi_k \underline{\mathcal{N}(x \mid \mu_k, \Sigma_k)}}$$

$$\gamma(z_{nk}) = p(z_k = 1 \mid x = x_n) = \frac{p(z_k = 1) \, p(x_n \mid z_k = 1)}{p(x = x_n)}$$

$$= \frac{p(z_k = 1) \, p(x_n \mid z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1) \cdot p(x_n \mid z_j = 1)}$$

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)}.$$

$X = \{x_1, x_2, \dots, x_N\}$ $\qquad$ $\cancel{\theta} =$

$$P(X) = P(x_1) \cdot P(x_2) \cdot \dots \cdot P(x_N)$$

$$\log(P(X)) = \sum_{n=1}^{N} \log(\underline{P(x_i)})$$

$$L(X \mid \mu, \Sigma, \pi) = \sum_{n=1}^{N} \log\left(\sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(x_n \mid \mu_j, \Sigma_j)\right)$$

$$\max_{\mu, \Sigma, \pi} L(X \mid \mu, \Sigma, \pi).$$

$$\nabla_{\mu_k} L = \vec{0}$$

$$\nabla_{\mu_k} L = \nabla_k \sum_{n=1}^{N} \log\left[\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)\right]$$

$$\frac{d}{dx}\log(x) = \frac{1}{x}$$

$$\frac{d}{dx}\log(f(x)) = \frac{1}{f(x)} \cdot \frac{d}{dx} f(x)$$

$$= \sum_{n=1}^{N} \frac{\pi_k \nabla_{\mu_k}\left[\mathcal{N}(x_n \mid \mu_k, \Sigma_k)\right]}{\left[\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)\right.}$$

$$\nabla_{\mu_k} \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma_k|}} e^{-\frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)}$$

$$\left|\right| \frac{d}{dx} e^{f(x)} = e^{f(x)} \cdot \frac{d}{dx} f(x)$$

$$= \underbrace{\frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma_k|}} e^{-\frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)}} \cdot \nabla\left[-\frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)\right]$$

$$\left|\right| A^T = A, \quad A \in \mathbb{R}^{n \times n}$$

$$\nabla_x x^T A x = 2Ax$$

$$= +\mathcal{N}(x_n \mid \mu_k, \Sigma_k)\Sigma_k^{-1}(x_n-\mu_k)$$

$$\nabla_{\mu_k} L = \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\left[\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)\right]} \Sigma_k^{-1}(x_n-\mu_k) = 0$$

$$= \sum_{n=1}^{N} \underbrace{\gamma(nk)}\, \Sigma_k^{-1} \underbrace{(x_n-\mu_k)} = 0$$

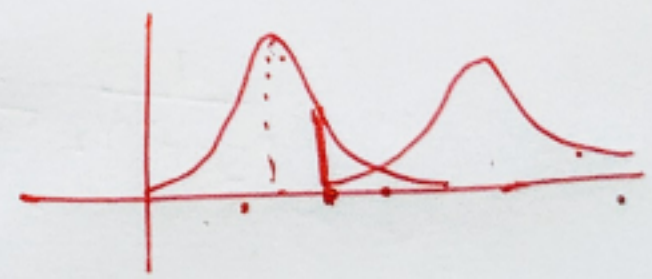
$$= \Sigma\, \Sigma_k^{-1} \sum_{n=1}^{N} \gamma(nk)(x_n-\mu_k) = 0$$

$$\sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k) = 0$$

$$\sum_{n=1}^{N} \gamma(z_{nk})\, x_n = \mu_k \sum_{n=1}^{N} \gamma(z_{nk})$$

$$\boxed{\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})\, x_n} \qquad \gamma(z_{nk}) = \underline{p(z_k = 1 \mid x_n)}$$

$$N_k = \sum_{n=1}^{N} \gamma(z_{nk})$$



$$\nabla_{\Sigma_k} L = 0$$

$$\nabla_{\Sigma_k} \sum_{n=1}^{N} \log\left[\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)\right]$$

$$= \sum_{n=1}^{N} \frac{\pi_k \nabla_{\Sigma_k}\left(\mathcal{N}(x_n \mid \mu_k, \Sigma_k)\right)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)}$$

$$\nabla_{\Sigma_k} = \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)} \left[-\frac{1}{2} \frac{1}{\sqrt{|\Sigma_k|}} \Sigma_k^{-1} + \frac{1}{2\sqrt{|\Sigma_k|}} \Sigma^{-1} (x_n-\mu_k)(x_n-\mu_k)^T \Sigma^{-1} \right]$$

$$= \frac{1}{2} \mathcal{N}(x_n \mid \mu_k, \Sigma_k) \left[-\Sigma_k^{-1} + \Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T \Sigma_k^{-1} \right]$$

$$\nabla_{\Sigma_k} L = \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)} \left[-\Sigma_k^{-1} + \Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T \Sigma_k^{-1} \right] = 0$$

$$\sum_{n=1}^{N} \gamma(z_{nk}) \left[-\Sigma_k^{-1} + \Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T \Sigma_k^{-1} \right] = 0$$

$$\Sigma_k \sum_{n=1}^{N} \gamma(z_{nk}) \Sigma_k^{-1} \Sigma_k = \Sigma_k \Sigma_k^{-1} \left[\sum_{n=1}^{N} \gamma(z_{nk})(x_n-\mu_k)(x_n-\mu_k)^T \right] \Sigma_k^{-1} \Sigma_k$$

$$\Sigma_k \left(\sum_{n=1}^{N} \gamma(z_{nk}) \right) = \sum_{n=1}^{N} \gamma(z_{nk})(x_n-\mu_k)(x_n-\mu_k)^T$$

$$\Sigma_k^{*} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(x_n-\mu_k)(x_n-\mu_k)^T$$

$$\nabla_{\Sigma_k} \left(\mathcal{N}(x_n \mid \mu_k, \Sigma_k) \right)$$

$$\nabla_{\Sigma_k} \left[\frac{1}{(2\pi)^{d/2} \sqrt{|\Sigma_k|}} e^{-\frac{1}{2}[x_n - \mu_k]^T \Sigma_k^{-1} [x_n - \mu_k]} \right]$$

$$\frac{d}{dx}(f \cdot g) = f \frac{dg}{dx} + g \frac{df(x)}{dx}$$

$$\frac{1}{(2\pi)^{d/2}} \left[\nabla_{\Sigma_k}\left[\frac{1}{\sqrt{|\Sigma_k|}}\right] e^{-\frac{1}{2}[x_n-\mu_k]^T \Sigma_k^{-1}[x_n-\mu_k]} + \frac{1}{\sqrt{|\Sigma_k|}} \nabla_{\Sigma_k} e^{-\frac{1}{2}[x_n-\mu_k]^T \Sigma_k^{-1}[x_n-\mu_k]} \right]$$

$$\nabla_{\Sigma_k}\left[d\,|\Sigma_k|^{-1/2} \right] \quad \Big\| \quad \frac{d\,|X|}{dX} = |X|\,X^{-1}$$

$$= -\frac{1}{2} |\Sigma_k|^{-3/2} \cdot |\Sigma_k| \cdot \Sigma_k^{-1}$$

$$= -\frac{1}{2} |\Sigma_k|^{-1/2} \Sigma_k^{-1}$$

$$\nabla_{\Sigma_k}\left[e^{-\frac{1}{2} p^T \Sigma_k^{-1} p} \right] = -\frac{1}{2} e^{-1/2\, p^T \Sigma_k^{-1} p} \, \nabla_{\Sigma_k}\left(p^T \Sigma_k^{-1} p \right) \quad \Big\| \quad \frac{d}{dX} a^T X^{-1} b = -X^{-1} a b^T X^{-1}$$

$$= +\frac{1}{2} e^{-1/2\, p^T \Sigma_k p} \, \Sigma_k^{-1} p p^T \Sigma_k^{-1}$$

$$\frac{\partial}{\partial \pi_k} L(x \mid \mu, \Sigma, \pi) = 0$$

$$\sum_{j=1} \pi_j =$$

$$\frac{\partial}{\partial \pi_k}\left[\sum_{n=1}^{N} \log\left(\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)\right)\right]$$

$$= \sum_{n=1}^{N} \frac{\mathcal{N}(x_n \mid \mu_j, \Sigma_j)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)}$$

$$\sum_{j=1}^{K} \pi_j = 1$$

$$f = L + \lambda\left(\sum_{j=1}^{K} \pi_j - 1\right)$$

$$\frac{\partial f}{\partial \pi_k} = \frac{\partial L}{\partial \pi_k} + \lambda = 0$$

$$\sum_{n=1}^{N} \gamma(z_{nk}) = -\lambda \pi_k$$

$$= N \pi_k$$

$$\pi_k = \frac{1}{N} \sum_{n=1}^{N} \gamma(z_{nk})$$

$$\sum_{k=1}^{K} \sum_{n=1}^{N} \frac{\mathcal{N}(x_n \mid \mu_j, \Sigma_j)\, \pi_k}{\sum_{j=1}^{N} \pi_j \mathcal{N}(x_n \mid \mu_j, \Sigma_j)} = -\lambda \underbrace{\sum_{k=1}^{K} \pi_k}_{=1}$$

$$\sum_{k=1}^{K} \sum_{n=1}^{N} \gamma(z_{nk}) = -\lambda = N \Rightarrow \lambda = -N$$