

## p.16

- Evaluating  $\log \sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z}; \theta)$  can be intractable. Suppose we have 30 binary latent features,  $\mathbf{z} \in \{0, 1\}^{30}$ . Evaluating  $\sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z}; \theta)$  involves a sum with  $2^{30}$  terms. For continuous variables,  $\log \int_{\mathbf{z}} p(\mathbf{x}, \mathbf{z}; \theta) d\mathbf{z}$  is often intractable.

$$\sum_{\mathbf{z}} p(x, \mathbf{z}; \theta) = \sum_{\mathbf{z}} [p(x | \mathbf{z}; \theta) p(\mathbf{z})]$$
 이런 식에 의해 joint distribution을 구할 수 있습니다.

$$p(x | \mathbf{z}; \theta)$$

likelihood는 미리 정한 모델에 의해서 계산할 수 있습니다,  
gaussian mixture model의 경우,  $\mathbf{z}$ 의 값(어떤 component에 속하느냐)에 따라 mean과 variance를 정해 얻을 수 있고,  
linear gaussian model의 경우,  $\mathbf{z}$ 를 linear transform한 벡터를 mean으로 하고, 미리 정한 variance를 통해 얻을 수 있으며,  
neural network model(VAE)의 경우, bernoulli나 gaussian의 파라미터를 네트워크를 통해 구하여 계산할 수 있습니다.

$$p(\mathbf{z}; \theta)$$

prior도 미리 정한 모델에 의해 계산됩니다.

gaussian mixture model의 경우,

categorical distribution으로 표현되며 해당 parameter들은  $\theta$ 안에 포함되고 트레이닝에 의해 바뀝니다.

linear gaussian model의 경우,

normal distribution으로 표현되며, 보통  $N(0, I)$ 로 결정하고 mean과 variance가 트레이닝에 의해 바뀌지 않습니다.

neural network model의 경우(VAE),

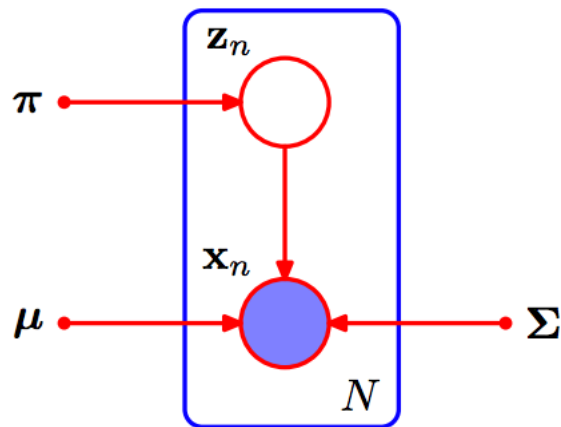
normal distribution으로 표현되며, 보통  $N(0, I)$ 로 결정하고 mean과 variance가 트레이닝에 의해 바뀌지 않습니다.

$\mathbf{z}$ 를 marginalize하는 계산이 불가능한(intractable) 경우는  $\mathbf{z}$ 의 차원이 매우 큰 경우 예를 들어 30차원인 경우,  $\mathbf{z}$ 가 0또는 1의 값을 가질 수 있는 binomial random variables이더라도  $2^{30}$ 가지의 수 만큼 summation을 해야하므로 실질적으로 불가능하다고 말하는 겁니다.

또는  $\mathbf{z}$ 가 continuous variables인 경우, linear gaussian model처럼 적분을 closed form으로 구할 수 있는 경우도 있지만,  
(Pattern Recognition and Machine Learning, p.93 참고)

$p(x|\mathbf{z})$ 가 non-linear로 표현되는 neural-network의 경우 쉽게 계산할 수 없으므로 intractable하다고 합니다.

## Gaussian Mixture Model



Prior : categorical

$$p(\mathbf{z}) = \prod_{k=1}^K \pi_k^{z_k}.$$

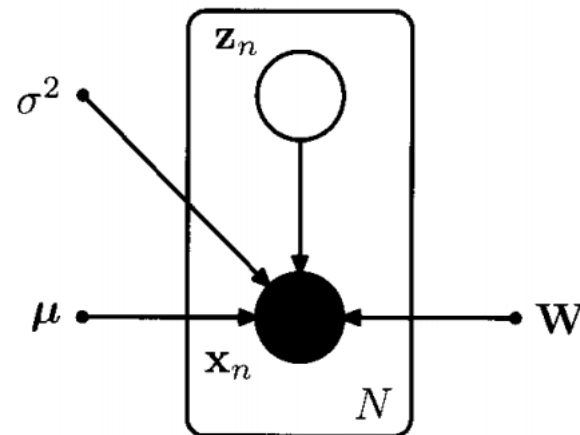
Likelihood : Gaussian

$$p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^K \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_k}$$

Marginal likelihood : Gaussian Mixture

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k).$$

## Linear Gaussian Model (probabilistic PCA)



Prior : Gaussian

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$$

Likelihood : Gaussian

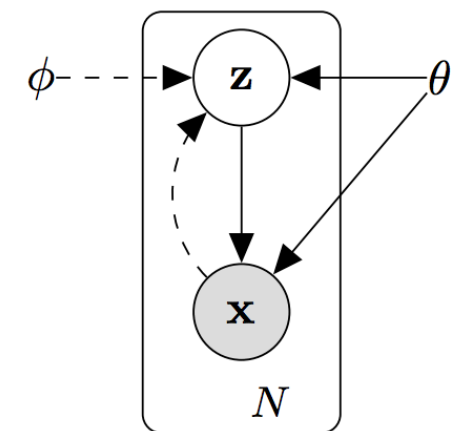
$$p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \boldsymbol{\mu}, \sigma^2\mathbf{I})$$

Marginal likelihood : Gaussian

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$$

$$\mathbf{C} = \mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}.$$

## Variational Auto-Encoder



Prior : Gaussian

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$$

Likelihood : Gaussian or bernoulli

$$p(x|z) = \mathcal{N}(x|\mu(z), \sigma(z))$$

*or*  $B(p_1(z), p_2(z), \dots, p_N(z))$

$\mu, \sigma, p_n$  are neural network functions

**Marginal likelihood : Intractable**

$$p(x) = ???$$