Ch 03. 다변량 정규 분포

1. 넘파이와 다차원 배열
2. 다변량 정규 분포
3. 2차원 정규 분포 시각화
4. 다변량 정규 분포의 최대 가능도 추정

3.1 넘파이의 다차원 배열

3.1.1 다차원 배열

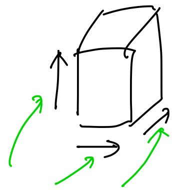
$\circ$ → 스칼라 → 0차원 텐서

$\begin{pmatrix} \circ \\ \circ \\ \circ \end{pmatrix}$ → 벡터 → 1차원 텐서

$\begin{pmatrix} \circ & \circ \\ \circ & \circ \end{pmatrix}$ → 행렬 → 2차원 텐서

→ 모두 텐서!

이때!



축! axis → 3차원 배열

3.1.3 원소별 연산

$$\begin{pmatrix} \circ & \circ \\ \circ & \circ \end{pmatrix} + \begin{pmatrix} \circ & \circ \\ \circ & \circ \end{pmatrix} = \begin{pmatrix} \circ & \circ \\ \circ & \circ \end{pmatrix}$$ ⇒ 원소별 합

$$\begin{pmatrix} \circ & \circ \\ \circ & \circ \end{pmatrix} \times \begin{pmatrix} \circ & \circ \\ \circ & \circ \end{pmatrix} = \begin{pmatrix} \circ & \circ \\ \circ & \circ \end{pmatrix}$$ ⇒ 원소별 곱 ⇒ 아다마르 곱 (Hadamard product)

3.1.4 벡터의 내적과 행렬 곱

$$\begin{pmatrix} \circ \\ \circ \\ \circ \end{pmatrix} \cdot \begin{pmatrix} \circ \\ \circ \\ \circ \end{pmatrix} = \triangle + \triangle + \triangle$$ ⇒ 벡터의 내적

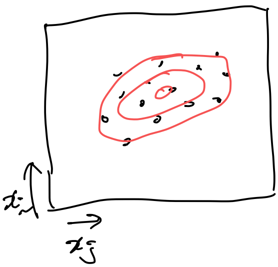
$$\begin{pmatrix} \circ & \circ \\ \circ & \circ \end{pmatrix} \cdot \begin{pmatrix} \circ & \circ \\ \circ & \circ \end{pmatrix} = \triangle\triangle + \triangle\triangle + \triangle\triangle + \triangle\triangle$$ ⇒ 행렬곱

3.2 다변량 정규 분포

3.2.1 다변량 정규 분포 공식

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \Rightarrow N(x; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

도식화



평균 벡터

$$= \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$$

공분산 행렬

$$= \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & & \\ \vdots & & \ddots & \\ \sigma_{p1} & & & \sigma_{pp} \end{pmatrix}$$

여기서!

$\sigma_{11}, \sigma_{22} \cdots \sigma_{pp}$ → 분산! ⇒ x_i가 얼마나 흩어졌는지 보여줘!

$\sigma_{12}, \sigma_{21}, \cdots$ → 공분산! ⇒ x_i와 x_j의 연관성을 보여줘!

2절. 공분산은 어떻게 계산?

분산 $= Var[x_i] = E[(x_i - \mu_i)^2]$ $\overset{\text{공분산?}}{\Longleftrightarrow}$ 공분산 $= Cov[x_i, x_j] = E[(x_i - \mu_i)(x_j - \mu_j)]$

요고 외워!

2절. 전치는?

$x = (1, 2, 3) \overset{\text{전치!}}{\Longrightarrow} x^T = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$

2절. 행렬식은?

이건 알죠? determinant!

2절. 역행렬은?

$AA^{-1} = A^{-1}A = I$ ⇒ inverse matrix!

3.4 다변량 정규 분포의 최대 가능도 추정

3.4.1 최대 가능도 추정하기

$$N(x;\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^D|\Sigma|}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

요기서! 샘플 $D = \{x^{(1)}, x^{(2)}, \cdots x^{(N)}\}$ 이면! 요 'D'가 샘플링될 확률은!

$$p(D;\mu,\Sigma) = N(x^{(1)};\mu,\Sigma)\,N(x^{(2)};\mu,\Sigma)\cdots N(x^{(N)};\mu,\Sigma)$$

$$= \prod_{n=1}^{N} N(x^{(n)};\mu,\Sigma)$$

요건 2장과 같아! → 즉! μ, Σ가 매개변수 일 때, 샘플 D를 얻은 확률 밀도!

따라서!

$$p(D;\mu,\Sigma) = L(\mu,\Sigma) = \text{가능도!} = \text{가능도 함수당!!}$$

혹은! ↳ $\log p(D;\mu,\Sigma)$

여기로 마찬가지! 최대 가능도

$$\frac{\partial L}{\partial \mu} = 0 \Rightarrow \hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} x^{(n)}$$

$$\frac{\partial L}{\partial \Sigma} = 0 \Rightarrow \hat{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}(x^{(n)}-\hat{\mu})(x^{(n)}-\hat{\mu})^T$$