8.6 다변수 정규분포

D차원 다변수정규분포 (MVN: multivariate Gaussian normal distribution)의 확률밀도함수는 평균벡터 μ와 공분산행렬 Σ라는 두개의 모수를 가지며, 다음과 같은 수식으로 정의한다.

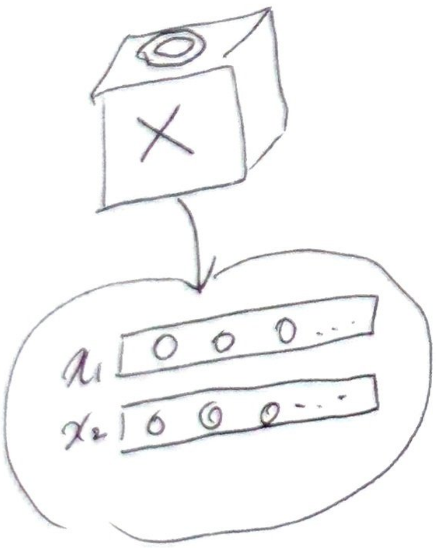
scalar × scalar = 확률밀도 scalar.

$$N(x;\mu,\Sigma) = \overbrace{\frac{1}{(2\pi)^{D/2}\underbrace{|\Sigma|}_{=\text{determinant of }\Sigma\,=\,\text{scalar}}^{1/2}}}^{\text{scalar}} \overbrace{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}^{\text{scalar}}$$

이 식에서 각 기호의 의미는 다음과 같다.

- $x \in \mathbb{R}^D$ 확률변수벡터
- $\mu \in \mathbb{R}^D$ 평균벡터
- $\Sigma \in \mathbb{R}^{D\times D}$ 공분산행렬

→ 양의 정부호인 대칭행렬이어야 함.
따라서 역행렬이 항상 존재한다.
Σ^{-1} = 정밀도행렬 = precision matrix = Λ



$N(x)$
⇒ x라는 벡터를 입력받아
x라는 벡터가 나올 확률을 출력한다.
input vector
output scalar

예제 for Σ^{-1}

다음과 같은 2차원 ($D=2$) 다변수정규분포를 생각하자. 2차원이므로 확률변수벡터는

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

만약 모수가 다음과 같다고 하자.

$$\mu = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

공분산행렬로부터 x_1과 x_2가 독립이라는 것을 알 수 있다. 확률밀도함수를 구하면 다음과 같다.

$$|\Sigma| = 1, \quad \Sigma^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$(x-\mu)^T \Sigma^{-1} (x-\mu) = \begin{bmatrix} x_1-2 & x_2-3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1-2 \\ x_2-3 \end{bmatrix}$$

$$= (x_1-2)^2 + (x_2-3)^2$$

$$\boxed{\mathcal{N}(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left((x_1-2)^2 + (x_2-3)^2\right)\right)}$$

$$p(x_1, x_2) = e^{(x_1-2)^2} e^{(x_2-3)^2}$$

$$= p(x_1)\, p(x_2)$$

$\therefore$ 독립

확률밀도함수값이 같은 등고선은 원이 된다.

$$(x_1-2)^2 + (x_2-3)^2 = r^2$$

$$\boxed{\text{scipy.stats.multivariate\_normal()}}$$

- mean = 평균벡터인수
- cov = 공분산행렬인수

예제 2

계수가 다음과 같을 때, 공분산행렬로부터 x_1과 x_2가 양의 상관관계가 있다는 것을 알 수 있다.

$$\mu = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 3 \\ 3 & 7 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{bmatrix}$$

$$※ \ \frac{\sigma_{12}}{\sqrt{\sigma_{11}^2 \sigma_{12}^2}} = \rho_{12} = \frac{1}{2}$$

이 때 확률밀도함수는 다음과 같다.

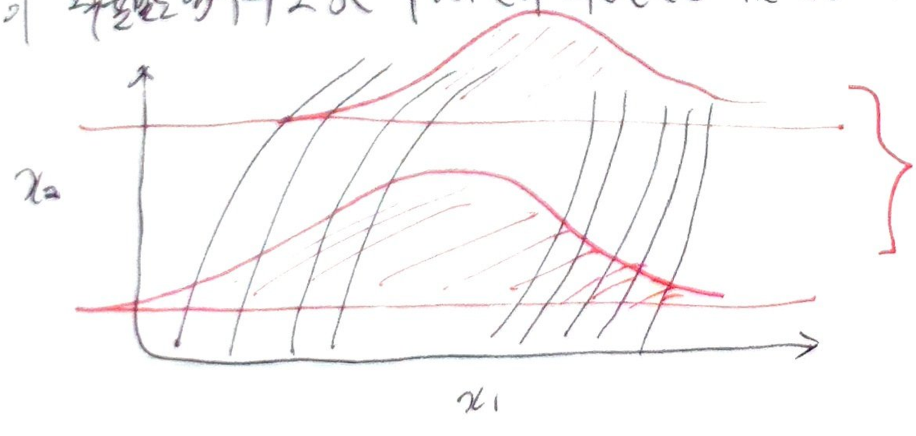
$$|\Sigma| = 5, \quad \Sigma^{-1} = \begin{bmatrix} \cancel{\sigma_{11}^2} \; 1.4 & \cancel{\sigma_{12}^2} \; -0.6 \\ \cancel{\sigma_{21}^2} \; -0.6 & \cancel{\sigma_{22}^2} \; 0.4 \end{bmatrix}$$

$$(x-\mu)^T \Sigma^{-1} (x-\mu) = \begin{bmatrix} x_1 - 2 & x_2 - 3 \end{bmatrix} \begin{bmatrix} 1.4 & -0.6 \\ -0.6 & 0.4 \end{bmatrix} \begin{bmatrix} x_1 - 2 \\ x_2 - 3 \end{bmatrix}$$

$$= \frac{7}{5}(x_1-2)^2 - \frac{6}{5}(x_1-2)(x_2-3) + \frac{2}{5}(x_2-3)^2$$

$$N(x_1, x_2) = \frac{1}{2\sqrt{5}\pi} \exp\left(\frac{7}{5}(x_1-2)^2 - \frac{6}{5}(x_1-2)(x_2-3) + \frac{2}{5}(x_2-3)^2 \right)$$

이 확률밀도함수의 모양은 다음과 같이 회전변환된 타원 모양이 된다.



상관관계 존재
($x_1 - x_2$)

다변수정규분포와 고윳값 분해

$\lambda_i > 0$

다변수정규분포의 공분산행렬 Σ은 양의 정부호인 대칭행렬이므로 대각화가능 (diagonalize)이다. 정밀도행렬 Σ^{-1}은 다음처럼 분해할수 있다.

- Λ = 고윳값행렬
- V = 고유벡터행렬

$$\Sigma^{-1} = V\Lambda^{-1}V^T$$

이를 이용하면 확률밀도함수는 다음처럼 좌표 변환할 수 있다.

$$\begin{aligned}
N(x) &\propto \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \\
&= \exp\left(-\frac{1}{2}(x-\mu)^T V\Lambda^{-1}V^T(x-\mu)\right) \\
&= \exp\left(-\frac{1}{2}(\boxed{V^T}(x-\mu))^T\Lambda^{-1}(\boxed{V^T}(x-\mu))\right) \\
&= \exp\left(-\frac{1}{2}(\underbrace{\boxed{V^{-1}}}_{V^T=V^{-1}}(x-\mu))^T\Lambda^{-1}(\underbrace{\boxed{V^{-1}}}_{V^T=V^{-1}}(x-\mu))\right)
\end{aligned}$$

이 식에서

$$\boxed{x' = V^{-1}(x-\mu)}$$

라고 하자. 이 식은 변환행렬 V^{-1}의 열벡터인 고유벡터를 새로운 축으로 가지도록 회전하고, μ벡터 방향으로 평행이동하는 것을 뜻한다.

$$\begin{bmatrix} g_1 & g_2 \end{bmatrix}^{-1} = \text{transformation matrix (변환행렬)}$$

basis vector (기저벡터)

→ 이전 basis에서 새로운 basis를 나타내는 g_1, g_2를 열로 가지는 행렬의 역행렬 V^{-1}를 곱하면, 새로운 basis를 기준으로 다시 정렬하는 것과 같다 = "좌표변환"

최종 확률밀도함수

— Λ는 고윳값 λ_i를 대각성분으로 가지는 대각행렬이므로, 새로운 좌표 x'에서 확률밀도함수는 타원이 된다. 타원의 반지름은 고윳값 크기에 비례한다. 반대로 이야기하면, 원래 좌표에서 확률밀도함수는 μ를 중심으로 갖고 고윳값에 비례하는 반지름을 가진 타원을 $\boxed{\text{고유벡터 방향으로 회전시킨 모양}}$이다.

$$N(x) \propto \exp\left(-\frac{1}{2} x'^{T} \Lambda^{-1} x'\right)$$

$$\propto \exp\left(\frac{x_1'^2}{\lambda_1^2} + \frac{x_2'^2}{\lambda_2^2} + \cdots + \frac{x_D'^2}{\lambda_D^2}\right)$$

V는 원래 고유벡터를 모아놓은 행렬!

$$V = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$$

$v_1 = g_1$, $v_2 = g_2$ … 기존의 고유벡터는 새롭게 변환(g_1, g_2)된 이후에 basis, 축이 된다.

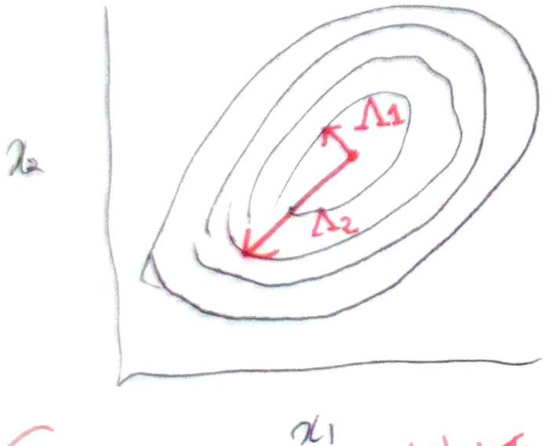
따라서 타원은 고유벡터 방향으로 회전한다.

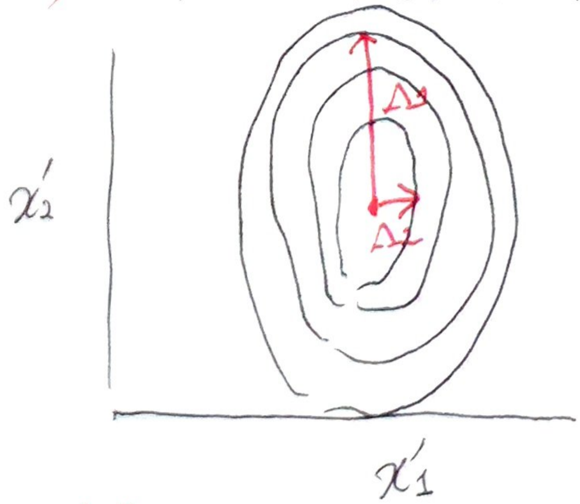
<확률밀도함수 예제>

독립X

x_1, x_2 좌표의 결합확률밀도함수 $\xRightarrow{\text{좌표변환}}$ x_1', x_2' 좌표의 결합확률밀도함수

"분리가능 // 독립"





$\left\{ \begin{array}{l} \text{공분산행렬의 고윳값} = \text{타원의 폭} \\ \text{공분산행렬의 고유벡터} = \text{타원의 방향} \end{array} \right\}$

* condition number $= \dfrac{\lambda_{max}}{\lambda_{min}}$,

크면 다중공선성으로 판단, 나쁘다! < 상관관계(독립x), scaling 문제.

다변수정규분포의 조건부확률분포

[정리] 다변수정규분포인 확률변수벡터 중, 어떤 원소의 값이 주어지면 다른 확률변수의 조건부확률분포는 다변수정규분포다.

⇒ 다변수정규분포 확률밀도함수를 자른 단면은 다변수정규분포가 된다.

예를 들어, 확률변수 X의 값 x를 두 벡터 x_1과 x_2로 나누었을 때,

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

x_2 값이 주어지면 (관측되면) X_1 만의 확률밀도함수가 다변수정규분포를 이루는 것을 증명하자.

x_1과 x_2에 따라 기댓값 벡터도 μ_1과 μ_2로 나뉘어진다.

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

공분산행렬 Σ도 다음처럼 나뉘어진다.

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

공분산행렬의 역행렬인 정밀도행렬 Λ도 마찬가지로 분할한다.

$$\Lambda = \Sigma^{-1} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}$$

이때, Σ와 Λ가 대칭행렬이므로 Λ_{11}과 Λ_{22}도 대칭행렬이고 $\Lambda_{12} = \Lambda_{21}$이다.

이를 적용하면

$$(x-\mu)^T \Sigma^{-1} (x-\mu) = (x_1-\mu_{1|2})^T \Lambda_{11} (x_1-\mu_{1|2}) + C(x_2, \mu, \Sigma)$$

가 된다. 이 식에서 조건부기댓값 $\mu_{1|2}$는

$$\mu_{1|2} = \mu_1 - \Lambda_{11}^{-1} \Lambda_{12} (x_2 - \mu_2)$$

이다. C는 x_1을 포함하지 않는 항을 가리키며 다음과 같다.

$$C = \mu_1^T \Lambda_{11} \mu_1 - 2\mu_1^T \Lambda_{12} (x_2-\mu_2) + (x_2-\mu_2)^T \Lambda_{22} (x_2-\mu_2) - (x_2-\mu_2)^T \Lambda_{12}^T \Lambda_{11}^{-1} \Lambda_{12} (x_2-\mu_2)$$

이 식에 지수함수를 적용하면

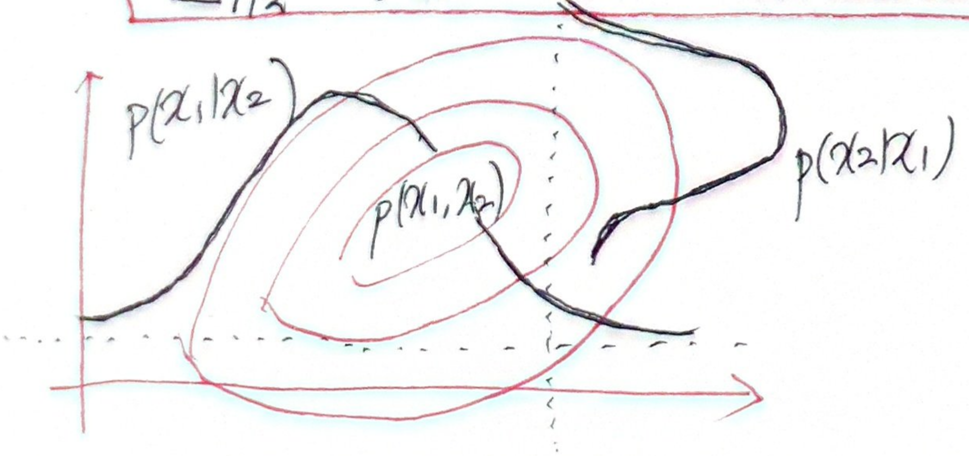
$$p(x_1|x_2) = C' \exp\left((x_1-\mu_{1|2})^T \Lambda_{11} (x_1-\mu_{1|2})\right)$$

가 된다. 이 식에서 $C' = \exp C$ 이다.

즉, x_2가 어떤 값으로 주어지면 x_1은 조건부기댓값 $\mu_{1|2}$와 조건부공분산행렬 $\Sigma_{1|2}$를 가지는 다변수정규분포가 된다. $\Sigma_{1|2}$는 분할행렬의 역행렬공식으로부터 다음과 같다.

$$\boxed{\Sigma_{1|2} = \Lambda_{11}^{-1} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}$$

단면을 잘라도 정규분포!



다변수정규분포의 주변확률분포

[정리] 다변수정규분포의 주변확률분포는 다변수정규분포다.

즉, 결합확률밀도함수를 어떤 확률변수의 값으로 적분하여 나머지 확률변수의 주변확률분포를 구하면 다변수정규분포이다. 예를 들어, x_1과 x_2로 이루어진 결합 확률밀도함수 $p(x_1, x_2)$를 x_2로 적분하면 x_1의 주변확률분포는 정규분포가 된다.

$$p(x_1) = \int p(x_1, x_2)\,dx_2 = \mathcal{N}(x_1; \mu_1, \Sigma_{11})$$

x_2의 주변확률분포의 기댓값은 원래 기댓값벡터 중 x_1 성분과 같고, 공분산행렬은 분할행렬 중 Σ_{11} 성분과 같다. (증명 생략)

하나의 축으로 압축해도 정규분포!