Multivariate Gaussian

두 벡터 확률 변수 x, y에 대해 다음과 같은 정리가 성립한다.

**Theorem 10.2 (Conditional PDF of Multivariate Gaussian)** *If* $\mathbf{x}$ *and* $\mathbf{y}$ *are jointly Gaussian, where* $\mathbf{x}$ *is* $k \times 1$ *and* $\mathbf{y}$ *is* $l \times 1$, *with mean vector* $[E(\mathbf{x})^T \, E(\mathbf{y})^T]^T$ *and partitioned covariance matrix*

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{C}_{yy} \end{bmatrix} = \begin{bmatrix} k \times k & k \times l \\ l \times k & l \times l \end{bmatrix} \tag{10.23}$$

*so that*

$$p(\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^{\frac{k+l}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp\left[-\frac{1}{2}\left(\begin{bmatrix} \mathbf{x} - E(\mathbf{x}) \\ \mathbf{y} - E(\mathbf{y}) \end{bmatrix}\right)^T \mathbf{C}^{-1}\left(\begin{bmatrix} \mathbf{x} - E(\mathbf{x}) \\ \mathbf{y} - E(\mathbf{y}) \end{bmatrix}\right)\right],$$

*then the conditional PDF* $p(\mathbf{y}|\mathbf{x})$ *is also Gaussian and*

$$E(\mathbf{y}|\mathbf{x}) = E(\mathbf{y}) + \mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}(\mathbf{x} - E(\mathbf{x})) \tag{10.24}$$

$$\mathbf{C}_{y|x} = \mathbf{C}_{yy} - \mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}. \tag{10.25}$$

만약 두 확률 변수 x, y에 대해서라면,

$$C = \begin{bmatrix} var(x) & cov(x,y) \\ cov(y,x) & var(y) \end{bmatrix}$$

$$E(y|x) = E(y) + \frac{cov(x,y)}{var(x)}(x - E(x))$$

$$var(y|x) = var(y) - \frac{cov(x,y)^2}{var(x)}$$

와 같이 표현 가능하다.

이 때, MMSE $\hat{y}$는 $E(y|x)$이므로

$$\hat{y} - E(y) = \frac{cov(x,y)}{var(x)}(x - E(x))$$

$$\frac{\hat{y} - E(y)}{\sqrt{var(y)}} = \frac{cov(x,y)}{\sqrt{var(x)}\sqrt{var(y)}} \frac{\hat{x} - E(x)}{\sqrt{var(x)}}$$

와 같이 정규화된 표현으로 표현 가능하다. 이 때 $\frac{cov(x,y)}{\sqrt{var(x)}\sqrt{var(y)}}$는 상관 계수 $\rho$이므로

$$\hat{y}_n = \rho x_n$$

이 된다.
분산을 살펴보면,

$$\begin{aligned} var(y|x) &= var(y)\left(1 - \frac{cov(x,y)^2}{var(x)var(y)}\right) \\ &= var(y)(1 - \rho^2) \end{aligned}$$

가 되는데 var(y)는 x에 독립적이므로 추정값의 성능은 상관 계수에만 의존한다는 것을 알 수 있다.

proof)

*Matrix Theorem

$$\begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix}\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} I-A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22}-A_{21}A_{11}^{-1}A_{12} \end{bmatrix} \text{--------------(1)}$$

$$\det\left(\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\right) = \det(A_{11})\det(A_{22}-A_{21}A_{11}^{-1}A_{12}) \text{--------------------------(2)}$$

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} I-A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix}\begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & (A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \text{-------(3)}$$

posterior pdf $p(y|x)$는 다음과 같이 계산할 수 있다.

$$p(y|x) = \frac{p(x,y)}{p(x)}$$

$$= \frac{\dfrac{1}{(2\pi)^{\frac{k+l}{2}}\det^{\frac{1}{2}}(C)}\exp\left[-\frac{1}{2}\left(\begin{bmatrix} x-E(x) \\ y-E(y) \end{bmatrix}\right)^T C^{-1}\left(\begin{bmatrix} x-E(x) \\ y-E(y) \end{bmatrix}\right)\right]}{\dfrac{1}{(2\pi)^{\frac{k}{2}}\det^{\frac{1}{2}}(C_{xx})}\exp\left[-\frac{1}{2}(x-E(x))^T C_{xx}^{-1}(x-E(x))\right]}$$

이 때, 식(2)에 의해

$$\det(C) = \det(C_{xx})\det(C_{yy}-C_{yx}C_{xx}^{-1}C_{xy})$$

이므로

$$p(y|x) = \frac{1}{(2\pi)^{\frac{l}{2}}\det^{\frac{1}{2}}(C_{yy}-C_{yx}C_{xx}^{-1}C_{xy})}\exp\left(-\frac{1}{2}Q\right)$$

$$Q = \left(\begin{bmatrix} x-E(x) \\ y-E(y) \end{bmatrix}\right)^T C^{-1}\left(\begin{bmatrix} x-E(x) \\ y-E(y) \end{bmatrix}\right) - (x-E(x))^T C_{xx}^{-1}(x-E(x))$$

가 된다.

여기서 식 (3)에 의해 $C^{-1}$은 다음과 같다.

$$C^{-1} = \begin{bmatrix} I-C_{xx}^{-1}C_{xy} \\ 0 & I \end{bmatrix}\begin{bmatrix} C_{xx}^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -C_{yx}C_{xx}^{-1} & I \end{bmatrix}$$

$$B = C_{yy}-C_{yx}C_{xx}^{-1}C_{xy}$$

따라서 Q는 다음과 같이 정리된다.

$$\begin{aligned} Q &= \left(\begin{bmatrix} x-E(x) \\ y-E(y) \end{bmatrix}\right)^T C^{-1}\left(\begin{bmatrix} x-E(x) \\ y-E(y) \end{bmatrix}\right) - (x-E(x))^T C_{xx}^{-1}(x-E(x)) \\ &= \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix}^T \begin{bmatrix} I-C_{xx}^{-1}C_{xy} \\ 0 & I \end{bmatrix}\begin{bmatrix} C_{xx}^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -C_{yx}C_{xx}^{-1} & I \end{bmatrix}\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} - \tilde{x}^T C_{xx}^{-1}\tilde{x} \\ &= \begin{bmatrix} \tilde{x} \\ \tilde{y}-C_{yx}C_{xx}^{-1}\tilde{x} \end{bmatrix}^T \begin{bmatrix} C_{xx}^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix}\begin{bmatrix} \tilde{x} \\ \tilde{y}-C_{yx}C_{xx}^{-1}\tilde{x} \end{bmatrix} - \tilde{x}^T C_{xx}^{-1}\tilde{x} \\ &= (\tilde{y}-C_{yx}C_{xx}^{-1}\tilde{x})^T B^{-1}(\tilde{y}-C_{yx}C_{xx}^{-1}\tilde{x}) \\ &= [y-(E(y)+C_{yx}C_{xx}^{-1}(x-E(x)))]^T[C_{yy}-C_{yx}C_{xx}^{-1}C_{xy}]^{-1}[y-(E(y)+C_{yx}C_{xx}^{-1}(x-E(x)))] \end{aligned}$$

그러므로 $y|x \sim N(E(y|x), C_{y|x})$이며

$$E(y|x) = E(y) + C_{yx} C_{xx}^{-1}(x - E(x))$$

$$C_{y|x} = C_{yy} - C_{yx} C_{xx}^{-1} C_{xy}$$

라는 것을 알 수 있다.

*Matrix Inversion Lemma

Suppose $A_{11}$ and $A_{22}$ are invertible and let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$

$$\begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}$$

X Y Z = W

$Y^{-1} = Z W^{-1} X$

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix} \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix}$$

$$= \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \\ -(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{bmatrix}$$

$$\begin{bmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{bmatrix} = \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & 0 \\ 0 & A_{22} \end{bmatrix}$$

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{bmatrix} \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & 0 \\ 0 & A_{22}^{-1} \end{bmatrix} \begin{bmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & -(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} \end{bmatrix}$$

$$\begin{aligned} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} &= A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} \\ (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} &= A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \\ (A_{11} + A_{12}A_{22}^{-1}A_{21})^{-1} &= A_{11}^{-1} - A_{11}^{-1}A_{12}(A_{22} + A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} \\ (A_{11} + A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} &= A_{11}^{-1}A_{12}(A_{22} + A_{21}A_{11}^{-1}A_{12})^{-1} \end{aligned}$$

## Matrix Inversion Lemma

$$\begin{aligned} (A - BC^{-1}D)^{-1} &= A^{-1} + A^{-1}B(C - DA^{-1}B)^{-1}DA^{-1} \\ (A - BC^{-1}D)^{-1}BC^{-1} &= A^{-1}B(C - DA^{-1}B)^{-1} \\ (A + BC^{-1}D)^{-1} &= A^{-1} - A^{-1}B(C + DA^{-1}B)^{-1}DA^{-1} \\ (A + BC^{-1}D)^{-1}BC^{-1} &= A^{-1}B(C + DA^{-1}B)^{-1} \end{aligned}$$