• 08강 공분산 행렬

→ 공분산의 개념 + 행렬의 기하학적 의미

PCA 와 가우시안의 핵심적인 개념

— 공분산 행렬

x가 퍼진 정도

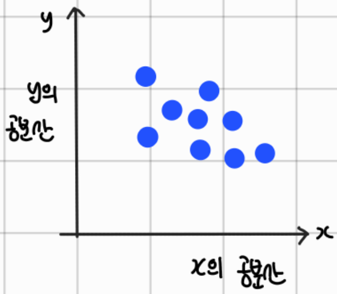
→ $cov(x, x) = V(x)$ // 자기자신의 cov는 분산

↱ x, y의 관계

$$COV = \begin{bmatrix} V(x) & cov(x, y) \\ cov(x, y) & V(y) \end{bmatrix}$$

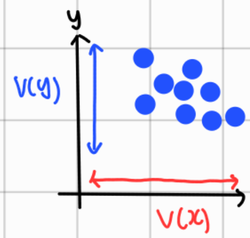
y가 퍼진 정보.

→ $cov(y, y) = V(y)$

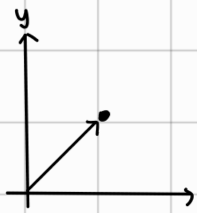


— 공분산 행렬의 기하학적 의미(1)

$$COV = \begin{bmatrix} V(x) & cov(x, y) \\ cov(x, y) & V(y) \end{bmatrix}$$



• 각각의 데이터에서 분산 뽑기



• 벡터(좌표)가 있을 때

• x는 V(x) 만큼

y는 V(y) 만큼 곱해서

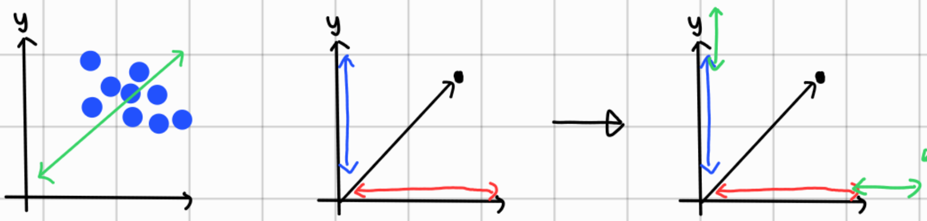
원래 벡터(좌표)를 mapping 해줌.

— 공분산 행렬의 기하학적 의미(2) :

x가 y에 영향주기

$$COV = \begin{bmatrix} V(x) & cov(x,y) \\ cov(x,y) & V(y) \end{bmatrix}$$

y가 x만큼 영향주기



cov는 음수도 가능하기 때문에 값이 늘어날지 줄어들지는 모르는 것.

- $V(x)$, $V(y)$ 때문에 늘어남
- 분산은 곱셈이라서 $V(x) \leq 1$ 일 때만 줄어들고 그 외에는 늘어나게 된다.

- x의 분산, y의 분산 그리고 서로의 관계까지 모두 포함한 데이터를 가지고 매핑하게 됨. 즉, 평범한 하나의 점이 두 데이터 간의 모든 정보를 반영해서 새로운 좌표가 되는 것

- 문제1) $X=[1,2,3]$, $Y=[2,3,1]$ 일 때, 공분산 행렬은?
↳ $m=2$ ↳ $n=2$

$$\sigma = \sqrt{\tfrac{2}{3}},\quad \sigma^2 = V(x) = \frac{2}{3},\quad V(y) = \frac{2}{3},\quad cov(x,y) = \frac{(1-2)(\cancel{2-2})^{0} + (\cancel{2-2})^{0}(3-2) + (3-2)(1-2)}{3} = -\frac{1}{3}$$

$$\therefore cov = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

-문제2) $(3,5)$는 공분산 행렬에 의해 mapping이 되면 어떻게 변하는가?

$$\begin{pmatrix} 3 & 5 \end{pmatrix} \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{pmatrix} = \begin{pmatrix} \frac{6}{3} - \frac{5}{3} & -\frac{3}{3} + \frac{10}{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{7}{3} \end{pmatrix}$$

↑ cov(x,y)