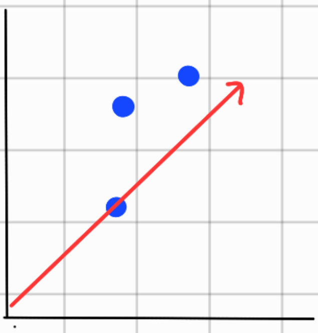
• 10강 _ 주성분 분석 (PCA)

Principal Componant Analysis
중요한 요소성분 분석

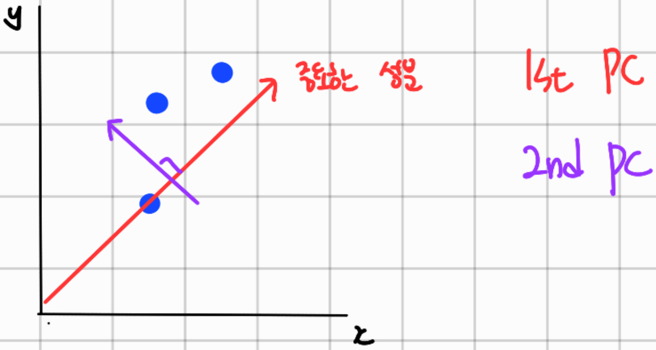
배경: 오버피팅문제를 해결해주기 위해



$$\underbrace{\begin{bmatrix} V(x) & cov(x,y) \\ cov(x,y) & V(y) \end{bmatrix}}_{\text{공분산 행렬}} \overset{\text{고유벡터}}{\begin{bmatrix} x \\ y \end{bmatrix}} = \underset{\text{고유값}}{\lambda} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\text{고유벡터}}$$

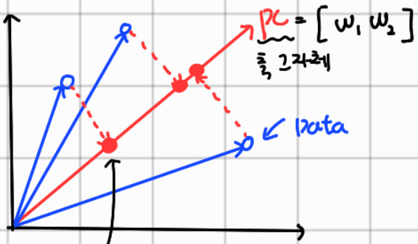
$\lambda \begin{bmatrix} x \\ y \end{bmatrix}$: 주성분 (PC, Principal Componene)

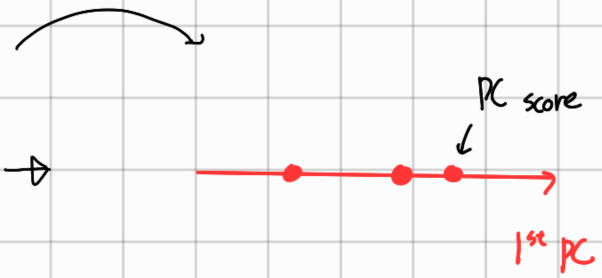
$$\lambda \begin{bmatrix} x \\ y \end{bmatrix} = \underset{\uparrow \ \text{가중치들}}{\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}}$$



- 기존 데이터를 PC에 정사영

2차원 Data를 옆으로 눕히면
1차원 Data가 됨.





PCscore : pc위에 올라간 값

원래의 값에서 내적하면 구할 수 있음.

$$\left[\begin{array}{cc} x & y \end{array}\right]\left[\begin{array}{c} w_1 \\ w_2 \end{array}\right] = w_1x + w_2y$$

식: $\left[\begin{array}{c} x \\ y \end{array}\right] = \left[\begin{array}{c} w_1 \\ w_2 \end{array}\right]$

2차원 $\xrightarrow[\text{정사영시킴}]{}$ 1차원

- PCA의 당위성

① **차원 축소** (PC score) → 차원을 축소하면 overfitting 문제가 해결됨. (주) 3차원 Data → 1차원으로 가능

$$[x\ y]\underbrace{\left[\begin{array}{c} w_1 \\ w_2 \end{array}\right]}_{\text{PC score 또는 가중치}} = w_1x + w_2y$$

② 데이터 의미 분석 & 새로운 변수 생성 (PC)

$z = w_1x + w_2y$

(예) x: 외모 0.5 ← 가중치
y: 성격 0.5

↓

z: 호감 이라는 Data 생성

- 문풀 ...

$x = [1, 3, 5]$, $y = [5, 6, 7]$

$m_x = 3$ $m_y = 6$

$V_x = \frac{8}{3}$ $V_y = \frac{2}{3}$ $\rightarrow$ $Cov = \frac{4}{3}$, $COV = \begin{pmatrix} \frac{8}{3} & \frac{4}{3} \\ \frac{4}{3} & \frac{2}{3} \end{pmatrix}$

고유벡터 $COV\vec{x} = \lambda\vec{x}$

$$\begin{pmatrix} \frac{8}{3}-\lambda & \frac{4}{3} \\ \frac{4}{3} & \frac{2}{3}-\lambda \end{pmatrix}\vec{x} = \lambda\vec{x}$$

$D=0$ 이되어야함.