

클린업 3주차

3팀 선형대수학

황정현
고경현
김지민
반경림
전효림

차원의 저주

차원
dimension

많은 정보를 가지고 있다

고차원일수록
좋다?

하지만!

변수가 많아질수록

overfitting(과적합)의 문제 발생 가능성 농후

변수 간 관련성이 높은 경우 공간 낭비

차원의 저주

변수 1

변수 1

고유값과 고유벡터

1 개념

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \quad n \times n \text{ 정방행렬}$$

eigenvector
고유벡터

$$A \neq 0 X = \lambda X$$

고유값
eigenvalue

$$Ax = \lambda x$$

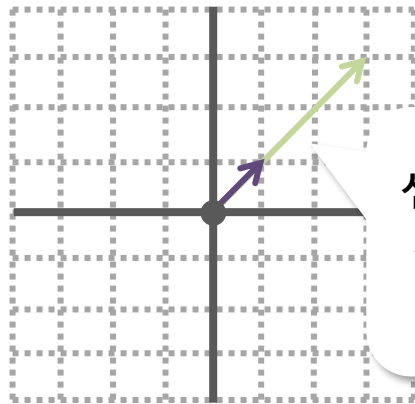
$$Ax - \lambda x = 0$$

$$(A - \lambda I)x = 0$$

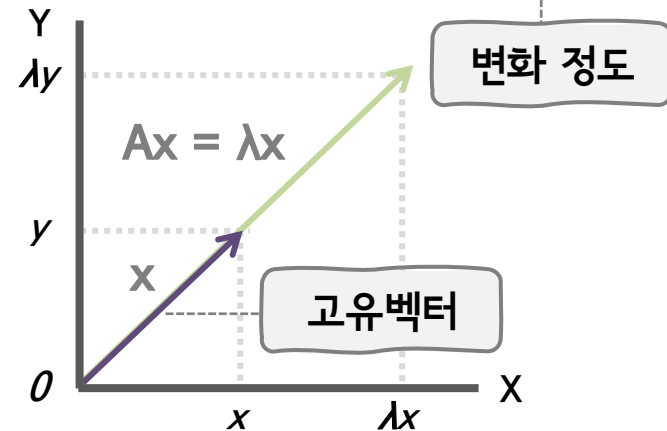
고유값과 고유벡터

3 선형변환의 관점

$$\begin{matrix} A & x \\ \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} & \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{matrix} = \begin{matrix} b \\ \begin{pmatrix} 3 \\ 3 \end{pmatrix} \end{matrix} = \begin{matrix} A & x \\ \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} & \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{matrix} = \begin{matrix} \lambda & x \\ 3 & \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{matrix}$$



선형변환을 취한 후에도
같은 직선 위에 놓이는
원점을 지나는 벡터



대각화와 고유값 분해

1 개념

$n \times n$ 정방행렬 A

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

고유벡터 행렬 P

$$\begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$$

행렬 A 의 선형독립인
고유벡터 n 개

고유값 대각행렬 D

$$\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

대각화

$$D = P^{-1} A P$$

고유값
분해

$$A = P D P^{-1}$$

공분산 행렬

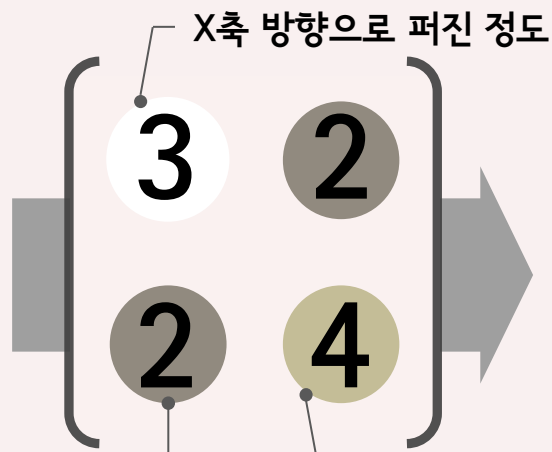
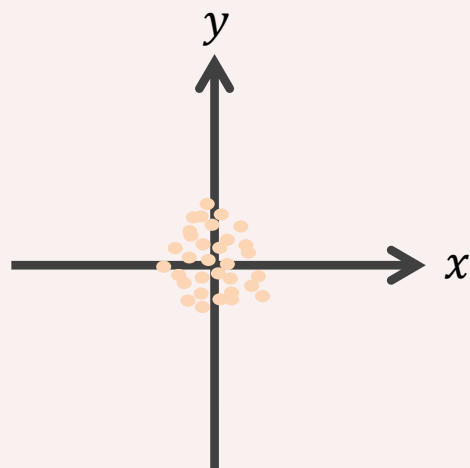
공분산 행렬

Covariance Matrix

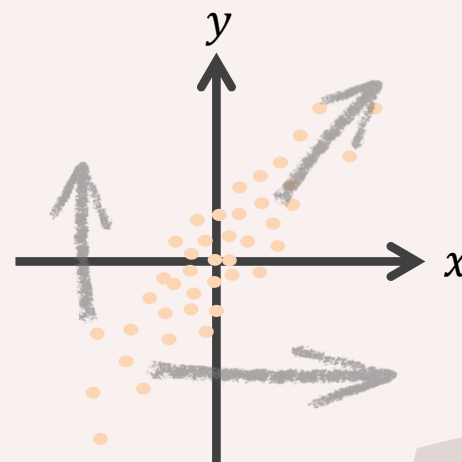
자료 간의 공분산을 행렬로 나타낸 것

~~~~~

선형변환으로 이해하기



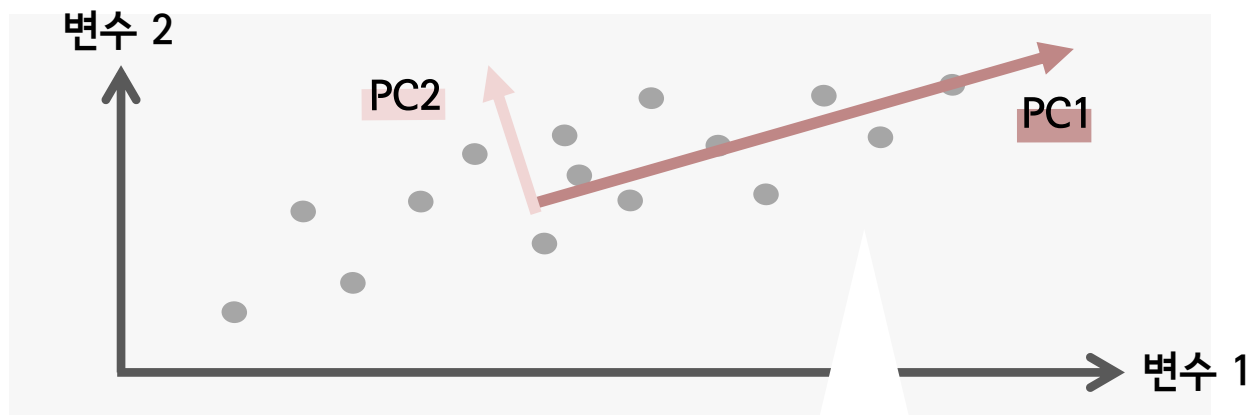
X, Y축 방향으로 함께 퍼진 정도



## 주성분 분석(PCA) 개념

*Principle Component Analysis*

데이터를 가장 잘 설명하는 주성분을 찾아내  
그 주성분이 이루는 공간으로 데이터를 **정사영**시켜 차원을 축소하는 방법



이 두 성분 중 데이터를 더 잘 설명하는 것은?



## 주성분 선택

PC 구하는 방법

공분산  
행렬

고유벡터 = PC

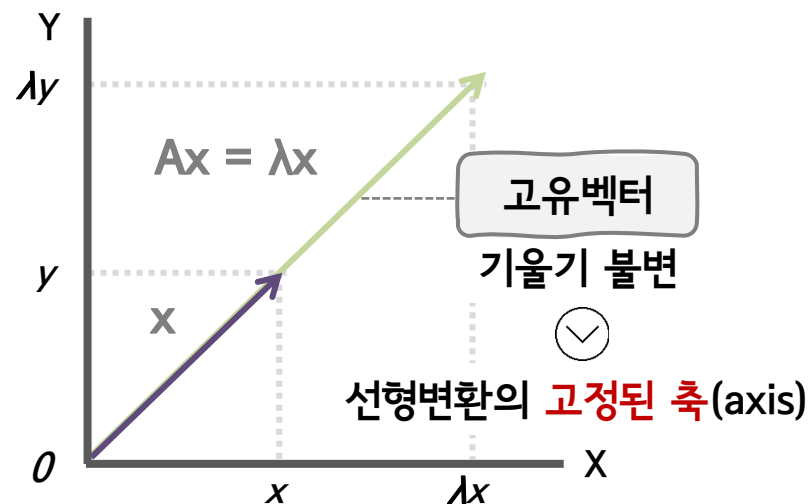
고유값 = 중요성

Q

고유벡터와 고유값을 사용하는 이유는 무엇인가요?

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} = 3 \begin{pmatrix} x \\ 1 \end{pmatrix}$$





## 특이값 분해(SVD) 이해하기

### 1 개념

- $m \times n$  크기의 일반적인 행렬에 대해 고유값 분해와 유사한 처리

$$\begin{array}{ccccccc}
 \begin{array}{c} n \\ m \end{array} \mathbf{A} & = & \begin{array}{c} m \\ m \end{array} \mathbf{U} & \times & \begin{array}{c} n \\ m \end{array} \mathbf{\Sigma} & \times & \begin{array}{c} n \\ n \end{array} \mathbf{V}^T \\
 m \times n \text{ 행렬} & & \text{선형변환 후의} & & \text{크기 변화} & & \text{선형변환 전의} \\
 & & m \times m \text{ 직교행렬} & & \text{나타내는} & & n \times n \text{ 직교행렬} \\
 & & & & m \times n \text{ 대각행렬} & &
 \end{array}$$

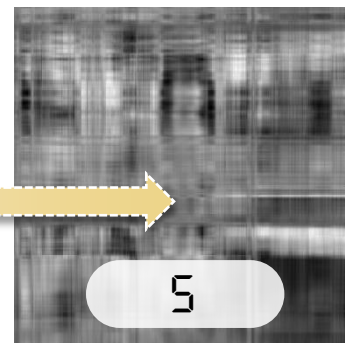
## 특이값 분해(SVD) 이해하기

## 3 이미지에 활용

정사각 사진의  
화질을 조정해보자!

$$\begin{array}{r} 2530 \\ \times \\ 2530 \end{array}$$

SVD를 통해  
적은 개수의  
**singular value**로  
사진 복원



정보량이 적어짐

## 개념

## 잠재요인분석

*Latent Semantic Analysis*

자연어 처리에서 문서 집합의 추상적 주제를 발견하고자  
사용하는 통계적 모델 (토픽모델링)



## 개념



인간의 의사결정은 계층적이고 상대적인 원칙을 따른다



## 계층화분석법

Analytic Hierarchy Process

의사결정문제가 다수의 평가 기준으로 이루어져 있을 때,  
평가 기준을 계층화한 뒤 이에 따라 중요도를 정해가는  
다기준 의사결정기법

쌍대비교의 반복

행렬을 이용한  
단계적 가중치 산정법

포인트



## 예시

| Charisma | 앙꼬  | 스누피 | K.K. |
|----------|-----|-----|------|
| 앙꼬       | 1   | 5   | 9    |
| 스누피      | 1/5 | 1   | 4    |
| K.K.     | 1/9 | 1/4 | 1    |

앙꼬가 K.K.보다  
9만큼 더  
카리스마 있다!

3번의 비교

Charisma



앙꼬



스누피



K.K.

Alternatives

## 예시

| 두 행렬의 곱 |     |   |   |     |     | 행간의 합 | 각 행의 비율 |       |   |      |
|---------|-----|---|---|-----|-----|-------|---------|-------|---|------|
| 1       | 5   | 9 | 1 | 5   | 9   | 53.24 | 0.75    |       |   |      |
| 1/5     | 1   | 4 | × | 1/5 | 1   | 4     | =       | 13.64 | > | 0.19 |
| 1/9     | 1/4 | 1 |   | 1/9 | 1/4 | 1     |         | 4.31  |   | 0.06 |

## 기준에 대한 가중치

|            |       |
|------------|-------|
| Experience | 0.547 |
| Education  | 0.127 |
| Charisma   | 0.270 |
| Age        | 0.056 |

∴ 최종 점수가 가장 높은 스누피가 바로 강아지대장!

|        | Experience | Education | Charisma | Age   | Goal  |
|--------|------------|-----------|----------|-------|-------|
| 양꼬     | 0.119      | 0.024     | 0.201    | 0.015 | 0.358 |
| 스누피    | 0.392      | 0.010     | 0.052    | 0.038 | 0.492 |
| K.K.   | 0.036      | 0.093     | 0.017    | 0.004 | 0.149 |
| Totals | 0.547      | 0.127     | 0.270    | 0.056 | 1.000 |
|        |            |           |          |       |       |
|        |            |           |          |       |       |

감사함멍

