

선형대수학팀

3팀
이재현
김규범
김민지
이정우
조혜현

INDEX

1. 주성분 분석(PCA)
2. 벡터 미적분학
3. 벡터 미적분학의 응용

차원의 저주 & 차원의 축소

차원

설명변수의 개수

Input 데이터에서 X 변수의 개수

차원 증가

 X 변수가 늘어남데이터를 더욱 잘 설명

공분산에 대한 선형대수에서의 해석

공분산

두 확률변수의 선형관계에 대한 정보를 주는 척도

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

$X = [x_1 \ x_2 \ \cdots \ x_n]^T$ 으로 구성된 행렬이며, x_i 가 각각 벡터일 때,

$$\text{공분산 행렬 } K_{XX} = \begin{bmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Cov}(X_n, X_n) \end{bmatrix}$$

$$= V(X_n)$$

공분산에 대한 선형대수에서의 해석

공분산 행렬을 통한 선형변환

$$\begin{bmatrix} V(X) & Cov(X, Y) \\ Cov(Y, X) & V(Y) \end{bmatrix}$$

분산 = 변수들이 퍼져 있는 정도
공분산 = 변수들의 공동적 움직임



분산과 공분산만큼
공간이 변화

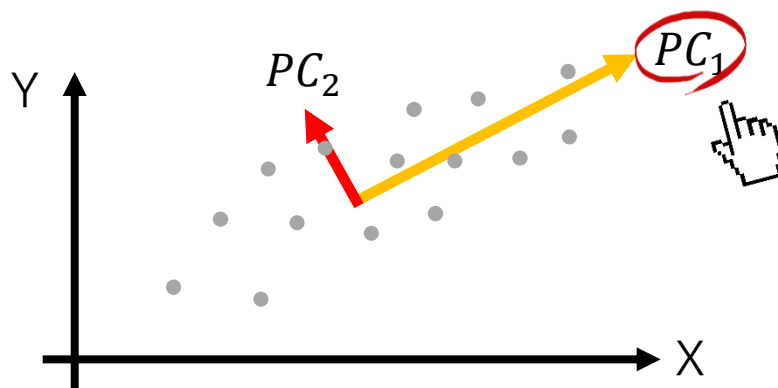
변수 간의 연관성
설명

변수들이 어떻게
분포됐는지 표현

주성분 분석 (PCA)

주성분 분석

데이터를 가장 잘 설명하는 주성분을 찾아 주성분이 이루는
공간으로 데이터를 정사영시켜 차원을 축소



PC_1 이 PC_2 보다 데이터가
더 많이 퍼져 있음

~~분산이 더 큰 주성분을 선택~~

주성분 선택



PCA결과의 설명되는 누적 분산량을 근거로 주성분의 개수 선택

	PC1	PC2	PC3	PC4
Standard Deviation	2.3925	0.44457	0.2522	0.0892
Proportion of Variance	0.8548	0.03294	0.0106	0.0034
Cumulative Proportion	0.8548	0.91273	0.9883	1.0000

일반적으로, 누적비율이 90%이상 되도록 선택

4차원 데이터 ➡ 2차원 데이터

미분과 Chain Rule

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x_1} \frac{\partial x_1}{\partial s} + \frac{\partial f}{\partial x_2} \frac{\partial x_2}{\partial s}$$

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x_1} \frac{\partial x_1}{\partial t} + \frac{\partial f}{\partial x_2} \frac{\partial x_2}{\partial t}$$



함수 f 가 $x_1(s, t)$, $x_2(s, t)$ 로 구성되어 있을 때,
 s, t 에 대한 f 의 도함수

미분과 Chain Rule

앞선 도함수를 행렬로 정리하면,

$$\begin{bmatrix} \frac{\partial f}{\partial s} & \frac{\partial f}{\partial t} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} \begin{bmatrix} \frac{\partial x_1}{\partial s} & \frac{\partial x_1}{\partial t} \\ \frac{\partial x_2}{\partial s} & \frac{\partial x_2}{\partial t} \end{bmatrix}$$



$$\begin{bmatrix} \frac{\partial f}{\partial s} & \frac{\partial f}{\partial t} \end{bmatrix} = \frac{\partial f}{\partial \vec{x}} \frac{\partial f}{\partial (s, t)}$$

Gradient란?

Gradient

- 어떤 함수 f 가 다중 변수 x_1, x_2, x_3, \dots 로 구성되어 있을 때,
- f 의 gradient 각 변수에 대하여 f 를 편미분 한 것을
행(또는 열)으로 나열한 것으로 표현
- 이때, gradient는 ∇f 혹은 $grad f$ 로 표기

Gradient란?



Gradient

x_1, x_2, x_3, \dots 로 구성된 어떤 함수 f 의 gradient

① 어떤 함수 f 가 다중 변수 x_1, x_2, x_3, \dots 로 구성되어 있을 때,

② f 의 gradient 각 변수에 대하여 f 를 편미분 한 것을

$$\nabla f = \text{grad } f = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \frac{\partial f}{\partial x_3} \quad \dots \right]$$

③ 이때, gradient는 ∇f 혹은 $\text{grad } f$ 로 표기

Gradient란?



Gradient

어떤 스칼라 함수 f 에 대해서

어떤 함수 f 가 다중 변수 x_1, x_2, \dots, x_n 에 의존되어 있을 때,
 gradient ∇f 의 방향은

f 의 gradient 각 변수에 대하여 f 를 편미분 한 것을
 f 가 가장 빨리 증가하는
 행(또는 열)으로 나열한 것으로 표현

이때, gradient ∇f 를 $grad f$ 로 표기
 방향을 가리킴


벡터함수란?

벡터함수
(Vector-Valued Functions)

$f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 과 벡터 $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ 로 주어졌다고 할 때,

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix} \in \mathbb{R}^m \text{으로 정의}$$

벡터함수의 Gradient

 실수함수 f_1, f_2, \dots 에 대해 $f = \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}$, 벡터 $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ 로 주어졌다고 하자.

벡터함수 f 를 특정 x_i 에 대해 편미분하면

$$\frac{\partial f}{\partial x_i} = \begin{bmatrix} \frac{\partial f_1}{\partial x_i} \\ \vdots \\ \frac{\partial f_m}{\partial x_i} \end{bmatrix}$$

벡터함수의 Gradient



앞선 식은 최종적으로 다음과 같이 전개됨

$$\frac{\partial f(x)}{\partial x} = \left[\frac{\partial f(x)}{\partial x_1} \quad \dots \quad \frac{\partial f(x)}{\partial x_n} \right] = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \dots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \dots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix}$$

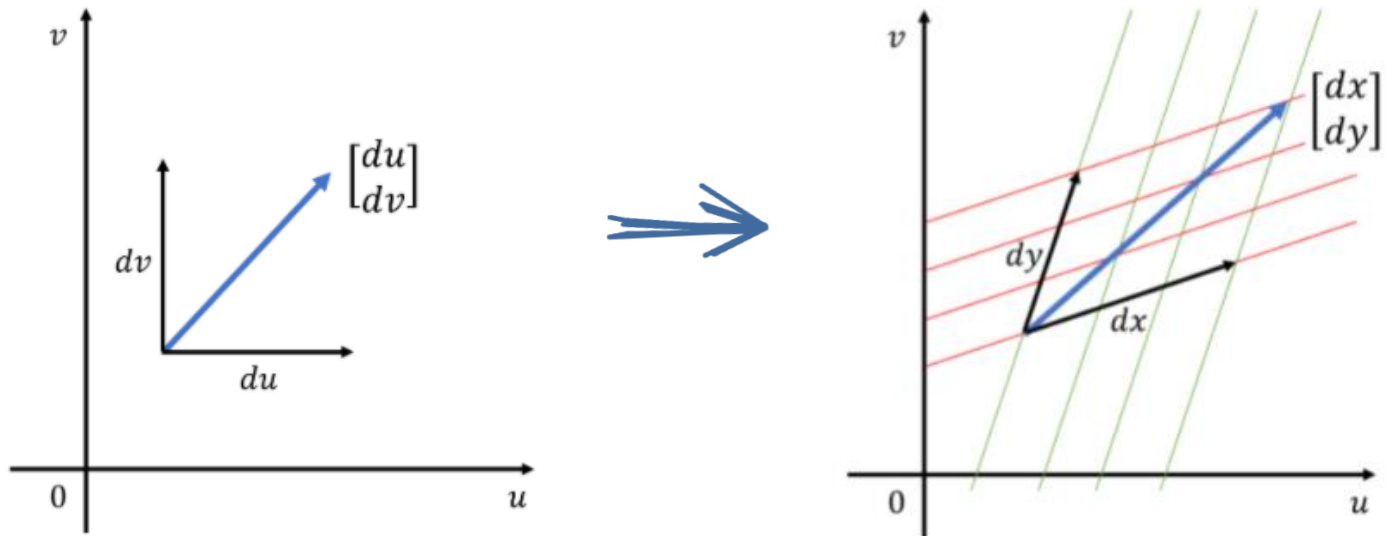
바로 위의 식을 벡터함수 f 의 gradient라고 정의하며,

$\nabla_x f$ 로 표기

자코비안(Jacobian)

자코비안

벡터함수 F 의 gradient
기하학적으로 **축 변환**의 역할을 수행

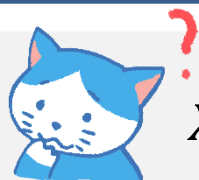


자코비안의 활용



수리통계학에서 축 변환을 위해 자코비안 행렬의 **행렬식** 사용
 자코비안의 행렬식은 **좌표축을 변환시켰을 때의 크기 변화**를 의미

$$f_{y_1, y_2}(y_1, y_2) = |J| f_{x_1, x_2}(g(y_1, y_2), h(y_1, y_2))$$



X의 pdf를 통해 Y의 pdf를 구해보자

독립변수 X_1, X_2 에 대해 $Y_1 = X_1, Y_2 = X_1 + X_2$ 일 때,

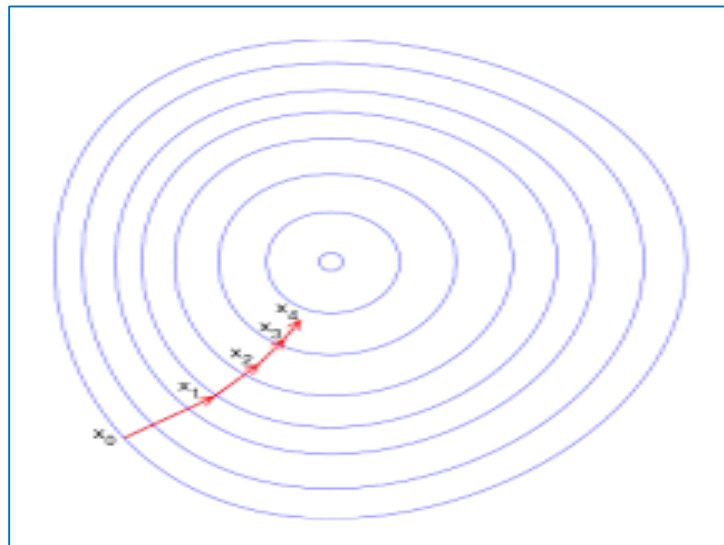
$f_{y_1, y_2}(y_1, y_2) = |J| f_{x_1, x_2}(g(y_1, y_2), h(y_1, y_2))$ 식에 대입

$$f_{y_1, y_2}(y_1, y_2) = |J| f_{x_1, x_2}(y_1, -y_1 + y_2) = f_{x_1, x_2}(y_1, -y_1 + y_2)$$

Gradient Descent

Gradient Descent

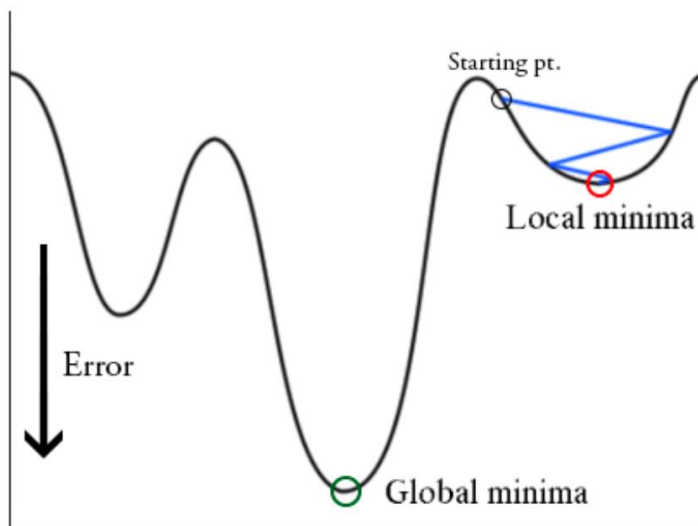
등고선에서 가장 빠르게 감소하는 방향으로 이동
Loss값을 최소화하기 위해 gradient에 (-1) 을 곱해서 알고리즘 운영



Gradient Descent

Gradient Descent

등고선에서 가장 빠르게 감소하는 방향으로 이동
Loss값을 최소화하기 위해 gradient에 (-1) 을 곱해서 알고리즘 운영



gradient의 값이
0이 되는 지점에서 멈춤
→ 하지만 Gradient Descent 방법은
local minimum에 빠질 수 있고
속도가 느리다는 단점 존재