

공분산·PCA·Mahalanobis 거리

학번 _____

이름 _____

오늘의 목표 데이터를 중심화하고 공분산의 고유축으로 차원축소와 표준화된 거리를 계산한다.

● 핵심 개념 빈칸 노트

- ☐ 데이터에서 평균을 빼는 과정을 _____ 라 한다.
- ☐ 공분산행렬은 대칭이고 방향별 _____ 을 담는다.
- ☐ PCA의 첫 주성분은 투영분산을 가장 크게 하는 공분산행렬의 _____ 이다.
- ☐ Mahalanobis 거리 제곱은 $d_M^2 = (x - \mu)^T$ _____ $(x - \mu)$ 이다.

따라 풀기

공분산의 주축

평균이 0인 네 점 $(2, 0), (-2, 0), (0, 1), (0, -1)$ 의 공분산을 $1/n$ 기준으로 구하자.

1단계 x 성분 제곱합은 8, y 성분 제곱합은 2, 교차곱 합은 0이다.

2단계 $\Sigma =$ _____ 이다.

3단계 가장 큰 분산 방향은 고윳값 2의 고유벡터 _____ 이다.

실수 방지 공분산을 계산하기 전에 평균을 빼야 하며, 수업에서 $1/n$ 과 $1/(n-1)$ 중 어느 정의를 쓰는지 확인한다.

공분산·PCA·Mahalanobis 거리

● 연습 문제 · 쉬운 계산에서 심화까지

1 기초

데이터 2, 5, 8의 평균과 중심화한 값을 구하시오.

2 기초

$\Sigma = \text{diag}(9, 1)$ 에서 분산이 가장 큰 주성분 방향을 쓰시오.

3 개념

PCA에서 데이터를 중심화하지 않으면 첫 방향이 무엇의 영향을 받을 수 있는지 설명하시오.

4 적용

점 (1, 1), (3, 3), (5, 5)의 평균과 $1/n$ 공분산행렬을 구하시오.

5 적용

앞 데이터의 첫 주성분 방향과 대응 고유값을 구하시오.

6 심화

$\mu = 0$, $\Sigma = \text{diag}(16, 1)$ 에서 $x = (4, 0)$ 과 $y = (0, 2)$ 의 유클리드 거리와 Mahalanobis 거리를 비교하시오.

나가기 전 한 문장 PCA와 Mahalanobis 거리가 공분산행렬을 서로 어떻게 다르게 사용하는지 쓰시오.

☐ 개념을 설명할 수 있다 ☐ 계산을 다시 확인했다 ☐ 질문이 있다