

Chapter 8 차원 축소

8.1 훈련 세트의 차원이 클수록 (과대적합/과소적합)의 위험이 커진다.

답: 과대적합

8.2 매니폴드 가정(매니폴드 가설)이란 무엇인가요?

답: 실제 고차원의 데이터셋이 더 낮은 저차원 매니폴드에 가깝게 놓여 있다는 가정이다.

8.3 점진적 PCA 알고리즘에 대해서 설명하세요.

답: 훈련세트를 미니배치로 나눈 뒤 IPCA 알고리즘에 한 번에 하나씩 주입한다. 이런 방식은 훈련 세트가 클 때 유용하고 온라인으로 PCA를 적용할 수 있다.

8.4 kPCA에서 커널과 하이퍼파라미터를 선택하는 방법 2가지에 대해 설명하세요.

답: 그리드 탐색을 사용하여 주어진 문제에서 성능이 가장 좋은 커널과 하이퍼파라미터를 선택할 수 있다. 또 다른 방법은 완전한 비지도 학습 방법으로 가장 낮은 재구성 오차를 만드는 커널과 하이퍼파라미터를 선택하는 방식이다.

8.5 LLE가 작동하는 방식에 대해서 간단히 설명하세요.

답: 각 훈련 샘플이 가장 가까운 이웃에 얼마나 선형적으로 연관되어 있는지 측정하고 국부적인 관계가 가장 잘 보존되는 훈련 세트의 저차원 표현을 찾는다.

8.6 다음 설명에 맞는 차원 축소 기법을 쓰시오.

(1) 샘플 간의 거리를 보존하면서 차원을 축소한다. ()

(2) 비슷한 샘플은 가까이, 비슷하지 않은 샘플은 멀리 떨어지도록 하면서 차원을 축소한다. 주로 시각화에 많이 사용되며 특히 고차원 공간에 있는 샘플의 군집을 시각화할 때 사용된다.

()

답: 다차원 스케일링(MDS), t-SNE