1. **차원축소의 필요성**

* 고차원 데이터는 계산량이 많고, 과적합의 위험이 높으며, 샘플 간 거리가 멀어지는 *차원의 저주(Curse of Dimensionality)* 현상이 발생함.
* 차원축소는 데이터를 저차원으로 변환하여 학습 가능성을 높이고, 시각화 및 저장 효율을 향상시킴.

1. **차원축소의 종류**

* **사영 기반 기법(Projection)**: 데이터를 직선적으로 낮은 차원에 투영.
* **다양체 학습(Manifold Learning)**: 고차원 데이터가 실제로는 저차원 곡면(다양체)에 존재한다는 가정하에 구조를 보존하면서 차원 축소.

1. **차원의 저주 해결**

* 차원이 증가할수록 데이터 간 거리가 멀어지고 일반화가 어려워짐.
* 해결책: 차원축소 또는 학습 데이터 수를 급격히 늘려야 하나, 후자는 현실적으로 어렵기 때문에 전자가 실질적 대안.

1. **PCA (Principal Component Analysis)**

* 데이터의 분산을 최대한 보존하는 방향(주성분)으로 데이터를 사영하는 기법.
* **첫째 주성분**: 분산 최대 보존 축.
* **둘째 주성분 이후**: 앞선 주성분들과 직교하며 남은 분산을 최대 보존.
* 특잇값 분해(SVD)로 계산 가능하며, Scikit-learn의 PCA 모델로 쉽게 구현 가능.

1. **설명 분산 비율과 차원 결정**

* 각 주성분이 설명하는 분산의 비율을 이용하여 적절한 차원 수 결정.
* 일반적으로 누적 분산 95%를 기준으로 주성분 개수를 선택함.
* 예시: MNIST 데이터셋 → 784차원 → 154차원으로 줄여도 95% 분산 유지 가능.

1. **PCA 응용**

* **데이터 압축**: 이미지 등에서 정보 손실 없이 저장 용량 감소 가능.
* **시각화**: 고차원 데이터를 2~3차원으로 축소해 군집 구조 등을 시각적으로 파악 가능.

1. **랜덤 PCA & 점진적 PCA**

* **랜덤 PCA**: SVD 계산을 근사적으로 처리해 빠르게 주성분 추출.
* **Incremental PCA (IPCA)**: 미니배치 단위로 데이터 처리 → 온라인 학습에 유용.

1. **임의 사영 (Random Projection)**

* **존슨-린덴슈트라우스 정리**에 기반: 임의로 사영해도 정보 손실이 크지 않음.
* **Gaussian / Sparse Random Projection**: 빠르고 메모리 효율적, 대규모 데이터에 적합.

1. **국소적 선형 임베딩 (LLE)**

* **비선형 다양체 학습 기법**의 대표주자.
* 전체 구조는 비선형이나, 국소적으로는 선형이라는 가정 하에 저차원 표현을 학습.
* 사영이 아닌 **국소 구조 보존**에 집중.

1. **기타 차원축소 기법**

* **MDS**: 거리 보존에 중점.
* **Isomap**: 거리 기반 다양체 복원.
* **t-SNE**: 고차원에서의 유사도를 저차원에 보존.
* **LDA**: 클래스 간 분산 최대화.
* **커널 PCA**: 비선형 구조를 커널 기법으로 처리.