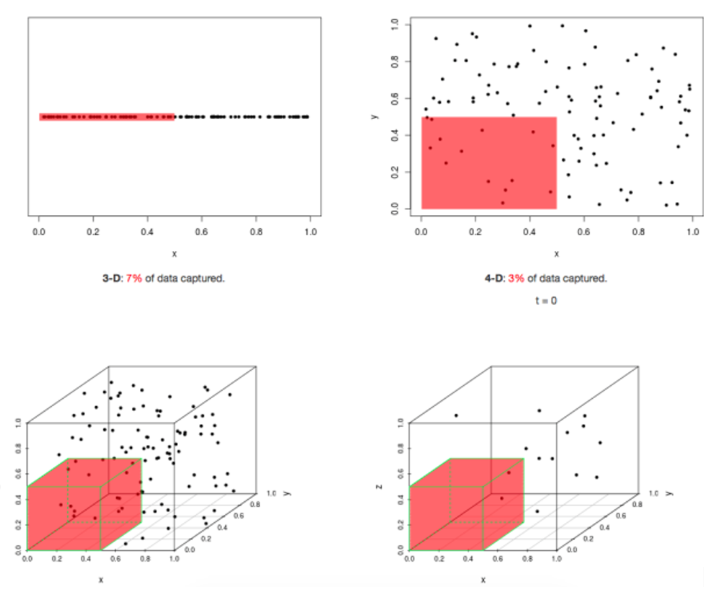
차원 축소

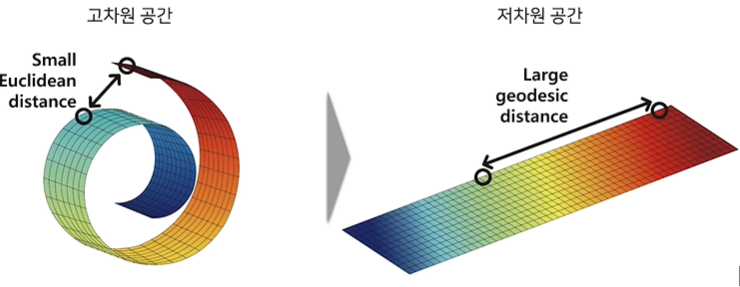
**차원의 저주** – 샘플 수가 늘어날수록 학습이 어려워지며 고차원 공간의 사상이 어려움.



고차원 데이터셋의 경우, 학습 난이도가 증가함. 이에 대비해 특성 수를 줄임으로써 차원 축소시켜 학습 가능하도록 함. – {데이터 유실, 데이터 시각화}

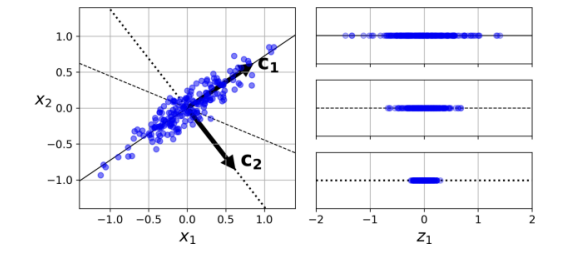
**투영** – n>d 일 때, n차원 공간의 존재하는 d차원 공간으로 투영해 표현

**매니폴드 학습** – n차원의 데이터를 n-1차원의 평면으로 축소하는 학습법으로 데이터를 잘 아우르는 sub-space(manifold)를 구하고 데이터를 투영시킴으로 차원 축소를 구현한다.



차원 축소 진행 시 분산 고려사항

**분산보존** – 투영 시에 가장 분산이 큰 축을 선택해야 데이터 손실이 가장 적음.



**PCA (주성분 분석)** – 훈련 세트 내에서 분산을 최대한 보존하는 축(특성)으로 SVD 사용해 탐색

차원의 데이터를 저차원의 데이터로 축소시키는 차원 축소 방법 중 하나로 여러 특성 중 중요한 몇가지만을 선택하는 기법

PCA 종류: 랜덤 PCA / 점진적 PCA / 커널 PCA

**차원축소를 하는 이유는?**

고차원으로 구성되어 있는 데이터셋은 시각화에 큰 어려움을 겪기 때문에 차원축소를 진행함으로 데이터 패턴을 인지하고 3차원 이내의 데이터로 시각화를 진행하기 위함이다.

또한, 특성을 줄이는 과정에 있어 중요한 특성만 남게 되므로 분산을 유지하면서도 불필요한 노이즈를 크게 줄이며 모델의 퍼포먼스 향상에 크게 기여하게 된다.

전반적으로 모델의 성능을 향상시키는 방향으로 메모리 사용률 감소, 데이터 압축, 시각화로 인한 데이터 패턴 분석 용이를 일궈낸다.