

# UMAP (Uniform Manifold Approximation)

• 거리  $\times$  확률 + 토폴로지 : 고차원 그래프를 생성  $\rightarrow$  저차원 매핑

• 특징 - 국소 구조 보존 : 가까운 데이터 포인트 간의 국소적인 유사성을 유지

- Topological Preservation

- Simplicial Complexes : 원의 반지름

너무 작으면, 작고 고립된 클러스터 생성  
너무 크면, 모든 게 다 연결  
 $\Rightarrow$  knn 으로 포인트마다 반지름을 선택해서 해결

- 높은 확장성과 효율성

• 장점 - 속도 현저히 빠름

- embedding 차원 크기에 대한 제한이 없어서 일반적인 차원 축소 알고리즘으로 적용 가능함

- 파라미터들 이해하기 쉬움

min\_dist { 작으면, 포인트들이 촘촘하게 무리지어  
크면, 포인트들이 느슨하게 퍼져  
n\_neighbors { 작으면, local structure에 집중  
크면, global structure에 집중

- 탄탄한 이론적 배경을 가지고 있음 : 리만 기하학 & 위상 수학에 기반

• Loss Function { 근접성 보존 부분 : comparable distance  
균질성 부분

## PaCMAP (Pairwise Controlled Manifold Approximation Projection)

• 거리  $\circ$  : 이웃 그래프 구성  $\rightarrow$  부분 대응 분석  $\rightarrow$  저차원 매핑

• 특징 - 부분 대응 관계 모델링 : 데이터 포인트 간의 상대적인 유사도를 고려

- 데이터 노이즈 대응 : 유용한 정보 유지

- 알고리즘의 효율성 : 빠른 처리 속도를 제공

• 장점 - Computational Time

- Structure Preservation { global : similar to TriMap  
local : similar to UMAP & t-SNE

• Loss Function { Near pairs를 서로 가깝게 하는 부분  
Mid-near pairs를 약하게 가깝게 하는 부분 : 데이터의 밀도 조정  
Further pairs를 멀게 하는 부분

