

UMAP

UMAP은 fuzzy topological structure를 사용하는 manifold learning이다.

다음의 세 가지 가정을 바탕으로 하고 있다.

1. 데이터가 리만 매니폴드 상에서 uniformly distributed되어 있다.
2. Riemannian metric이 locally constant하다.
3. 매니폴드는 국소적으로 연결되어 있다.

알고리즘 단계

1. 퍼지 위상 표현을 구성
2. cross entropy로 측정된 퍼지 위상 표현에 최대한 가깝게끔 최적화

주요 파라미터

1. n_neighbors: local structure와 global structure 간의 balance를 결정. 데이터의 매니폴드 구조를 학습할 때 몇 개의 이웃을 살펴볼 것인가.
2. min_dist: 저차원 표현에서 점들 간의 최소 거리.
3. n_components: 몇 차원으로 임베딩할 것인지 결정
4. metric: input data의 ambient space에서의 거리 계산 척도

PaCMAP

Reference

<https://umap-learn.readthedocs.io/en/latest/>