

## 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

지역 구조와 전역 구조를 모두 보존하기 위해 고안되었다. neighbor pairs, mid-near pair, further pairs 세 종류의 쌍을 사용하여 저차원 임베딩을 최적화한다. 전역 구조를 보존하기 위해 더 많은 이웃 데이터를 고려하기보다 mid-near pair를 사용하여 먼저 전역 구조를 캡처한 뒤 지역 구조를 refine한다.

### 주요 파라미터

1. n\_components: 축소 결과 차원
2. n\_neighbors: KNN 그래프 상에서 고려하는 이웃의 수.
3. MN\_ratio: 이웃 수에 대한 mid-near pairs의 비율.  $n_{MN} / n_{neighbors}$
4. FP\_ratio: 이웃 수에 대한 further pairs의 비율.  $n_{FP} / n_{neighbors}$

### Reference

<https://umap-learn.readthedocs.io/en/latest/>  
<https://github.com/YingfanWang/PaCMAP>