**UMAP과 PaCMAP**

UMAP과 PaCMPA은 각각 데이터의 복잡한 구조를 저차원으로 잘 보존하는 동시에 시각화를 지원하는 알고리즘이다. UMAP은 Uniform Manifold Approximation and Projection의 약어로 데이터의 차원 축소와 시각화를 위한 비선형적 기법이다. 기존의 차원 축소 기법들은 데이터의 복잡한 구조를 보존하지 못하는 경우가 많았고 선형적인 변환만을 고려하여 비선형 구조를 가진 데이터의 표현에 어려움을 겪었다. 이에 따라 UMAP은 비선형구조를 잘 보존하는 차원 축소기법으로써 주목을 받았다. 또한, UMAP은 데이터의 복잡한 구조를 보존하면서 저차원 공간으로 매핑하는데 사용된다. UMAP의 주요 아이디어는 데이터 포인트 간의 유사도를 보존하는 방식으로 저차원으로 매핑함으로써 데이터의 구조를 보다 잘 유지하는 것에 있다.

UMAP의 작동원리 3가지는 다음과 같다. 1) 근접이웃찾기. 각 데이터 포인트 주변의 근접 이웃을 찾는다. 이는 데이터 포인트 간의 거리 또는 유사도를 측정하여 이루어진다. 2) 이웃 유사도 보존. 근접 이웃 간의 유사도를 저차원으로의 매핑과 비교하여 보존한다. 3) 저차원 매핑 학습. 저차원 공간에서 데이터 포인트 간의 유사도를 최대한 보존하도록 저차원으로의 매핑을 조정한다. UMAP의 장점은 비선형 구조를 잘 보존하면서도 시각화 가능한 저차원 표현을 제공함에 있다. 또한 대용량 데이터셋에서도 적용 가능하며, 빠른 속도와 확장성을 보장한다.

PaCMAP은 Parameter-free Convergence-aware Manifold Approximation and Projection의 약어로 UMAP과 유사한 차원축소와 시각화 알고리즘이다. 그러나 PaCMAP은 UMAP의 하이퍼파라미터에 의존하지 않고도 더 빠른 수렴 속도를 제공하려는데 초점을 맞추었다. PaCMAP의 특징과 작동원리는 다음과 같다. 1) 파라미터 프리(Parameter-free): PaCMAP은 사요자가 조정해야 하는 하이퍼파라미터 없이도 최적의 결과를 얻을 수 있도록 설계되었다. 2) 수렴 속도 향상: PaCMAP은 초기 매핑에서 데이터 포인트 간의 거리와 근접 이웃을 활용하여 더 빠른 수렴을 도모한다. 3) 확장성: 대량의 데이터에 대해서도 빠른 처리 속도와 우수한 품질을 제공한다. 그러나 PcMAP도 몇 가지 한계점을 가지고 있다. 그 한계는 5가지로 축약할 수 있다. 1) 샘플 수와 데이터의 차원에 민감한 영향, 2) 최적화 과정의 복잡성, 3) 데이터 구조의 복잡성 처리의 한계, 4) 제약된 하이퍼파라미터 없이 모든 경우에 부적합, 5) 데이터 노이즈에 민감. 즉, PaCMAP은 빠른 수렴 속도와 비선형 구조를 잘 보존하지만, 이 역시도 모든 상황에서 완벽한 결과를 제공하는 것이 아니므로 데이터셋의 특성과 사용자의 요구에 따라 적절하게 활용되어야 한다.