2019150445 신백록

1. 188쪽의 wine data를 이용하여 MDS, local MDS, LLE, Isomap, t-SNE의 성능을 비교하라.

Wine data의 전체에 대해서 logistic regression을 실시하면 accuracy가 1이 나온다.

텍스트이(가) 표시된 사진

자동 생성된 설명

MDS로 training을 하면 91%의 acc가 나온다. 원래 데이터가 13개의 특성 변수를 가지고 있었던 것과 비교하면 특성 변수가 2개뿐인 MDS가 꽤 많은 정보를 보존하고 있다는 것을 알 수 있다. 원래 데이터가 나름의 선형성을 가지고 있어서 유클리디안 거리를 계산하는 MDS에서 괜찮은 결과가 나온 것이다. Plot을 해봐도 몇몇 group2 점들을 제외하곤 나름 괜찮게 separation 되어있다.

텍스트이(가) 표시된 사진

자동 생성된 설명

다음으론 k-nearest neighbors를 이용하는 LLE로 차원축소를 진행하였다. Neighbor 숫자에 영향을 많이 받기에 for문을 통하여 차원축소 후 logistic regression 결과가 가장 괜찮은 neighbor을 결정했고, 67개의 neighbor로 대략 90%의 accuracy가 나온다.

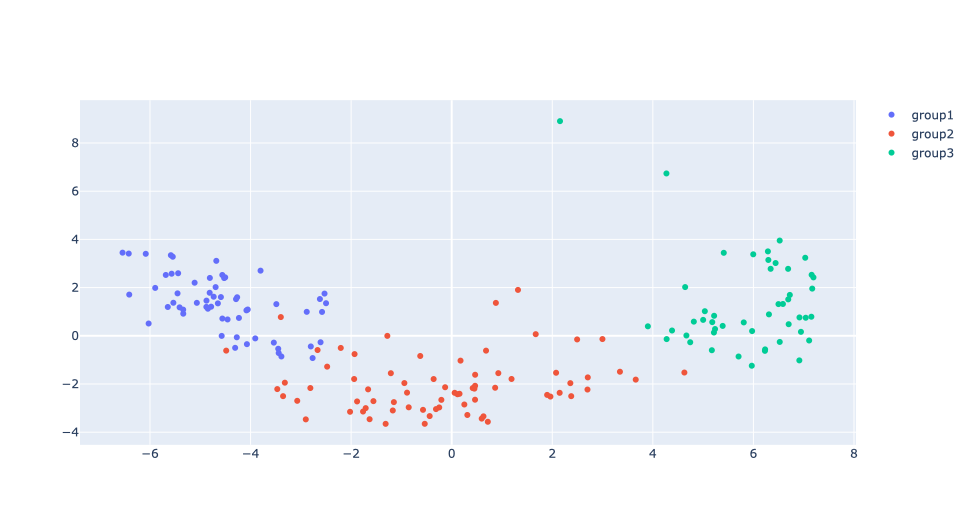
텍스트이(가) 표시된 사진

자동 생성된 설명

다음엔 지도면상의 최소거리를 이용하는 Isomap으로 차원축소를 하였다. 마찬가지로 for 문을 통하여 neighbor을 결정하였고, 21개의 nearest neighbor로 차원축소를 진행하는 것이 성능이 가장 괜찮게 나왔다. 무려 98%의 accuracy가 나왔다. Plot을 해서 보면 세 그룹이 잘 separation되어있는 것을 볼 수 있다. 큰 정보의 손실 없이 13개의 feature을 2개의 feature로 차원축소가 잘 된 것이다.

텍스트이(가) 표시된 사진

자동 생성된 설명



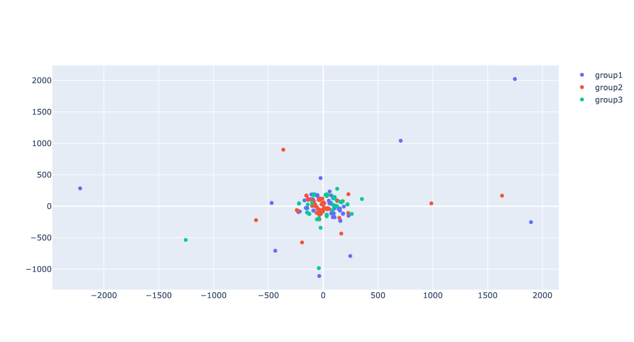
마지막으로 t-sne를 이용하여 차원축소를 진행하였고, perplexity=39인 데에서 대략 99%의 accuracy가 나왔다. T-sne가 classification을 위한 정보를 가장 잘 보존하였다.

텍스트이(가) 표시된 사진

자동 생성된 설명

2. t-SNE에서 perplexity의 영향을 논의하라.

Perplexity=1인 경우에는 차원축소가 일어나긴 하지만 아무런 정보를 얻을 수 없을 정도로 데이터가 몰려있는 것을 볼 수 있다. Logistic regression을 시행한 뒤의 Accuracy도 40%정도 나온다.



여기서 perplexity를 1만 올려도 90%의 accuracy가 나오며 Class별로 분류가 조금씩 되는 것이 보인다.



Perplexity=10정도부터 t-sne로 차원축소한 데이터의 분포가 차원축소 전의 데이터의 분포를 잘 보존하고 있어 plot과 acc 모두 준수하게 나오는 것 같다.

