

LLE(Locally Linear Embedding)의 함수관계에 대한 다층퍼셉트론 학습

Training of Locally Linear Embedding using Multilayer Perceptrons

오상훈
목원대학교

Sang-Hoon Oh
Mokwon Univ.

요약

고차원의 데이터를 저차원으로 차원축소 하는 것은 일반적인 문제에서 항상 나타난다. 이때, 고차원에서 데이터 간의 인접한 관계를 유지하면서 저차원으로 변형시켜주는 방법으로 LLE(Locally Linear Embedding)이 제안되었다. 이 방법은 비록 최적의 해를 찾아주지만, 학습되지 않은 데이터가 주어지면 다시 전체 데이터를 상대로 처리를 하여야 한다. 이 논문에서는, 주어진 학습데이터만을 상대로 LLE의 관계를 수행할 수 있는 다층퍼셉트론을 학습시켜, 학습되지 않은 데이터가 입력되는 경우 다층퍼셉트론의 출력으로 LLE 처리를 하는 방법을 제안한다.

Abstract

LLE(Locally Linear Embedding) has been proposed to compute low dimensional, neighborhood preserving embeddings of high dimensional data. Here, we should perform whole processes of LLE when untrained patterns are presented. In this paper, we propose a training of MLPs(Multilayer Perceptrons) to perform the mapping of LLE from high dimensional data to low dimensional ones.

I. 서론

자연현상에서 다루는 문제들은 대부분 아주 고차원의 데이터들이다. 이런 데이터들을 처리하여 유용한 정보를 얻어내는 다음 단계의 처리를 위해서는 먼저 고차원의 데이터를 저차원 데이터로 변형시키는 과정이 우선적으로 수행되어야 한다. 약 3만개의 청신경섬유와 백만 개의 시신경섬유를 통한 정보를 처리하는 사람 역시 이러한 과정을 거치는 것으로 여겨진다. 이러한 차원 축소는 여러 형태로 제안되었다.

ICA(Independent Component Analysis)는 일종의 Project Pursuit Learning으로써, 데이터들이 고유하게 간직하고 있는 구조를 찾아내는 방법으로 해석될 수 있다[1]. 이 방법은 처리 결과 얻어지는 데이터들의 p.d.f.(probability density function)가 서로 독립이 되도록 하는 부분에 초점이 맞추어져 있다. 또한, NMF(Nonnegative Matrix Factorization) 역시 고차원의 데이터를 저차원의 데이터로 변형시키는 과정 중에 특정 부위에 대한 특징들을 추출해내는 방법으로 제안되었다[2].

이들 방법과 달리 고차원 데이터들의 통계적인 특성과 기하학적 구조를 활용하는 방법으로 Isomap과 LLE(Locally Linear Embedding)이 제안되었다. Isomap은 고차원 상의 데이터들의 pariwise distance가 유지되도록 저차원 데이터로

변형시키는 방법이며[3], LLE는 고차원 상의 neighborhood 관계가 저차원에서 유지되도록 하는 방법이다[4].

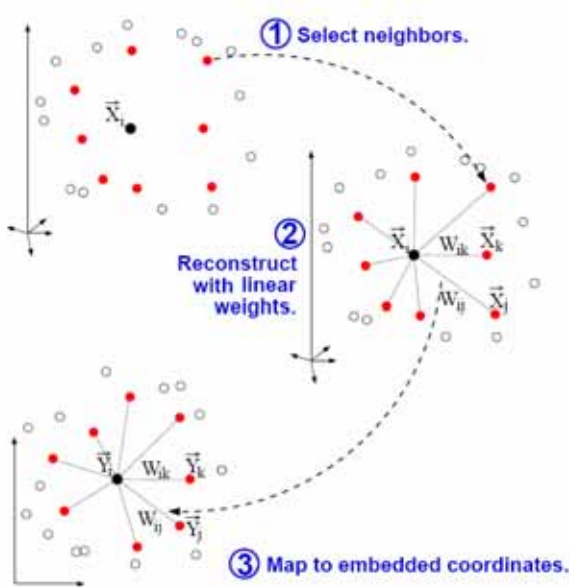
이 논문은 LLE에 대하여 초점을 맞추었다. LLE는 고차원의 입력 데이터에 대하여 회전, 스케일 변경, 및 이동에 영향을 받지 않는 차원 축소 방법으로, eigenvector에 근거한 방법이다[4]. 이 경우, 주어진 데이터에 대한 분석을 통하여 차원이 축소된 데이터로 변환을 시켜주는 데, 이 분석에 포함되지 않는 데이터가 주어진 경우에는 전체 데이터에 대하여 분석을 새롭게 수행하여야 하는 단점이 있다.

한편, 다층퍼셉트론은 충분한 수의 중간층 노드가 주어지면 임의의 함수를 학습을 통하여 근사화시킬 수 있다. 즉, 주어진 고차원 데이터에 LLE를 적용시켜 차원이 축소된 데이터가 얻어진 경우, 이들 데이터들을 다층퍼셉트론에 학습시키면, 이 다층퍼셉트론은 LLE의 기능을 하는 함수를 근사화한 것 이므로 학습시키지 않은 고차원 데이터의 차원 축소된 데이터를 얻을 수 있을 것이다.

II. LLE(Locally Linear Embedding)

LLE는 고차원에서 인접한 벡터들이 차원 축소된 저차원에서도 인접한 관계가 유지되도록 저차원 벡터를 계산해내는 방

법이다. 여기서는 이 과정을 간략히 설명한다[4].



▶▶ 그림 1. LLE(Locally Linear Embedding) 절차도.

먼저, D 차원의 실수 벡터 \vec{X}_i 가 N 개 주어졌다고 하자. (단계1) 이들 데이터들을 대상으로 국소(local)한 특성을 파악하는 과정으로 K-NN(Nearest Neighbor)를 수행한다. (단계2) 다음으로, 임의의 \vec{X}_i 가 인접한 K개의 벡터들의 합으로 표시되는 가중치 W_{ij} 들을]

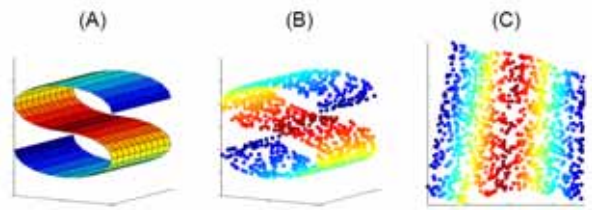
$$E(W) = \sum_i |\vec{X}_i - \sum_j W_{ij} \vec{X}_j|^2 \quad (1)$$

가 최소가 되도록 구해낸다. 여기서, 가중치 W_{ij} 는 \vec{X}_j 가 \vec{X}_i 의 구성에 기여하는 정도를 나타낸다. 그리고, W_{ij} 는 K-NN에 속하지 않는 벡터들의 관계에서는 0의 값을 가지도록 하여 sparseness의 성질이 지니도록 한다. 또한, $\sum_j W_{ij} = 1$ 이 되도록 하는 제한도 가해진다.

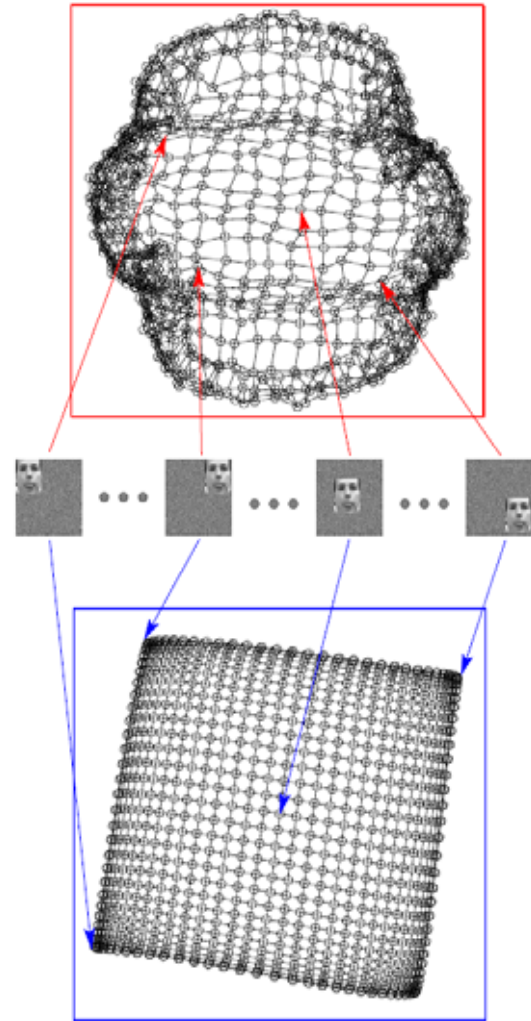
(단계 3) 마지막 단계로 단계2에서 구해진 W_{ij} 를 고정시킨 상태에서 \vec{X}_i 를 저차원으로 매핑시킨 d 차원의 \vec{Y}_i 를 구해낸다. 이 과정에서는

$$\Phi(Y) = \sum_i |\vec{Y}_i - \sum_j W_{ij} \vec{Y}_j|^2 \quad (2)$$

가 최소로 되도록 하는 \vec{Y}_i 를 구해내는 것이다. 이 과정에서 \vec{X}_i 는 관여하지 않는다. 그림 1은 이러한 과정을 도식적으로 그린 것이다.



▶▶ 그림 2. LLE(Locally Linear Embedding) 적용예



▶▶ 그림 3. 얼굴 영상에 적용한 결과

그림 2는 LLE의 효과를 보여주는 한 예이다. (A)는 3차원 공간 상에 존재하는 데이터들을 보여주며, 이를 샘플링한 데이터들이 (B)에 그려져 있다. 이 샘플링된 데이터들에 대하여 LLE를 수행하여 얻어진 2차원 데이터들이 (C)에 나타난 바와 같다. 원래 데이터들이 지닌 인접한 관계가 LLE 결과에서도 유지되고 있음을 알 수 있다.

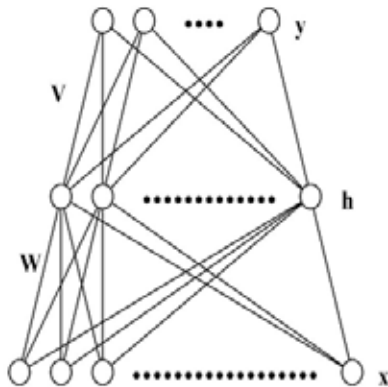
그림 3은 하나의 얼굴 영상을 잡음영상 위의 여러 위치에 있도록 한 $N=961$ 개의 영상에 대하여 PCA(Principal Component Analysis)와 LLE를 적용한 결과를 나타낸 것이

다. 여기서 $D=3009$ 이며, $K=4$ 를 적용하여 2차원의 차원 축소된 데이터($d=2$)를 얻은 것을 아래 그림에 그렸다. 잡음 영상에서 얼굴 부분이 차지하는 위치의 관계가 LLE의 결과에 잘 나타나 있는 것을 볼 수 있다. 한편 위의 그림은 PCA를 적용한 결과인데, 얼굴 영상의 인접한 특성을 제대로 반영하지 못한 것을 볼 수 있다.

III. LLE의 MLP에 의한 학습

LLE는 고차원의 데이터를 인접한 관계가 유지되도록 하면서 차원축소를 하는 성능이 뛰어나다. 그렇지만, 주어진 데이터 전체를 대상으로 분석을 수행하여 그 결과를 얻어내는 것이므로, 분석 대상에 포함되지 않은 새로운 데이터가 주어진 경우에는 이 데이터를 포함하여 새롭게 LLE를 수행하여야 하는 번거로움이 있다. 한편, MLP(Multilayer Perceptron)는 충분한 수의 중간층 노드가 주어지면 임의의 함수를 근사화할 수 있으므로, 여기서는 고차원 데이터의 LLE에 의해 얻어진 저차원 데이터들을 대상으로 MLP 학습을 수행하여 고차원에서 저차원으로 매핑 시키는 관계를 학습하는 방법을 제시한다.

이 경우, 학습시키고자 하는 입력 데이터는 고차원 데이터이며, 그 출력의 목표 데이터는 LLE에 의해 얻어진 저차원 데이터들이 된다.



▶▶ 그림 4. 다층퍼셉트론 구조

MLP의 학습 방법은 다음과 같다. MLP가 N 개의 입력 뉴런과 H 개의 중간층 뉴런 및 M 개의 출력 뉴런들로 구성되어 있다고 하자. 어떤 입력패턴 $\mathbf{x} = [x_1, x_2, \dots, x_N]$ 이 MLP에 입력되면, j 번째 중간층 뉴런의 값은

$$h_j = f(\hat{h}_j) = \tanh(\hat{h}_j/2), j = 1, 2, \dots, H \quad (3)$$

와 같이 주어진다. 여기서 $f(\cdot)$ 는 중간층 뉴런의 시그모이드 비선형 함수이며

$$\hat{h}_j = \sum_{i=0}^N w_{ji} x_i \quad (4)$$

는 중간층 뉴런에 입력되는 가중치 합이다. w_{ji} 는 x_i 와 h_j 를 연결하는 중간층 가중치이며 $x_0=1$ 로 주어지기에 w_0 는 바이어스라고 불린다. 같은 형태로 k 번째 출력 뉴런에 입력되는 가중치 합은

$$\hat{y}_k = \sum_{j=0}^H v_{kj} h_j, k = 1, 2, \dots, M \quad (5)$$

이고, v_{kj} 는 h_j 와 y_k 를 연결하는 출력층 가중치이고, $h_0=1$ 이며 v_{k0} 는 바이어스이다.

P 개의 학습패턴 $\mathbf{x}^{(p)} (p = 1, 2, \dots, P)$ 와 이들의 출력층 목표벡터 $\mathbf{t}^{(p)} = [t_1^{(p)}, t_2^{(p)}, \dots, t_M^{(p)}]$ 가 주어지면, 가중치들은 출력층에서

$$E^{out} = \frac{1}{2} \sum_{p=1}^P \sum_{k=1}^M (t_k^{(p)} - y_k^{(p)})^2 \quad (6)$$

로 주어지는 MSE(mean squared error)를 최소화시키도록 변경된다. 즉, EBP(Error Back-Propagation) 알고리즘에 의해 MLP의 출력층 각 노드에 연결된 출력층 가중치는

$$\Delta v_{kj} = -\eta \frac{\partial E^{out}}{\partial v_{kj}} = \eta \delta_k^{(p)} h_j^{(p)} \quad (7)$$

에 따라 학습된다. 여기서,

$$\delta_k^{(p)} = -\frac{\partial E^{out}}{\partial y_k^{(p)}} = (t_k^{(p)} - y_k^{(p)}) f'(y_k^{(p)}) \quad (8)$$

이다. 또한, 중간층 노드에 연결된 중간층 가중치는

$$\Delta w_{ji} = -\eta \frac{\partial E^{out}}{\partial w_{ji}} = \eta f'(\hat{h}_j^{(p)}) x_i^{(p)} \sum_{k=1}^M v_{kj} \delta_k^{(p)} \quad (9)$$

에 따라 학습된다.

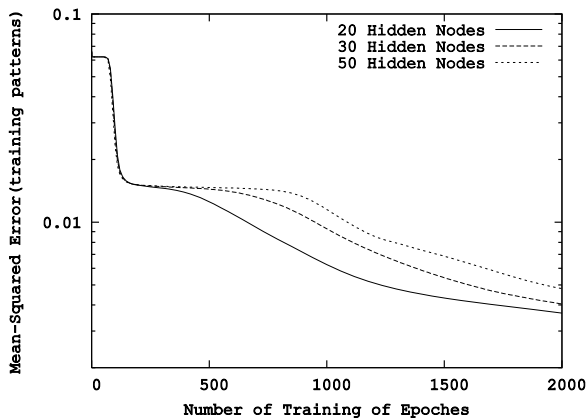
IV. 시뮬레이션

LLE의 매핑관계를 MLP로 학습시키기 위하여 대상 문제를 음성데이터로 정하였다. 원래, 음성과 입술모양을 함께 인식에 사용하기 위하여 만들어진 데이터에서 음성 부분만을 대상으로 LLE를 수행하여 MLP의 입력과 목표값에 해당하는 데이터를 얻어낸 것이다.

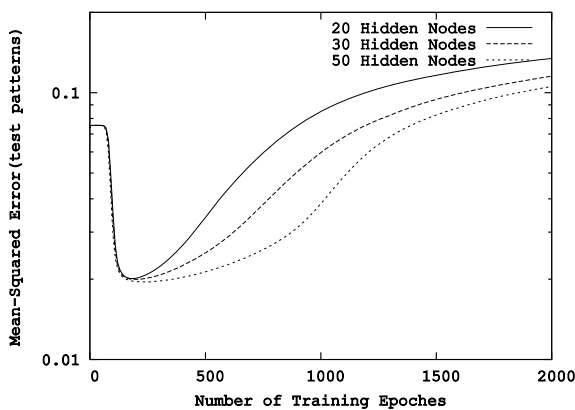
음성 데이터는 한명이 영어로 0부터 9까지를 5번씩 발음한

데이터를 16kHz로 샘플링한 것을 입술 영상과의 일치시키기 위하여 14.985kHz로 리샘플링한 것을 대상으로 삼았다. 이 음성 데이터에 대하여 프레임사이즈를 500 샘플로하고 250샘플이 겹치도록 이동을 시켜서 얻어지는 13차 MFCC(Mel-Frequency Cepstrum Coefficient) 중에서 C0를 제외한 나머지 12차 MFCC를 3개의 프레임 연속으로 연결시킨 36차 MFCC 데이터들을 745개 구하였다. 이를 대상으로 LLE를 수행하여 3차원으로 축소된 데이터들을 구해내어 MLP의 목표값으로 사용하였다.

이들 데이터에 대하여 임의로 선택하여 596개를 학습패턴으로 사용하였으며, 129개를 시험패턴으로 사용하였다. 학습에 사용된 MLP는 입력노드 36개 출력노드 3개로 하였으며, 중간층 노드는 20, 30, 50개로 변형시켜 MLP의 학습 결과를 확인하였다. 이때 학습률은 0.006을 사용하였다.



(a) 학습패턴에 대한 MSE 오차



(b) 시험패턴에 대한 MSE 오차

▶▶ 그림 5. LLE데이터를 대상으로 한 MLP 학습

그림 5는 MLP 학습과정에서 얻어지는 학습패턴에 대한 MSE(Mean-Squared Error)와 시험패턴에 대한 MSE를 그린 것이다. 그림 5(a)에서 보는 바와 같이 학습패턴에 대한 오

차는 학습이 지속될수록 꾸준히 감소하는 것을 볼 수 있지만, (b)에 나타난 시험패턴에 대한 MSE를 보면 중간층 노드 수의 변동에 상관 없이 0.02 수준에서 더 이상 감소하지 못하고 다시 증가하는 것을 볼 수 있다. 즉, LLE의 매핑관계를 충분히 반영하기에는 학습패턴 및 시험패턴의 개수가 부족하여, 학습패턴에 대한 MSE가 줄어들지만 시험패턴에 대한 MSE는 상당히 큰 수준에서 더 이상 좋은 성능을 얻을 수 없음을 알 수 있다. 이러한 관계에 대한 시뮬레이션을 하기 위해서는 학습패턴이 시험패턴의 확률 분포를 제대로 반영하는 더 많은 수의 데이터들을 얻을 수 있는 문제를 대상으로 삼아야 하겠다.

V. 결론

이 논문에서는 LLE의 매핑관계를 MLP에 학습시키는 방법을 제안하고 이를 시뮬레이션으로 구현하여 보았다. 그렇지만, 시뮬레이션으로 삼은 데이터들이 LLE매핑 관계를 충분히 반영할 만큼의 개수가 확보되지 않은 관계로 학습이 제대로 이루어지지 않았다.

감사의 글

시뮬레이션에 필요한 LLE 데이터를 제공하여 주신 오도관 씨에게 감사드립니다.

참고 문헌

- [1] T.-W. Lee, et al., "A unifying information-theoretic framework for independent component analysis," *Computers & Mathematics with Applications*, vol. 31, no. 11, pp. 1-21, March 2000.
- [2] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788-791, 1999.
- [3] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 22, pp. 2319-2323, Dec. 2000.
- [4] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 22, pp. 2323-2326, Dec. 2000.