퍼지 및 지능시스템학회 논문지 2004, Vol. 14, No. 7, pp. 847-851

변형된 혼합 밀도 네트워크를 이용한 비선형 근사

Nonlinear Approximations Using Modified Mixture Density Networks

조원희*, 박주영[1)]
Won-Hee Cho and Jooyoung Park

* Jatco Korea Engineering Co.
고려대학교 제어계측공학과

요 약

Bishop과 Nabney에 의해 소개된 기존의 혼합 밀도 네트워크(Mixture Density Network)에서는 조건부 확률밀도 함수의 매개변수들(parameters)이 하나의 MLP(multi-layer perceptron)의 출력 벡터로 주어진다. 최근에는 변형된 혼합 밀도 네트워크(Modified Mixture Density Network)라고 하는 이름으로 조건부 확률밀도 함수의 선분포(priors), 조건부 평균(conditional means), 그리고 공분산(covariances) 등이 각각 독립적인 MLP의 출력벡터로 주어지는 경우를 다룬 연구가 보고된 바 있다. 본 논문에서는 조건부 평균이 입력에 관해 선형인 경우를 위한 버전에 대한 이론과 매트랩 프로그램 개발을 다룬다. 본 논문에서는 우선 일반적인 혼합 밀도 네트워크에 대해 간단히 설명하고, 혼합 밀도 네트워크의 출력인 다층 퍼셉트론의 매개변수를 각각 다른 다층 퍼셉트론에서 학습시키는 변형된 혼합 밀도 네트워크를 설명한 후, 각각 다른 다층 퍼셉트론을 통해 매개변수를 얻는 것은 동일하나 평균값은 선형함수를 통해 얻는 혼합 밀도 네트워크 버전을 소개한다. 그리고, 모의실험을 통하여 이러한 혼합 밀도 네트워크의 적용가능성에 대해 알아본다.

Abstract

In the original mixture density network(MDN), which was introduced by Bishop and Nabney, the parameters of the conditional probability density function are represented by the output vector of a single multi-layer perceptron. Among the recent modifications of the MDNs, there is the so-called modified mixture density network, in which each of the priors, conditional means, and covariances is represented via an independent multi-layer perceptron. In this paper, we consider a further simplification of the modified MDN, in which the conditional means are linear with respect to the input variable together with the development of the MATLAB program for the simplification. In this paper, we first briefly review the original mixture density network, then we also review the modified mixture density network in which independent multi-layer perceptrons play an important role in the learning for the parameters of the conditional probability, and finally present a further modification so that the conditional means are linear in the input. The applicability of the presented method is shown via an illustrative simulation example.

Key words : 가우시안 혼합 모델(gaussian mixture model), 혼합 밀도 네트워크(mixture density network), 비선형 근사(nonlinear approximation), 다층 퍼셉트론(multi-layer perceptron)

1. 서 론

다층 퍼셉트론(MLP, multi-layer perceptron)과 같은 일반적인 전방향 신경망은 비선형 근사(nonlinear approximation)에 광범위하게 사용되고 있으며 이러한 근사는 일반적으로 목표값(target)의 분산이 상수이고, 목표 값의 조건부 확률밀도 함수(conditional probability density function)가 일정한 크기의 분산을 갖는 단일 가우시안 형태 (the single gaussian of constant variance)라는 점을 가정한다. 근래에는, Bishop 등에 의하여 일반적인 전방향 신경망과는 달리 입출력 데이터의 연관성을 보다 세밀하게 표현할 수 있는 새로운 형태의 신경망인 혼합 밀도 네트워크(MDN, mixture density network)가 제시된 바 있다[1-4]. 혼합 밀도 네트워크는 비상수(non-constant) 분산이나 가우시안 형태가 아닌 임의의 분포를 사용해 모델링할 수 있는 특징이 있다. 그리고 이 네트워크에서는 조건부 확률밀도 함수의 매개변수들(parameters)이 하나의 다층 퍼셉트론(multi-layer perceptron)의 출력 벡터 형태로 주어지는 형태로 구현된다[2].

그동안 혼합 밀도 네트워크는 여러 분야에 적용되어 왔으며[3-8], 필요에 따라 다양한 형태로 수정되기도 하였다. 이러한 수정본들 중 본 논문에서 고려하는 형태는 Schittenkopf 등[5]에 의해 제시된 변형된 혼합 밀도 네트워크(Modified Mixture Density Network)이다. 이 네트워크의 주요 개념은 원래의 혼합 밀도 네트워크와 같이 다층 퍼셉트론을 조건부 확률밀도 함수의 각 변수들을 예측하는데 사용하되, 조건부 확률밀도 함수의 선분포(priors), 조건부 평균, 그리고 분산값 등을 모두 제공하는 하나의 다층 퍼셉트론을 사용하는 대신 이들 각각을 위하여 독립적인 다층 퍼셉트론을 사용한다는 점이다.

접수일자 : 2004년 9월 30일
완료일자 : 2004년 11월 20일
1) 교신저자

다음에서는 Schittenkopf 등의 변형된 혼합 밀도 네트워크의 구조를 대상으로 하여 특별히 조건부 평균이 입력에 관해 선형인 경우를 위한 혼합 밀도 네트워크 형태의 방법론을 도출하고자 한다. 이러한 모델은 국소적으로 선형적인 특성을 띄되 분산 등이 복잡한 변화 양상을 보이는 경우에 대해 효과적인 비선형 근사를 수행할 수 있는 가능성을 가질 것으로 기대된다. 본 논문에서는 이러한 경우의 네트워크를 편의상 개선된 혼합 밀도 네트워크라고 명하고 이를 위한 MATLAB 프로그램을 개발하였다.

본 논문의 구성은 다음과 같다. 2장에서는 일반적인 혼합 밀도 네트워크에 대해 설명하고 3장에서는 혼합 밀도 네트워크의 출력인 다층 퍼셉트론의 매개변수를 각각 독립된 다층 퍼셉트론에서 학습시키는 변형된 혼합 밀도 네트워크와 각각 다른 다층 퍼셉트론을 통해 매개변수를 얻는 것은 동일하나 평균값은 선형함수를 통해 얻는 개선된 혼합 밀도 네트워크를 소개한다. 4장에서는 모의실험을 통해 3장에서 소개한 개선된 혼합 밀도 네트워크를 적용한 결과를 살펴본다. 그리고, 마지막으로 5장에서는 결론을 제시하고 향후 과제에 대해 논의한다.

2. 혼합 밀도 네트워트

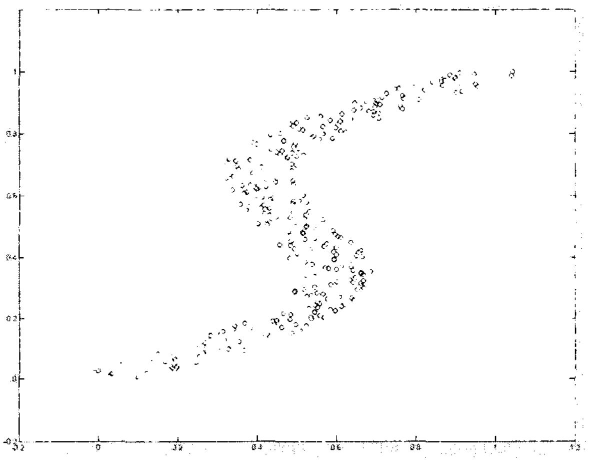


그림 2.1. 전형적인 전방향 신경망(feedforward neural network)을 이용하여 근사하기 힘든 종류의 학습 데이터 예 [2]

일반적으로 다층 퍼셉트론을 학습하는 목적은 주어진 학습 데이터를 충분히 정확하게 근사할 수 있는 다층 퍼셉트론 형태의 비선형 함수를 찾는 데 있다. 즉, 조건 밀도 함수 $p(t|x)$에 관한 정보를, 하나의 다층 퍼셉트론 기반 근사함수를 중심으로 표현하는 것이 학습의 기본 목적이라 할 수 있다. 그러나, 학습 데이터가, [2]에서 옮겨온 그림 2.1과 같은 유형의 분포를 갖는 경우에는, 이들을 하나의 다층 퍼셉트론만을 이용하여 나타내는 방식에는 상당한 한계가 있다. 즉, 학습 데이터의 분포에 따라서는 하나의 신경망에 의존하는 함수근사 전략은 불충분한 표현능력을 갖는 문제점을 갖게 되는 것이다. 그림 2.1의 경우에는, 입력값 x가 구간 [0.4 0.6] 부분에 속할 때에는 출력의 특성은 다중이다. 따라서, 우리가 이전에 사용했던 단일 가우시안 형태의 분포로는 조건 밀도 함수 $p(t|x)$를 효과적으로 표현할 수 없다. 이러한 종류의 문제는 여러 분야의 이공계 관련 문제에서 자주 관찰되며, 이러한 문제를 효과적으로 다루기 위해서는 일반화된 조건 확률밀도(general conditional probability density)를 이용한 모델링 구조가 필요하게 된다. 특히, 몇 개의 확률밀도함수가 혼합된 형태인 혼합 밀도 네트워크를 사용하되 혼합 밀도함수의 매개변수 Z를 입력 벡터 x의 함수로 설정하면, 이러한 종류의 문제를 효과적으로 다룰 수 있게 된다. 일반적으로, 매개변수 Z는 입력 x의 변화에 대해 복잡한 변화 특성을 보이므로, x와 $Z(x)$의 관계식을 모델링 하는 데에는 universal approximation property를 소유한 다층 퍼셉트론[9]이 효과적으로 사용될 수 있다. 혼합 밀도 네트워크에서 사용하는 조건부 확률 모델은 다음 식과 같다.

$$p(t|x) = \sum_{j=1}^{M} \alpha_j(x)\phi_j(t|x) \tag{2.1}$$

위의 식에서 보여진 바와 같이, 혼합 밀도 네트워크의 조건부 확률밀도를 위한 식은 기본적으로 커널 밀도함수 ϕ_j의 일차결합(linear combination) 형태로 그려진다. 그리고, 커널 밀도함수를 위해서는 일반적으로 다음의 가우시안 함수가 사용된다[2]:

$$\phi_j(t|x) = \frac{1}{(2\pi)^{c/2}\sigma_j(x)^c}\exp\left\{-\frac{\| t-\mu_j(x) \|^2}{2\sigma_j(x)^2}\right\} \tag{2.2}$$

식 (2.1)에서 M은 커널 밀도함수의 개수이고, $\sigma_j(x)$는 x에 따라 정규분포의 공분산이 어떻게 변화하는가를 보여주는 함수이다. 그리고, $\alpha_j(x)$는 각 x에 대해서 음이 아닌 수이며 총합이 1이 되는 혼합 계수(mixing coefficient)이며, 사전확률(prior probabilities)의 의미를 갖는다. 식 (2.2)에서 벡터 $\mu_j(x)$는 j 번째 커널의 중심을 의미한다. 이상에서 설명된 바와 같이, 혼합 밀도 네트워트의 기본 전략은 전체 학습 데이터의 입출력 상관관계를, 각 매개변수가 신경망으로 표현되는 조건부 가우시안 커널의 가중 평균 형태를 이용하여 표현하는 것이다. 이러한 전략을 그림으로 표현하면 그림 2.2와 같다[2]:

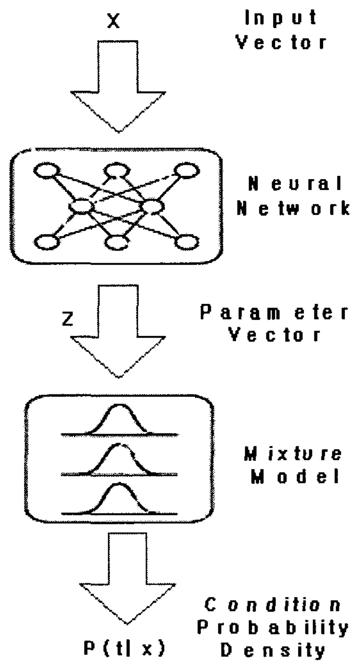


그림 2.2. 혼합 밀도 네트워크의 기본개념 [2]

(2.1)과 (2.2)식의 매개변수 $\alpha_j(x)$, $\mu_j(x)$, $\sigma_j(x)$를 위해 [2]에서 사용한 신경망은 그림 2.3과 같이 은닉층(hidden layer)이 1개인 다층 퍼셉트론이다. 학습 데이터의 독립변수 부분이 이 다층 퍼셉트론의 입력이 되고, 출력층(output layer)은 모든 매개변수가 모두 망라된 출력노드들로 이루어 진다. 본 논문에서는 이러한 신경망의 기본 구조를 보다 단순하게 변화시킴으로써, 입출력변수의 상관관계가 지역적으로 혹은 전역적으로 선형인 모드가 혼합된 종류의 학습 문제의 목적에 보다 부합되는 형태의 결과를 얻는데 중점을 두고 있다. 이에 관한 상세한 내용은 다음 장에서 설명된다.

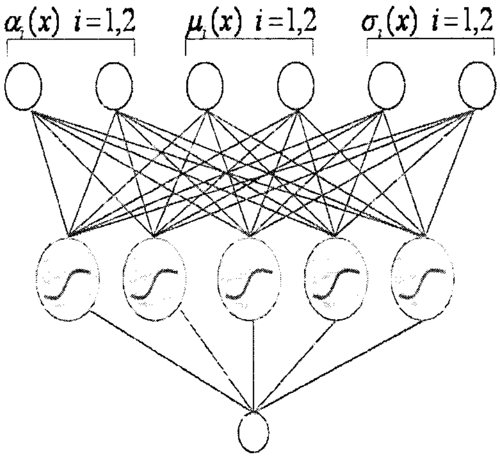


그림 2.3. 일반적인 혼합 밀도 네트워크에 사용되는 은닉층이 1개인 퍼셉트론 [2]

3. 변형된 혼합 밀도 네트워크

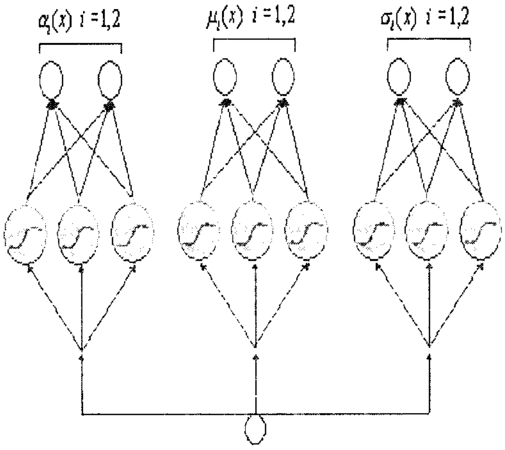


그림 3.1. 변형된 혼합 밀도 네트워크의 신경망 부분 [5]

[5]에서 제시된 변형된 혼합 밀도 네트워크(Modified Mixture Density Network)는, 그림 3.1에서와 같이 혼합 밀도 네트워크의 선분포(priors), 조건부 평균(conditional means), 그리고 공분산(covariances) 등이 각각 독립적인 다층 퍼셉트론의 출력벡터로 주어진다. 이러한 혼합 밀도 네트워크에서는 조건부 확률밀도 함수가 다음과 같이 모델링된다[5].

$$p(x_t|x_{t-1},\ldots,x_{t-m}) = \sum_{i=1}^{n} \alpha_{i,t}\, g(\mu_{i,t},\ \sigma^2_{i,t}) \qquad (3.1)$$

$$g(\mu_{i,t},\sigma^2_{i,t}) = \frac{1}{\sqrt{2\pi\sigma^2_{i,t}}}\exp\left\{-\frac{(x_t-\mu_{i,t})^2}{2\sigma^2_{i,t}}\right\} \qquad (3.2)$$

그리고, 식 (3.1), (3.2)에서 사용된 매개변수 $\alpha_{i,t}$, $\mu_{i,t}$과 $\sigma^2_{i,t}$ 등은 다음과 같이 MLP 신경망을 이용하는 형태로 구해진다:

$$\alpha_{i,t} = s(\tilde{\alpha}_{i,t}) = \frac{\exp(\tilde{\alpha}_{i,t})}{\sum_{j=1}^{n}\exp(\tilde{\alpha}_{j,t})} \qquad (3.3)$$

$$\tilde{\alpha}_{i,t} = MLP_{1,i}(x_{t-1},\ldots,x_{t-m}) \qquad (3.4)$$

$$\mu_{i,t} = MLP_{2,i}(x_{t-1},\ldots,x_{t-m}) \qquad (3.5)$$

$$\sigma^2_{i,t} = (MLP_{3,i}(x_{t-1},\ldots,x_{t-m}))^2 \qquad (3.6)$$

여기에서, 식 (3.3)의 소프트맥스(softmax) 함수 $s(\tilde{\alpha}_{i,t})$는 하중(weight) $\alpha_{i,t}$가 양이 되고 그 합이 1이 되도록 한다. 그리고, 식 (3.6)의 $\sigma^2_{i,t}$는 분산값이 항상 양수가 됨을 보장하게 된다 결과적으로 각각의 다층 퍼셉트론은 동일한 m 차원의 입력 $x_{t-1},\ldots,x_{t-m}$을 받고 서로 다른 n 차원의 출력을 한다. n 의 개수는 가우시안 구성요소의 개수와 동일하다. 각각의 MLP는 다음과 같은 입출력 관계식을 갖는다:

$$\begin{aligned} &MLP_i(x_{t-1},\ldots,x_{t-m}) \\ &= \sum_{j=1}^{h} v_{ij}\tanh\Big(\sum_{k=1}^{m} w_{jk}\,x_{t-k} + c_j\Big) + b_i \end{aligned} \qquad (3.7)$$

여기에서 i는 출력 뉴런을 위한 색인이며, h는 은닉 뉴런의 개수, w_{jk}와 v_{ij}는 첫 번째와 두 번째 층의 하중(weight)이며, c_j와 b_i는 첫 번째와 두 번째 층의 바이어스이다. 혼합 밀도 네트워크의 각 신경망을 위한 매개변수 값들은 식 (3.8)과 같이 우도(likelihood) 함수에 음의 로그를 취한 성능지수를 scaled conjugate gradient 기법이나 conjugate gradient 기법 등의 비선형 최적화 풀이법을 사용하여 최소화하여 구하게 된다.

$$L = -\frac{1}{N}\log \Pi_{t=m+1}^{m+N}\, p(x_t|x_{t-1},\ldots,x_{t-m}) \qquad (3.8)$$

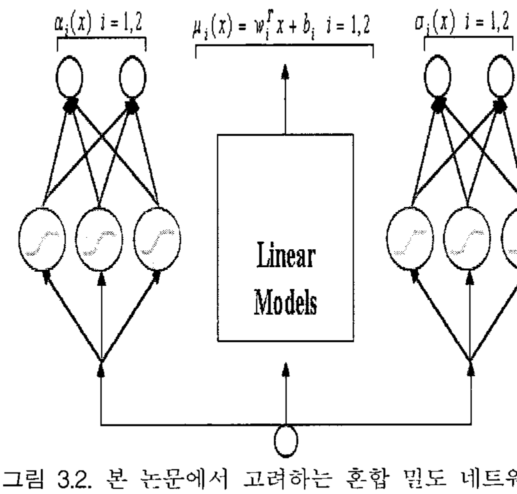


그림 3.2. 본 논문에서 고려하는 혼합 밀도 네트워크의 신경망 부분

본 논문에서 고려하는 개선된 혼합 밀도 네트워크(improved mixture density network)는 선분포(priors), 조

건부 평균(conditional means), 그리고 공분산(covariances) 등이 각각 독립적인 다층 퍼셉트론의 출력벡터로 주어진다는 점에서는 변형된 혼합 밀도 네트워크와 유사한 성질을 가지고 있다. [5]의 변형된 혼합 밀도 네트워크와 다른 점은 조건부 평균이 입력에 관하여 선형인 경우를 고려한다는 점이다. 이것은 Takagi-Sugeno 퍼지모델처럼 국소적 선형 모델의 혼합으로 전체 매핑을 표현하는 종류의 문제 해결에 유용할 것으로 예상된다. 고려된 혼합 밀도 네트워크에서는 평균값이 다음과 같이 선형 표현을 취하게 됨에 유의하자:

$$\mu_{i,t} = w_i^T x_t + b \tag{3.9}$$

따라서, 본 논문에서 고려하는 혼합 밀도 네트워크의 매개변수를 구하기 위한 신경망은 그림 3.2와 같이 고전적인 다층 퍼셉트론과 선형 모델이 혼합된 형태를 띄되, 각 종류의 매개변수, 즉, 혼합계수, 조건부 평균, 및 분산에 관한 정보는 각각 독립된 부분의 네트워트에 의해 구해지게 된다. 본 논문에서는 [2]의 매트랩 프로그램을 대폭 수정하여 이상에서 설명한 개선된 혼합 밀도 네트워크 기능을 수행하는 매트랩 프로그램을 확립하였다. 이 프로그램의 적용에 관한 내용은 다음 절에서 소개된다.

4. 모의 실험

모의실험에서는 다음의 두 식에 의해 발생된 데이터가 섞여있는 상황을 고려한다.

$$x = 0.015t + 0.27 + error_1 \tag{4.1}$$

$$x = -0.02t + 0.285 + error_2 \tag{4.2}$$

이 학습 데이터는 그림의 좌측 부분에서는 첫 번째 선형 모드가 우세하고 우측 부분에서는 두 번째 선형 모드가 우세하되, 두 모드가 비교적 널리 퍼져 있는 특성을 갖는다. 따라서, 비교적 전역적으로 섞여 있는 두 선형 모드의 개략적 모양을 추출할 수 있는지가 본 예제의 주요 관점이 된다.

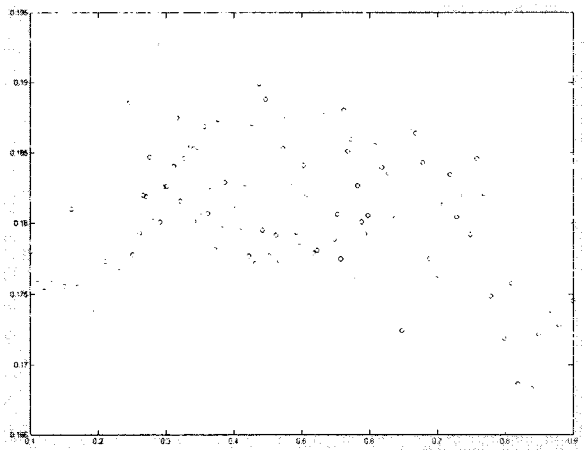


그림 4.1. 모의실험에서 사용된 데이터

그림 4.1의 데이터를 본 논문에서 고려한 개선된 혼합 밀도 네트워크를 이용하여 발견한 각 모드에 대한 정보가 그림 4.2에서 보여졌다. 이 그림에서 실선으로 그려진 부분은 각 모드의 선형 특성이고, 점선으로 그려진 부분은 구해진 분산 관련 정보를 나타낸다. 구해진 결과를 분석해보건대, 각 선형

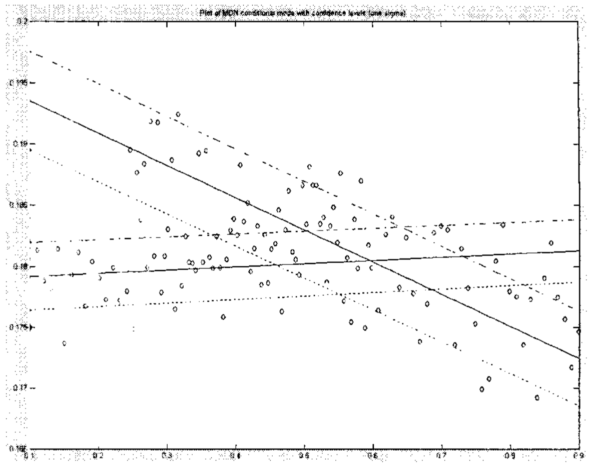


그림 4.2. 본 논문의 혼합 밀도 네트워크를 구현한 매트랩 프로그램이 생성한 결과

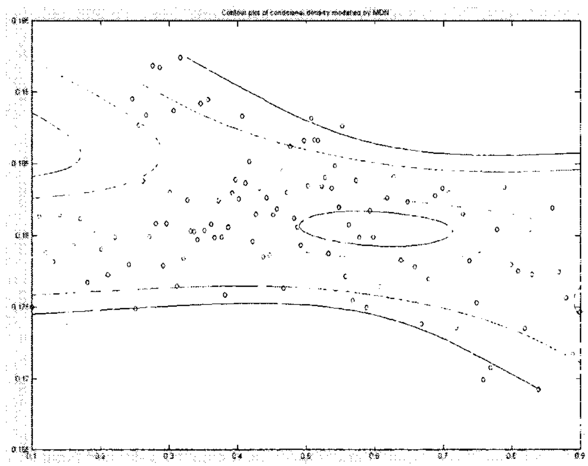


그림 4.3. 본 논문에서 고려한 개선된 혼합 밀도 네트워크에 의해 구해진 조건 밀도

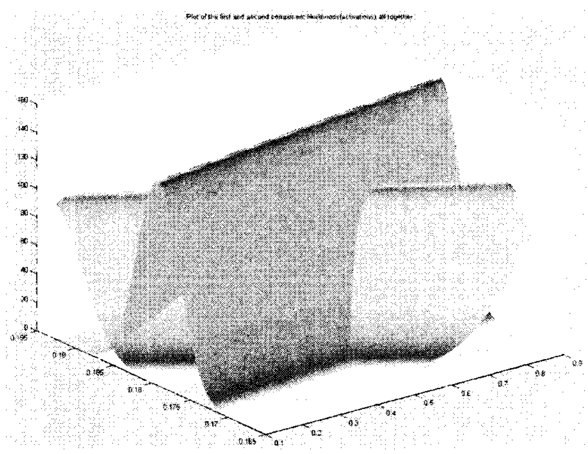


그림 4.4. 각 component likelihood의 도시

모드의 기울기가 학습 데이터의 생성에 사용된 경우와 정확하게 일치하고 있지는 못하지만, 대략적인 두 모드의 차이점 등은 잘 부각하고 있음을 관찰할 수 있다. 따라서, 본 논문의 방법론은 모의 실험의 데이터의 내재적 특성을 어느 정도 잘 파악할 수 있는 결과를 제시한다고 할 수 있다고 결론지을 수 있다. 참고로, 전체적인 조건밀도함수와 각 component likelihood가 각각 그림 4.3과 4.4에서 보여졌다.

그리고, 성능 비교를 위하여 [5]의 변형 혼합 밀도 네트워크를 적용한 결과도 그림 4.5에서 보여졌다. 그림 4.2와 4.5의 비교로부터 본 논문에서 고려한 방법론이 기존의 방법론 보다 의미 있는 결과를 생성하였음을 관찰할 수 있다.

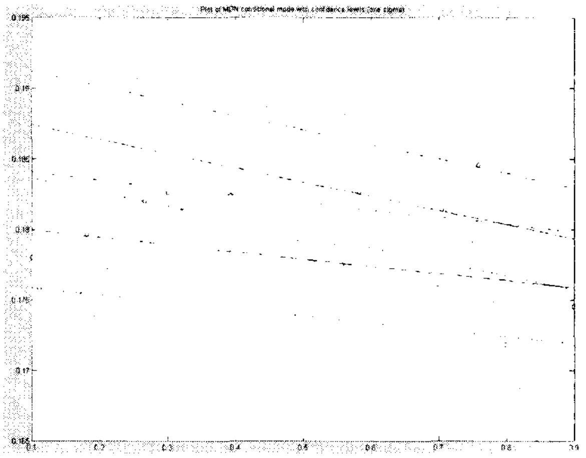


그림 4.5. 변형 혼합 밀도 네트워크를 구현한 매트랩 프로그램이 생성한 결과

5. 결론 및 향후 과제

Bishop의 혼합 밀도 네트워크는 모든 파라메터의 추정을 하나의 다층 퍼셉트론만을 사용하여 학습을 하는 방법을 취하였다. 본 논문에서 고려한 혼합 밀도 네트워크는 [5]에서와 같이 매개변수 각각에 대해 분리된 다층 퍼셉트론을 사용하는 방법을 채용하고 출력이 평균인 곳의 다층 퍼셉트론 부분을 선형모델로 대치하는 방안을 고려하였다. 이것은 주어진 데이터가 선형 모드의 혼합일 경우에 데이터의 원형복원에 더 우세하고 본래 함수의 함수식을 찾는 것이 가능하다는 장점이 있기 때문이다. 또한, 이러한 성질은 Takagi-Sugeno 퍼지모델의 identification과 같은 종류의 문제 해결에 도 활용할 여지가 있을 것으로 예상된다. 향후과제로는 현재의 학습 데이터보다 더 많은 개수와 더 복잡한 양상의 띠는 데이터에 대한 분석 등을 들 수 있다.

참 고 문 헌

[1] C. M. Bishop, Mixture Density Network, Neural Computing Research Group Report, February 1994.

[2] I. T. Nabney, NETLAB: Algorithms for Pattern Recognition, Springer, 2001.

[3] C. M. Bishop, Neural Network for Pattern Recognition, Oxford University Press, 1999.

[4] P. Moerland, Mixture Models for Unsupervised and Supervised Learning, Ph.D. thesis, Computer Science Department of the Swiss Federal Institute of Technology, Lausanne, 2000.

[5] C. Schittenkopf, G. Dorffner, and E. J. Dockner, "Identifying Stochastic Processes with Mixture Density Networks," Adaptive Information Systems and Modelling in Economics and Management Science Working Paper No.11, 1998.

[6] L. U. Hjorth and I. T. Nabney, "Regularisation of mixture density networks," Proceedings of Ninth International Conference on Artificial Neural Networks(ICANN'99), vol. 2, pp. 521-526, 1999.

[7] L. U. Hjorth and I. T. Nabney, "Bayesian training of mixture density networks," Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, pp. 455-461, 2000.

[8] K. Richmond, "Mixture density networks, human articulatory data and acoustic-to-articulatory inversion of continuous speech," Proceedings of Workshop on Innovation in Speech Processing, pp. 259-276, 2001.

[9] S. Haykin, Neural Networks: A Comrehensive Foundation, Second Edition, Prentice Hall, 1998.

저 자 소 개

조원희(Won-Hee Cho)

2002년 : 고려대학교 제어계측공학과 졸업(학사)

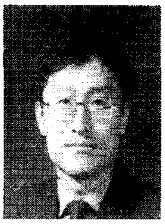
2004년 : 고려대학교 대학원 제어계측공학과 졸업(석사)

2004년~현재 : Jatco Korea Engineering Co. 연구원

관심분야 : SVM 응용, 강화학습

Phone : 011-9098-8970

E-mail : loadneo@hanmail.net



박주영(Jooyoung Park)

1983년 : 서울대학교 전기공학과 졸업(학사)

1985년 : 한국과학기술원 졸업(석사)

1985년3월~1988년7월 : 한국전력 월성원자력발전소 근무

1992년 : University of Texas at Austin, 전기 및 컴퓨터공학과 졸업(박사)

1992년8월~1993년2월 : 한국전력 전력경제연구실 선임전문원

1993년3월~현재 : 고려대학교 과학기술대학 제어계측공학과 교수

관심분야 : 신경망이론, 지능시스템, 비선형시스템

E-mail : parkj@korea.ac.kr