**여러가지 조건 변화에 따른 performance 차이 비교(MLP)**

2015년 8월 5일

김의성, 선한결, 나기현

1. Error (MSE vs SSE)
2. Hidden layer 개수
3. Mini batch mode VS pattern mode
4. Hidden node 개수
5. Learning Rate
6. Error Max
7. **“Error 처리방식(MSE vs SSE)” 측면에서 본 분석결과**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Training Set | **에러처리방식** | **Error Max** | **Learning Rate** | **Hidden Node Number** | **mini Batch** | **Hidden Layer Number** | **Result1**  **(속도)**  **(평균)** | **Result2**  **( 정확도)** |
| 100 | SSE | 0.01 | 0.1 | 20 | 10 | 1 | Epoch: 3050  Time : 54 초 | 1 : 0.98이상  0 : 0.01이하 |
| 100 | MSE | 0.01 | 0.1 | 20 | 10 | 1 | Epoch: 100  Time : 1.8 초 | 1 : 0.62~0.98 (대부분 0.85이상)  0 : 0.05 이하 |
| 100 | MSE | 0.001 | 0.1 | 20 | 10 | 1 | Epoch: 350  Time : 7 초 | 1 : 0.96 이상  0 : 0.02 이하 |
| 100 | MSE | 0.0001 | 0.1 | 20 | 10 | 1 | Epoch: 2900  Time : 51초 | 1 : 0.98 이상  0 : 0.005이하 |
| 10000 | **MSE** | 0.01 | 0.1 | 20 | 1 | 1 | Epoch: 100  Time :180 초 | 1: 0.99 이상  0: 0.001 이하 |

* 같은 Error Max 값으로 프로그램을 돌렸을 때, SSE 방식으로 했을 때는 Epoch의 수는 더 많지만(느리다) 정확도가 더 높다. 반대로 MSE 방식으로 했을 때는 Epoch의 수가 적지만(비교적 많이 빠르다) 정확도가 상대적으로 낮다.
* MSE방식은 기본적으로 빠르지만 데이터의 개수가 적으면 몇개의 정확도가 낮다. 하지만 Error Max를 변화시키지 않더라도 데이터의 개수가 많아지면 많아질 수록 정확도가 높아진다.
* SSE 방식은 기본적으로 느리지만 데이터의 개수의 상관 없이 정확도가 높다. MSE에 비교해서 상당히 느리다.

**2) Hidden Layer 개수 측면에서 본 분석결과**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **에러처리방식** | **Error Max** | **Learning Rate** | **Hidden Node Number** | **mini Batch** | **Hidden Layer Number** | | **Result1**  **(속도)**  **(평균)** | **Result2**  **(정확도)** |
| MSE | 0.01 | 0.1 | 20 | 10 | | 1 | Epoch: 100  Time : 1.8 초 | 1 : 0.62~0.98  (대부분 0.85이상)  0 : 0.05 이하 |
| MSE | 0.01 | 0.1 | 20 | 10 | | 2 | Epoch: 340  Time : 6.5 초 | 1 : 0.60~0.98  (대부분 0.85이상)  0 : 0.05 이하 |
| MSE | 0.01 | 0.1 | 20 | 10 | | 3 | Epoch: 1995  Time : 38 초 | 1 : 0.84~0.97  (대부분 0.90이상)  0 : 0.005이하 |

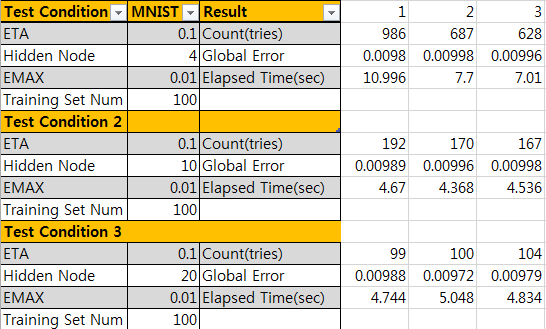
* Hidden Layer 개수가 증가할 수록 계산을 마치는데 시간이 오래 걸린다. 하지만 Hidden Layer를 3개로 했을 경우 정확도가 확연하게 높아지는 것을 확인할 수 있었다. 같은 조건에서 Hidden Layer가 4이상에서는 Error가 수렴하지 않았다.

**3) Mini batch mode 10개 VS pattern mode 측면에서 본 분석결과**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **training set 개수** | **에러처리방식** | **Error Max** | **Learning Rate** | **Hidden Node Number** | **mini Batch** | **Hidden Layer 개수** | **Result1**  **(속도)**  **(평균)** | **Result2**  **(정확도)** |
| 100 | **MSE** | 0.01 | 0.1 | 20 | 1 | 1 | Epoch: 100  Time : 3 초 | 1 : 0.54~0.96  0 : 0.05 이하 |
| 100 | **MSE** | 0.01 | 0.1 | 20 | 10 | 1 | Epoch: 97  Time : 1.8 초 | 1 : 0.62~0.96  0 : 0.05 이하 |
| 1000 | **MSE** | 0.01 | 0.1 | 20 | 1 | 1 | Epoch: 47  Time : 13 초 | 1 : 0.93~0.99  0 : 0.01 이하 |
| 1000 | **MSE** | 0.01 | 0.1 | 20 | 10 | 1 | Epoch: 43  Time : 8 초 | 1 : 0.90~0.99  0 : 0.01 이하 |
| 10000 | **MSE** | 0.01 | 0.1 | 20 | 1 | 1 | Epoch: 100  Time : 180 초 | 1: 0.99 이상  0: 0.001 이하 |
| 10000 | **MSE** | 0.01 | 0.1 | 20 | 10 | 1 | Epoch: 110  Time: 310 초 | 1: 0.99이상  0: 0.001 이하 |

- full batch mode 일 때는 수렴이 안되었다. 트레이닝 개수를 각각 100개, 1000개, 10000개 넣어봤을 때 실행해서 비교를 해보았는데, pattern mode(mini batch = 1) 일 때보다 mini-batch(=10) 일 때가 결과값에 수렴하는 속도가 더 빨랐다.

**4) Hidden Node 개수 변화에 따른 변화**



동일 Error Max(0.01) 하에서 node개수의 변화에 따라 수행 시간의 차이가 나는지를 비교했다. 대체적으로 node 개수가 많아지면 수행 횟수는 대체적으로 줄었으나, 수행 시간은 일정 node 개수 이상이 되면 큰 차이가 나지 않음을 발견할 수 있었다.

**5) Learning Rate 측면에서 본 분석결과**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **training set 개수** | **에러처리방식** | **Error Max** | **Learning Rate** | **Hidden Node Number** | **mini Batch** | **Hidden Layer 개수** | **Result1**  **(속도)**  **(평균)** | **Result2**  **(정확도)** |
| 100 | **MSE** | 0.01 | 0.05 | 10 | 10 | 1 | Epoch: 340  Time : 9.8 초 | 1 : 0.91  0 : 0.1 이하 |
| 100 | **MSE** | 0.01 | 0.1 | 10 | 10 | 1 | Epoch: 164  Time : 7.71 초 | 1 : 0.91  0 : 0.1 이하 |
| 100 | **MSE** | 0.01 | 0.5 | 10 | 10 | 1 | Epoch: 44  Time : 4.43 초 | 1 : 0.92  0 : 0.1 이하 |
| 100 | **MSE** | 0.01 | 1 | 10 | 10 | 1 | Epoch: 43  Time : 9.93 초 | 1 : 0.94  0 : 0.1 이하 |
| 100 | **MSE** | 0.01 | 2 | 10 | 10 | 1 | 진동(수렴 안함) | 진동(수렴 안함) |

다른 모든 조건을 동일하게 설정하고, Learning rate 값을 변화시켰을 때, 일정 선까지는 값이 증가할수록 수렴하는 시간이 감소함을 알 수 있다. 그러나 그 이상으로 계속 증가하게 되면 오히려 시간이 증가하게 되고, 결국에는 진동하거나 발산하는 결과를 보여준다. 그에 비해서, 정확도는 Learning rate 값의 변화의 영향을 받지 않는 결과로 봐서, 정확도와는 아무런 상관관계가 없음을 알 수 있다.

**5) Error Max**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **training set 개수** | **에러처리방식** | **Error Max** | **Learning Rate** | **Hidden Node Number** | **mini Batch** | **Hidden Layer 개수** | **Result1**  **(속도)**  **(평균)** | **Result2**  **(정확도)** |
| 100 | **MSE** | 0.001 | 0.1 | 10 | 10 | 1 | Epoch: 813  Time : 27.7 초 | 1 : 0.97  0 : 0.03 이하 |
| 100 | **MSE** | 0.01 | 0.1 | 10 | 10 | 1 | Epoch: 196  Time : 8.96 초 | 1 : 0.92  0 : 0.05 이하 |
| 100 | **MSE** | 0.05 | 0.1 | 10 | 10 | 1 | Epoch: 106  Time : 6.1 초 | 1 : 0.82  0 : 0.3 이하 |
| 100 | **MSE** | 0.1 | 0.1 | 10 | 10 | 1 | Epoch: 62  Time : 5 초 | 1 : 0.71  0 : 0.3 이하 |
| 100 | **MSE** | 0.5 | 0.1 | 10 | 10 | 1 | Epoch: 0  Time : 3.1 초 | 1 : 0.11  0 : 0.3 이하 |

다른 조건들이 동일한 상황에서, Error Max 값을 변화시키며 변화를 관찰하였다. Error Max 값이 증가할수록 학습 속도는 증가하지만, 정확도는 급격하게 감소함을 볼 수 있다. 일정 이상 감소하면 더 이상 “학습”이라는 역할을 수행한다고 보기 힘들 정도의 낮은 정확성을 보여준다. 이런 결과로 볼 때, Error Max 값은 학습 속도와는 비례하지만, 정확도 값과는 반비례 함을 알 수 있다.