

밑바닥부터 시작하는 딥러닝

모델 구현 결과

한국외대 이예진

목차

1. 2층 신경망

2. 학습 기술 적용

3. 합성곱 신경망 CNN

4. CBOW 모델

5. 개선된 CBOW

6. 순환 신경망 RNN

7. LSTM

8. Seq2seq

9. 어텐션 갖춘 Seq2seq

1. 2층 신경망 (1권 - ch4, 5)

전제 : 신경망에는 적응 가능한 가중치와 편향이 있고, 이 가중치와 편향을 훈련 데이터에 적응하도록 조정하는 과정이 학습.

신경망 학습은 다음과 같이 4단계로 수행.

1단계 - 미니배치 ↗ 교차 엔트로피 오차, 미니배치 학습

훈련 데이터 중 일부를 무작위로 가져옴. 이렇게 선별된 데이터를 미니배치라 하며, 그 미니배치의 손실 함수 값을 줄이는 것이 목표.

2단계 - 기울기 산출 ↗ 오차역전파법

미니배치의 손실 함수 값을 줄이기 위해 각 가중치 매개변수의 기울기를 구함. 기울기는 손실 함수의 값을 가장 작게 하는 방향 제시.

3단계 - 매개변수 갱신 ↗ 경사하강법

가중치 매개변수를 기울기 방향으로 아주 조금 갱신

4단계 - 반복

1~3단계를 반복.

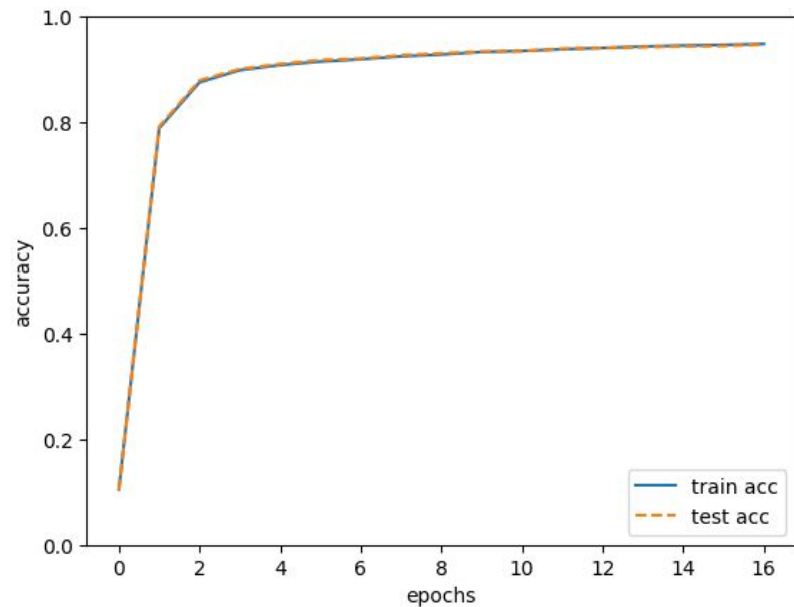
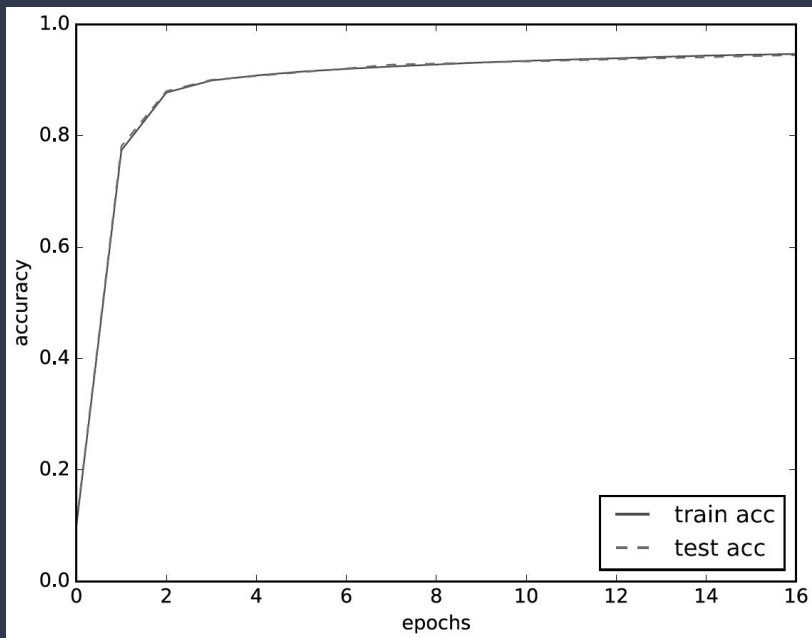
표 5-2 TwoLayerNet 클래스의 메서드

메서드	설명
<code>__init__(self, input_size, hidden_size, output_size, weight_init_std)</code>	초기화를 수행한다. 인수는 앞에서부터 입력층 뉴런 수, 은닉층 뉴런 수, 출력층 뉴런 수, 가중치 초기화 시 정규분포의 스케일
<code>predict(self, x)</code>	예측(추론)을 수행한다. 인수 x는 이미지 데이터
<code>loss(self, x, t)</code>	손실 함수의 값을 구한다. 인수 x는 이미지 데이터, t는 정답 레이블
<code>accuracy(self, x, t)</code>	정확도를 구한다.
<code>numerical_gradient(self, x, t)</code>	가중치 매개변수의 기울기를 수치 미분 방식으로 구한다(앞 장과 같음).
<code>gradient(self, x, t)</code>	가중치 매개변수의 기울기를 오차역전파법으로 구한다.

● MNIST 데이터셋



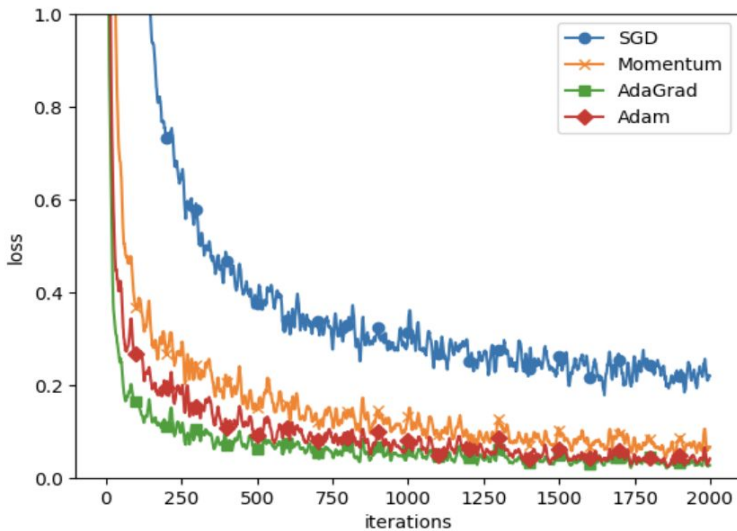
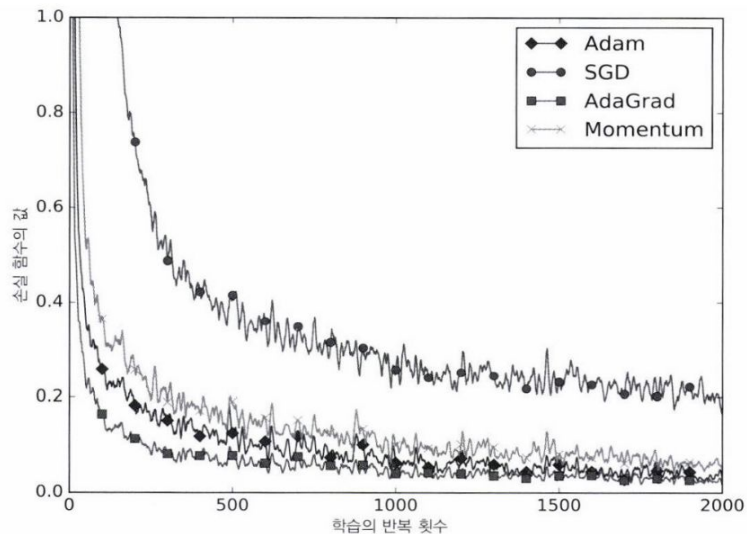
< 교재 결과 >



-> 유사한 결과, 에폭이 진행될수록 정확도 모두 좋아짐,
훈련 데이터와 시험 데이터의 정확도에 차이가 없음 (오버피팅x)

2. 학습 기술 적용 (1권 - ch6)

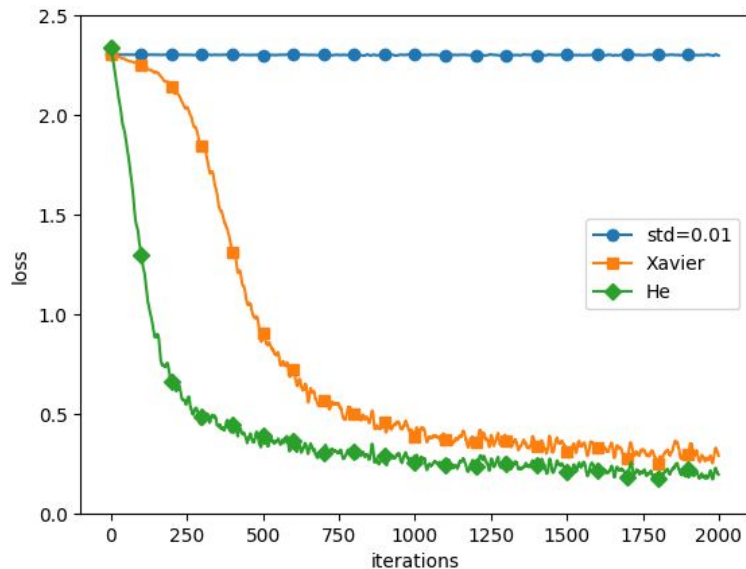
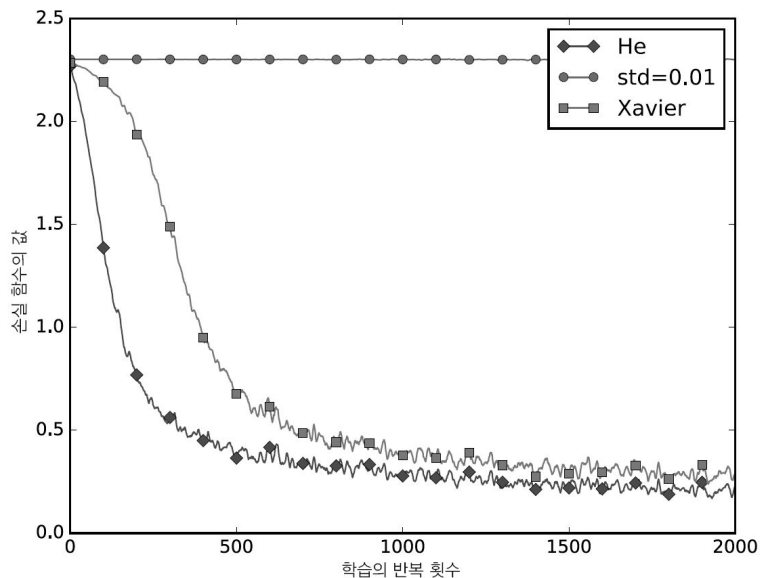
(1) 매개변수 갱신 방법 비교 - MNIST 데이터셋



-> 유사한 결과, SGD의 학습 진도가 가장 느림

2. 학습 기술 적용 (1권 - ch6)

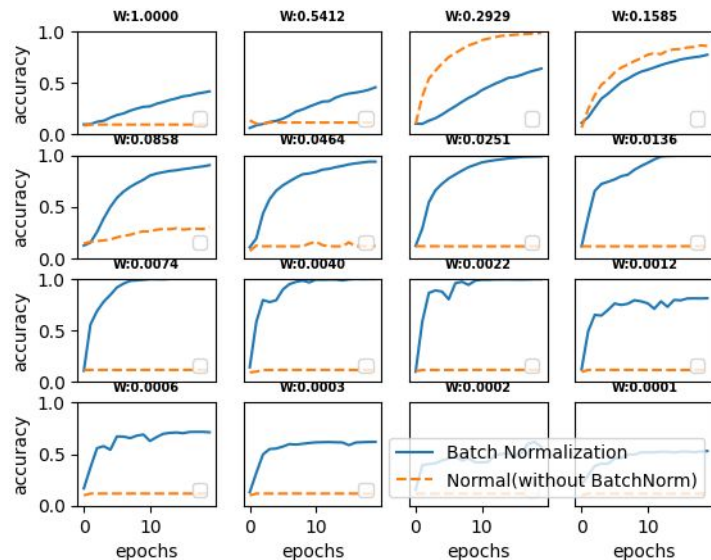
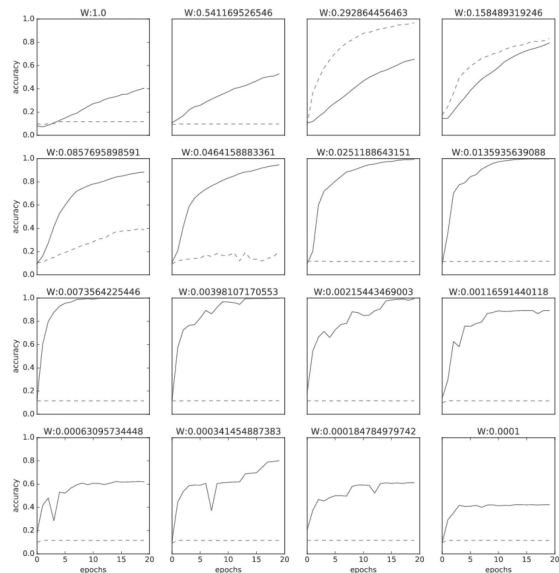
(2) 가중치의 초깃값 비교 - MNIST 데이터셋



-> 유사한 결과, 활성화 함수 ReLU -> He 초깃값일 때 학습 가장 빠름

2. 학습 기술 적용 (1권 - ch6)

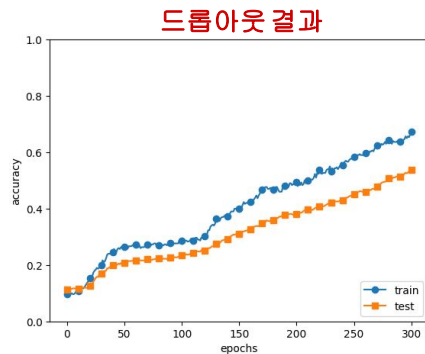
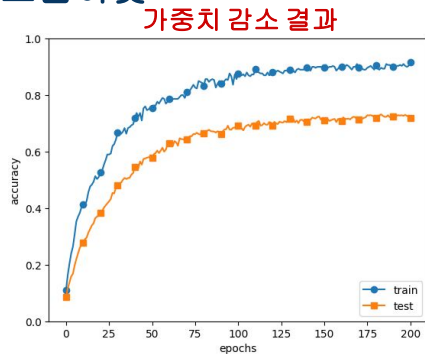
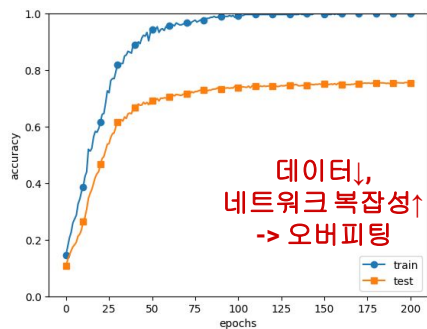
(3) 배치 정규화 효과 - MNIST 데이터셋



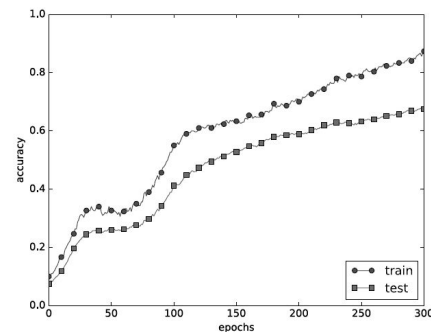
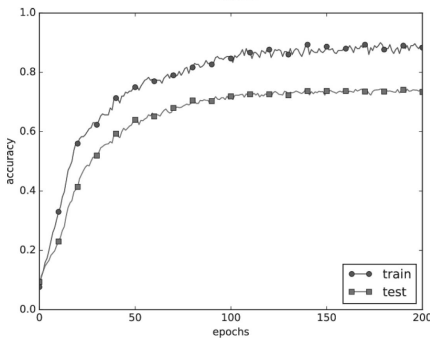
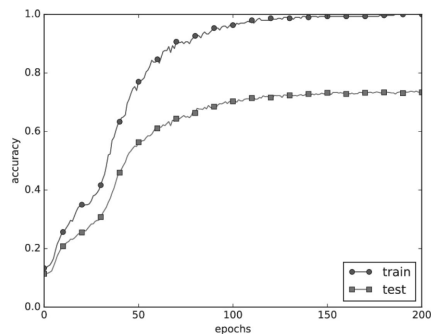
-> 유사한 결과, 가중치 초깃값에 크게 의존하지 않고 학습 빨라짐

2. 학습 기술 적용 (1권 - ch6)

(4) 오버피팅 -> 가중치 감소 -> 드롭아웃



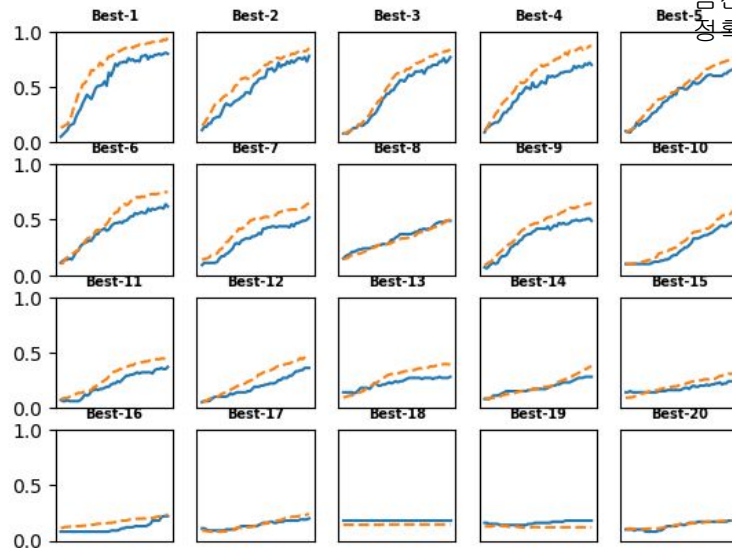
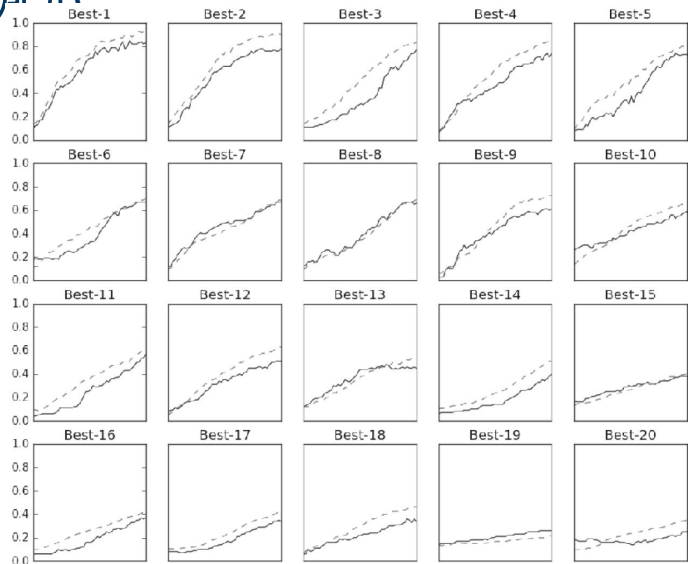
-> 훈련 데이터와
시험 데이터에 대한
정확도 차이 감소,
훈련 데이터에
대한 정확도↓



2. 학습 기술 적용 (1권 - ch6)

(5) 하이퍼파라미터 최적화 (학습률, 가중치 감소)

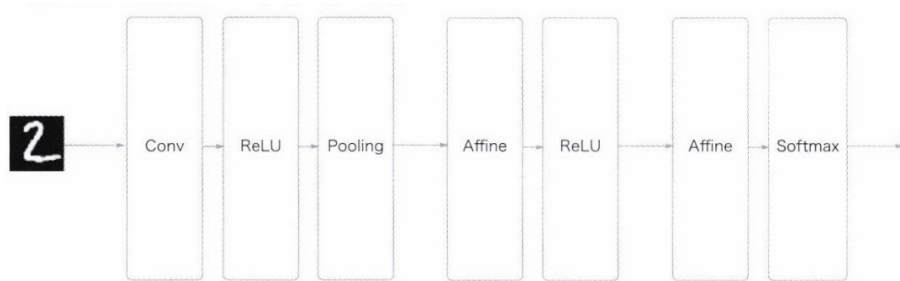
이러



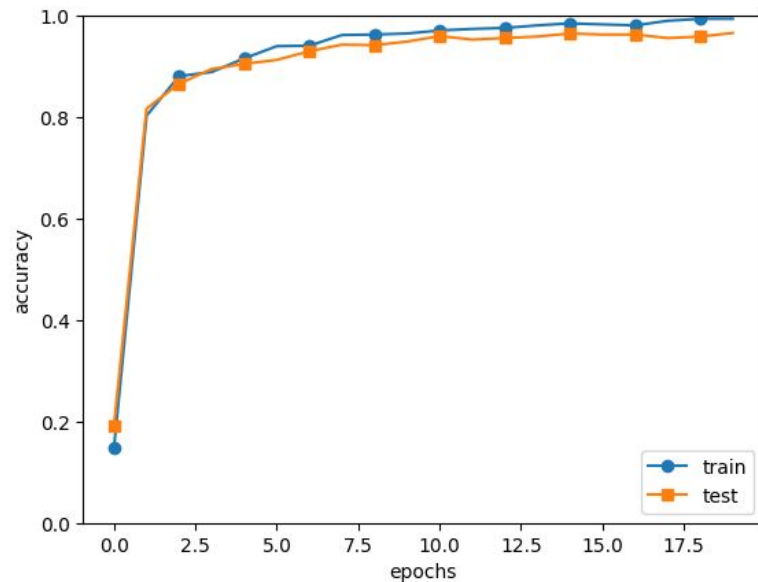
실선 - 검증 데이터
정확도
점선 - 훈련 데이터
정확도

-> 검증 데이터의 학습 추이를
정확도가 높은 순서로 나열

3. CNN (1권 - ch7)

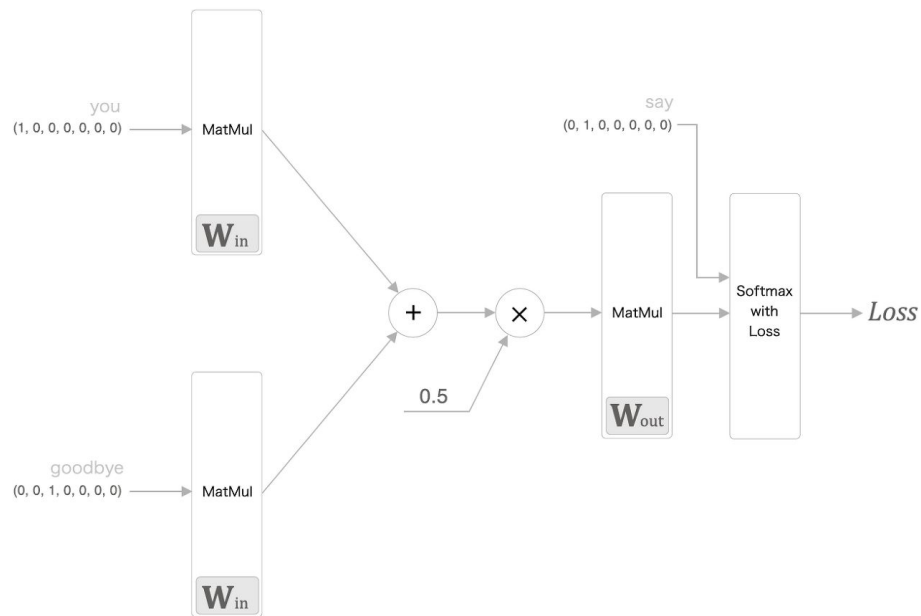


- train-5000, test-1000개로 데이터 줄임
- 매개변수 갱신 - Adam
- 학습률 : 0.001



-> 내 결과 훈련 데이터에 대한 정확도 : 99.40%, 시험 데이터에 대한 정확도 : 96.60%
-> 교재 결과 훈련 데이터에 대한 정확도 : 99.82%, 시험 데이터에 대한 정확도 : 98.96%

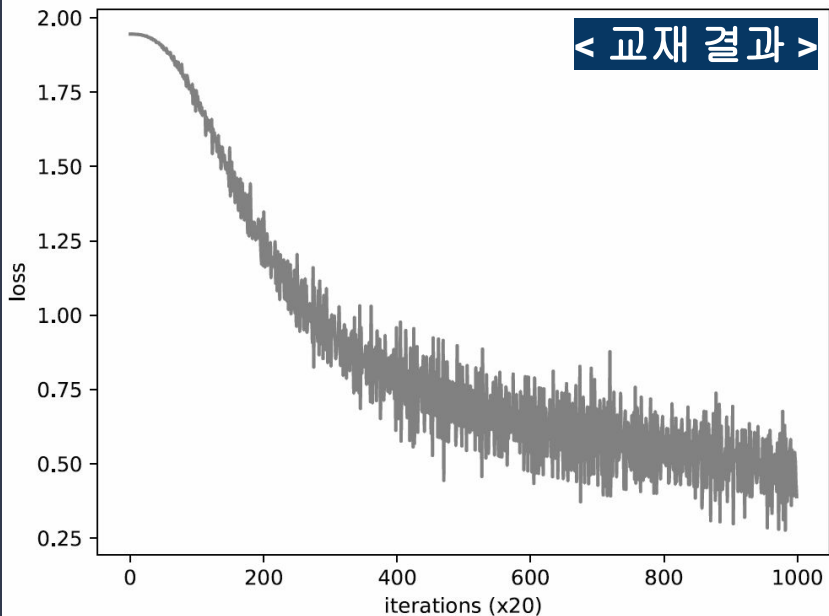
4. CBOW (2권 - ch3)



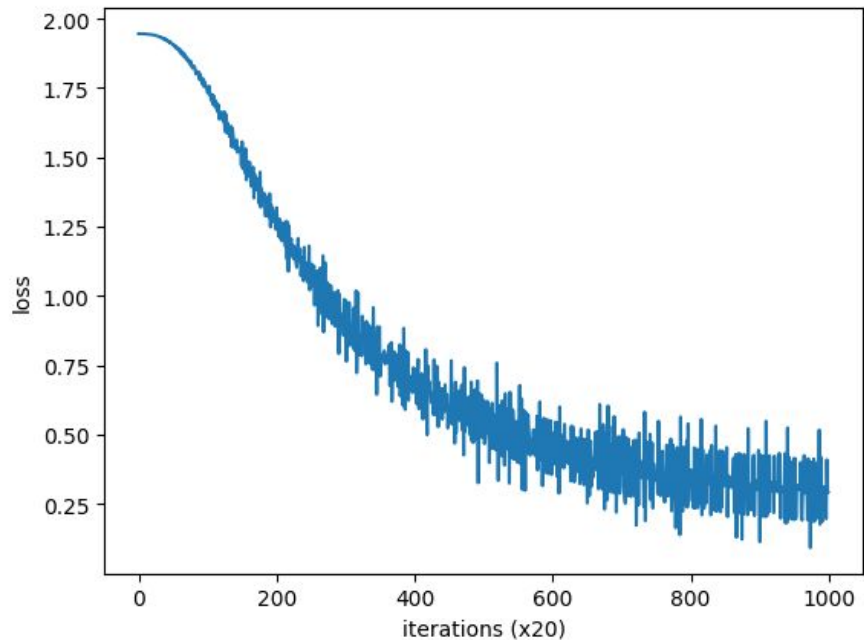
- **Trainer** 클래스 사용
- 매개변수 갱신 - Adam
- **window_size = 1**
- **hidden_size = 5**

Data : 'You say goodbye and I say hello.'

그림 3-21 학습 경과를 그래프로 표시(가로축은 학습 횟수, 세로축은 손실)



```
you [-0.9031807 -1.0374491 -1.4682057 -1.3216232 0.93127245]
say [ 1.2172916 1.2620505 -0.07845993 0.07709391 -1.2389531 ]
goodbye [-1.083403 -0.8826921 -0.33428606 -0.57201311 0.0488235 ]
and [ 1.0244362 1.0160093 -1.6284224 -1.6400533 -1.0564581 ]
i [-1.0642933 -0.9162385 -0.31357735 -0.57308311 0.041875 ]
hello [-0.9018145 -1.035476 1.4629668 -1.3058501 0.9280102]
. [1.0985303 1.1642815 1.4365371 1.3974973 -1.0714306]
```

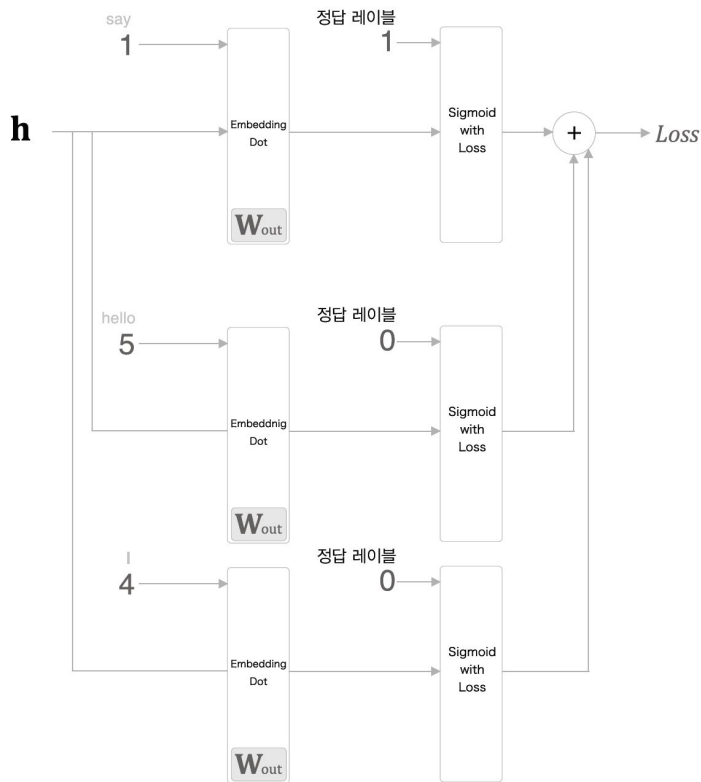


```
you [ 1.1629325 -0.9632237 1.0220319 -0.9973298 1.7453 ]
say [-0.6323519 1.2298591 0.02795994 1.1977036 1.124968 ]
goodbye [ 0.76684 -0.95298827 1.010701 -0.98307246 -0.8858874 ]
and [-1.5050563 0.99208957 1.668874 0.9707058 0.74477434 ]
i [ 0.7717827 -0.9794516 1.0309044 -0.99582094 -0.89843786 ]
hello [ 1.151623 -0.95551836 1.0052384 -0.9810654 1.769742 ]
. [ 1.6542563 1.1570057 -1.4423819 1.1536727 1.1364125 ]
```

-> 유사한 결과, 학습이 거듭될수록 손실이 줄어듦

-> 단어 ID의 분산 표현(밀집벡터) 출력

5. 개선된 CBOW (2권 - ch4)



개선 ① - Embedding 계층
개선 ② - Negative Sampling

- Data : PTB
- 매개변수 갱신 - Adam
- window_size = 5
- hidden_size = 100

< 거리 가까운 단어 출력 >

```
[query] you  
we: 0.6103515625  
someone: 0.59130859375  
i: 0.55419921875  
something: 0.48974609375  
anyone: 0.47314453125
```

```
[query] year  
month: 0.71875  
week: 0.65234375  
spring: 0.62744140625  
summer: 0.6259765625  
decade: 0.603515625
```

```
[query] car  
luxury: 0.497314453125  
arabia: 0.47802734375  
auto: 0.47119140625  
disk-drive: 0.450927734375  
travel: 0.4091796875
```

```
[query] toyota  
ford: 0.55078125  
instrumentation: 0.509765625  
mazda: 0.49365234375  
bethlehem: 0.47509765625  
nissan: 0.474853515625
```

교재 매개변수 사용 결과

```
[query] you  
we: 0.75537109375  
i: 0.70751953125  
they: 0.63525390625  
your: 0.5888671875  
someone: 0.5830078125
```

```
[query] year  
month: 0.85498046875  
week: 0.7900390625  
summer: 0.75146484375  
spring: 0.7509765625  
decade: 0.701171875
```

```
[query] car  
truck: 0.61376953125  
window: 0.60791015625  
cars: 0.57275390625  
luxury: 0.56884765625  
trucks: 0.5322265625
```

```
[query] toyota  
nissan: 0.61572265625  
cars: 0.6103515625  
beretta: 0.6103515625  
minivans: 0.60595703125  
tires: 0.6005859375
```

-> 같은 성격의 단어 출력

< 거리 가까운 단어 출력 >

```
[query] you  
we: 0.6103515625  
someone: 0.59130859375  
i: 0.55419921875  
something: 0.48974609375  
anyone: 0.47314453125
```

```
[query] year  
month: 0.71875  
week: 0.65234375  
spring: 0.62744140625  
summer: 0.6259765625  
decade: 0.603515625
```

```
[query] car  
luxury: 0.497314453125  
arabia: 0.47802734375  
auto: 0.47119140625  
disk-drive: 0.450927734375  
travel: 0.4091796875
```

```
[query] toyota  
ford: 0.55078125  
instrumentation: 0.509765625  
mazda: 0.49365234375  
bethlehem: 0.47509765625  
nissan: 0.474853515625
```

```
[query] you  
we: 0.75537109375  
i: 0.70751953125  
they: 0.63525390625  
your: 0.5888671875  
someone: 0.5830078125
```

```
[query] year  
month: 0.85498046875  
week: 0.7900390625  
summer: 0.75146484375  
spring: 0.7509765625  
decade: 0.701171875
```

```
[query] car  
truck: 0.61376953125  
window: 0.60791015625  
cars: 0.57275390625  
luxury: 0.56884765625  
trucks: 0.5322265625
```

```
[query] toyota  
nissan: 0.61572265625  
cars: 0.6103515625  
beretta: 0.6103515625  
minivans: 0.60595703125  
tires: 0.6005859375
```

교재 매개변수 사용 결과

-> 같은 성격의 단어 출력

<유추 문제>

```
[analogy] king:man = queen:?  
woman: 5.16015625  
veto: 4.9296875  
ounce: 4.69140625  
earthquake: 4.6328125  
successor: 4.609375
```

```
[analogy] take:took = go:?  
eurodollars: 5.5625  
points: 4.25  
began: 4.09375  
comes: 3.98046875  
oct.: 3.90625
```

```
[analogy] car:cars = child:?  
children: 5.21875  
average: 4.7265625  
yield: 4.20703125  
cattle: 4.1875  
priced: 4.1796875
```

```
[analogy] good:better = bad:?  
more: 6.6484375  
less: 6.0625  
rather: 5.21875  
slower: 4.734375  
greater: 4.671875
```

교재 매개변수 사용 결과

```
[analogy] king:man = queen:?  
woman: 5.26953125  
carolinas: 5.1328125  
a.m: 5.0234375  
toxin: 5.015625  
lady: 4.90234375
```

```
[analogy] take:took = go:?  
eurodollars: 5.5625  
're: 4.4140625  
went: 4.40234375  
came: 4.19140625  
were: 4.16796875
```

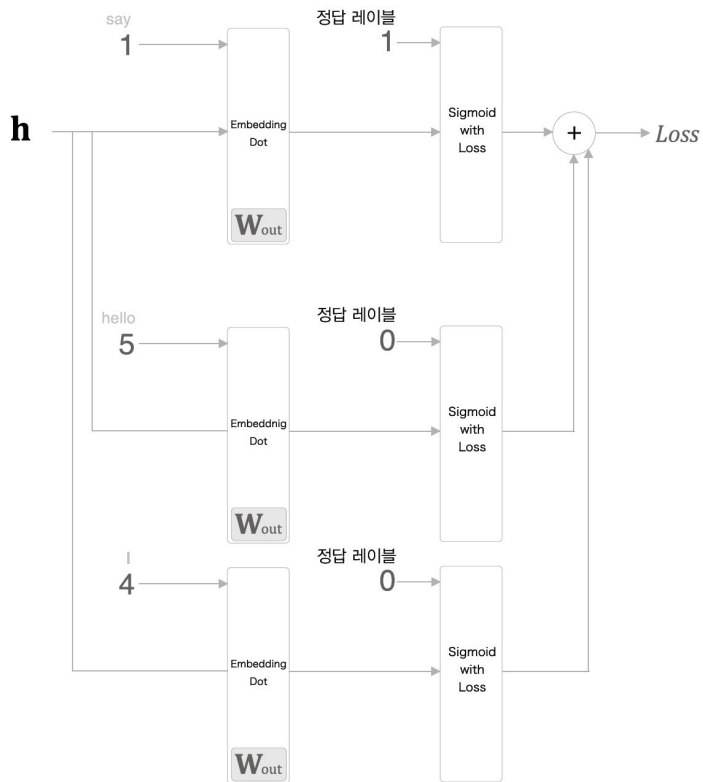
```
[analogy] car:cars = child:?  
a.m: 6.70703125  
rape: 6.23046875  
children: 5.3203125  
incest: 5.12890625  
women: 5.109375
```

```
[analogy] good:better = bad:?  
rather: 6.00390625  
more: 5.7890625  
less: 5.23046875  
greater: 4.46875  
fewer: 4.25390625
```

-> 4문제 중 1문제 정답

-> 교재) 4문제 중 3문제 정답

6. RNN (2권 - ch5)



개선 ① - Embedding 계층
개선 ② - Negative Sampling

- Data : PTB
- 매개변수 갱신 - Adam
- window_size = 5
- hidden_size = 100