

딥러닝 스터디 2주차

학습 관련 기술들 Part 3~5

변영무

순서

6-3. 배치 정규화

6-4. 오버피팅

6-5. 하이퍼 파라미터 최적화

6-3. 배치 정규화

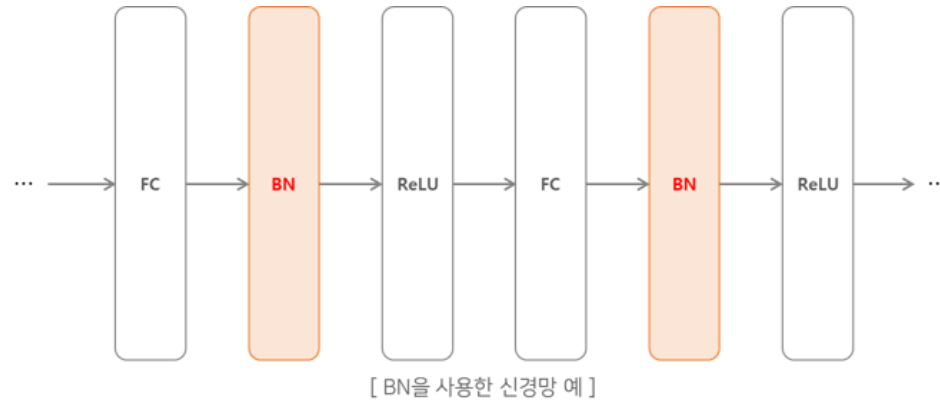
- 활성화 값을 어떻게 고르게 분포시킬까?

-> Xavier, He 초기값

- 활성화 값을 고르게 퍼지도록 **강제**하자

-> **배치 정규화**

6-3. 배치 정규화란?



$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad // \text{ mini-batch mean}$$
$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\mathcal{B}})^2 \quad // \text{ mini-batch variance}$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}}$$
$$y_i \leftarrow \gamma \hat{x}_i + \beta$$

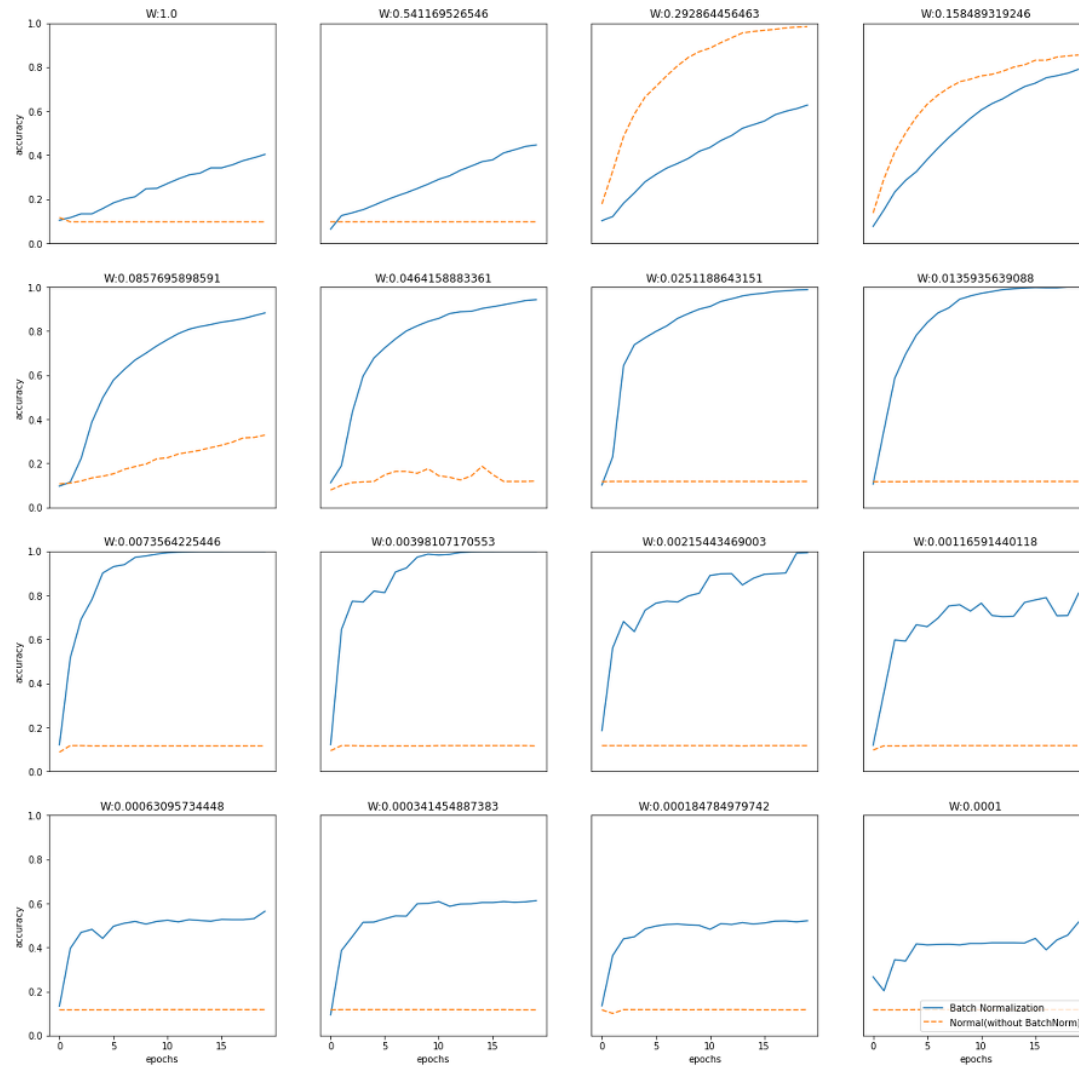
// normalize

// scale and shift

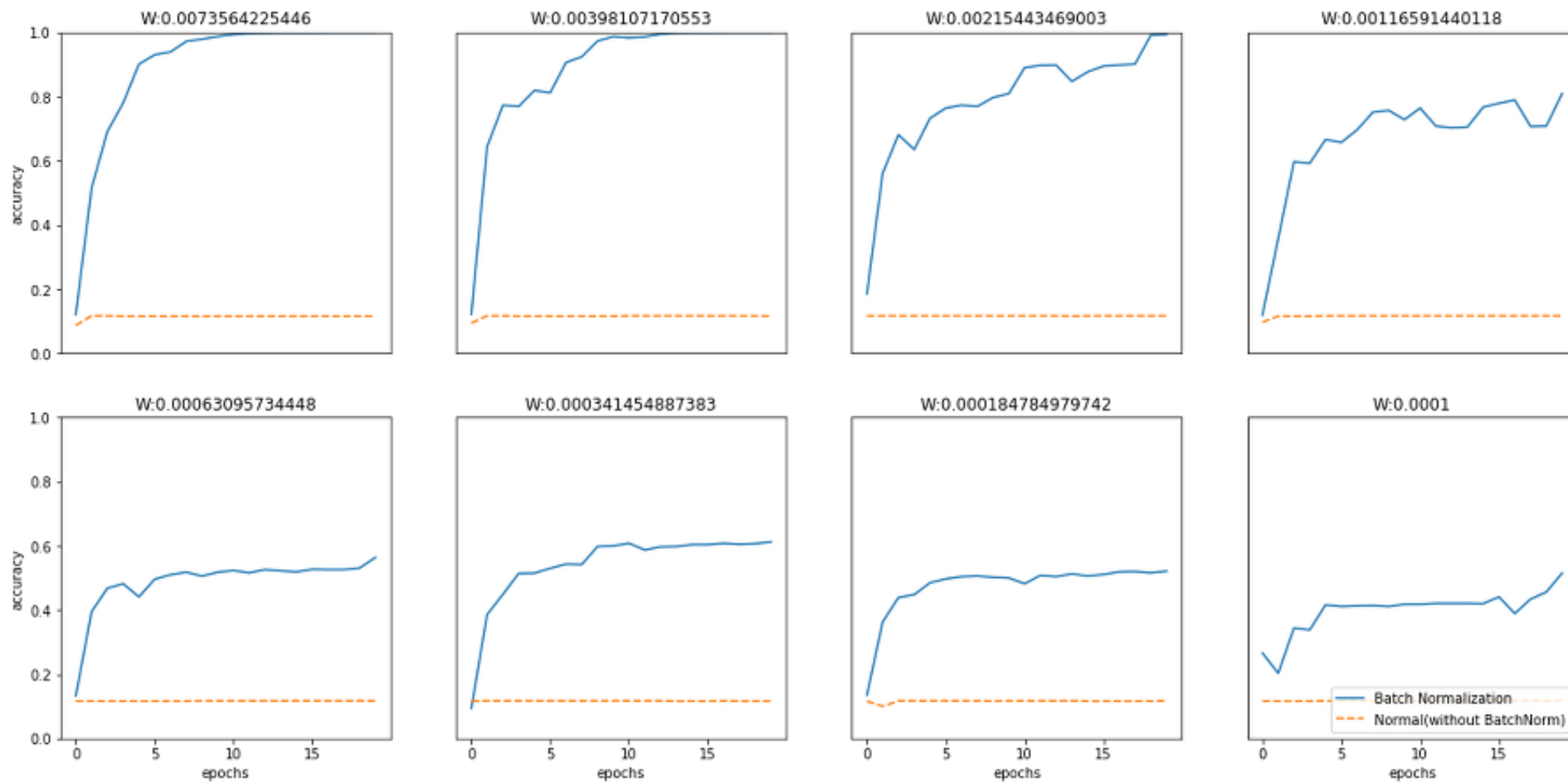
6-3. 배치 정규화의 장점

- 학습 속도 개선
- 초기값 선택에 의존하지 않는다
- 오버피팅 감소

6-3. 배치 정규화의 효과



6-3. 배치 정규화의 효과



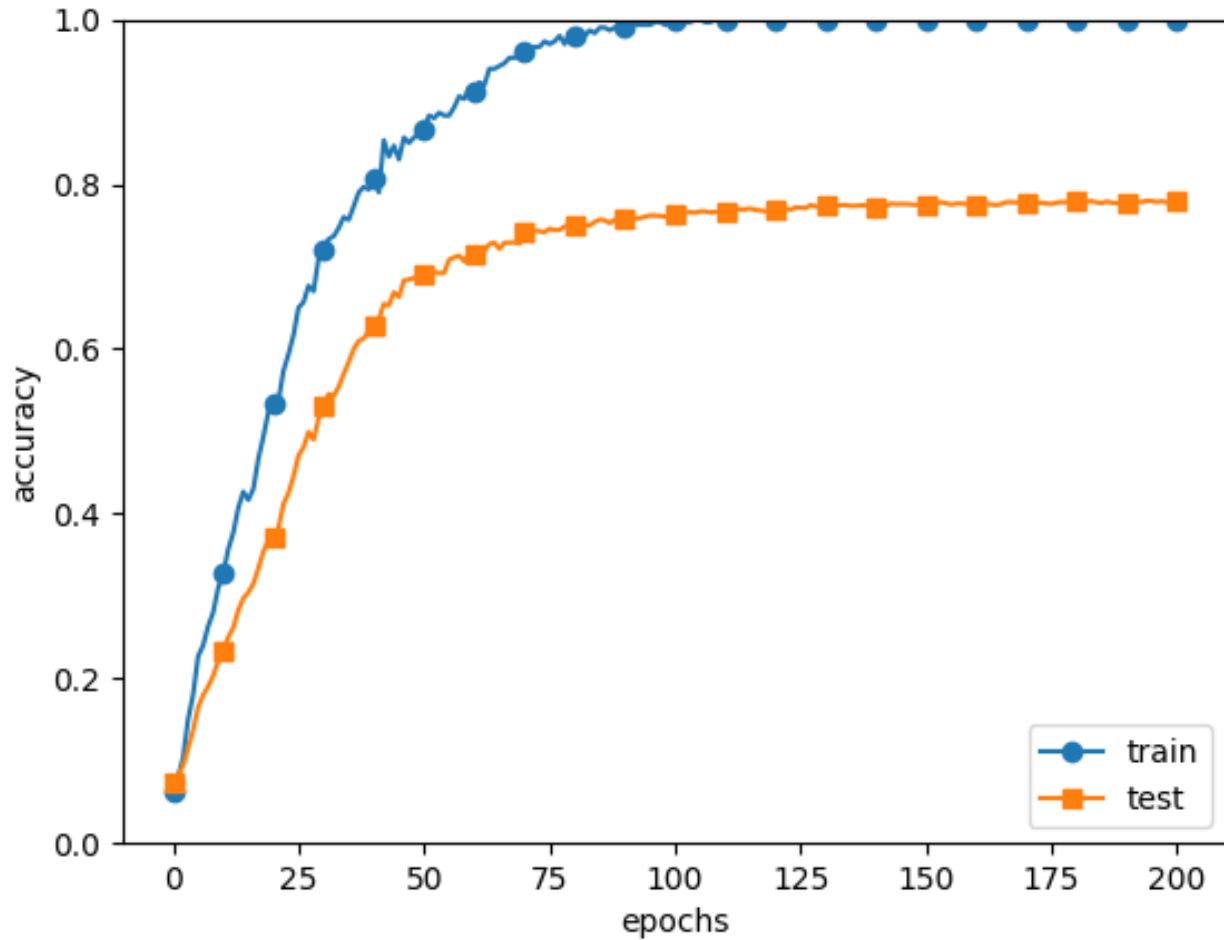
초기값이 잘못 설정된 경우

6-4. 오버피팅

Train set에 지나치게 적응된 상태

- 매개변수가 많고 표현력이 높은 모델
- 훈련 데이터가 적을 때

6-4. 오버피팅 그래프



- 300개의 Data Set
- 7 Layer Network

6-4. 오버피팅 해결 방법

- 가중치 감소
- 드롭 아웃

6-4. 오버피팅 - 가중치 감소

- 오버 피팅의 원인: 큰 가중치

-> 큰 가중치에는 상응하는 패널티를 주자

6-4. 오버피팅 - 가중치 감소

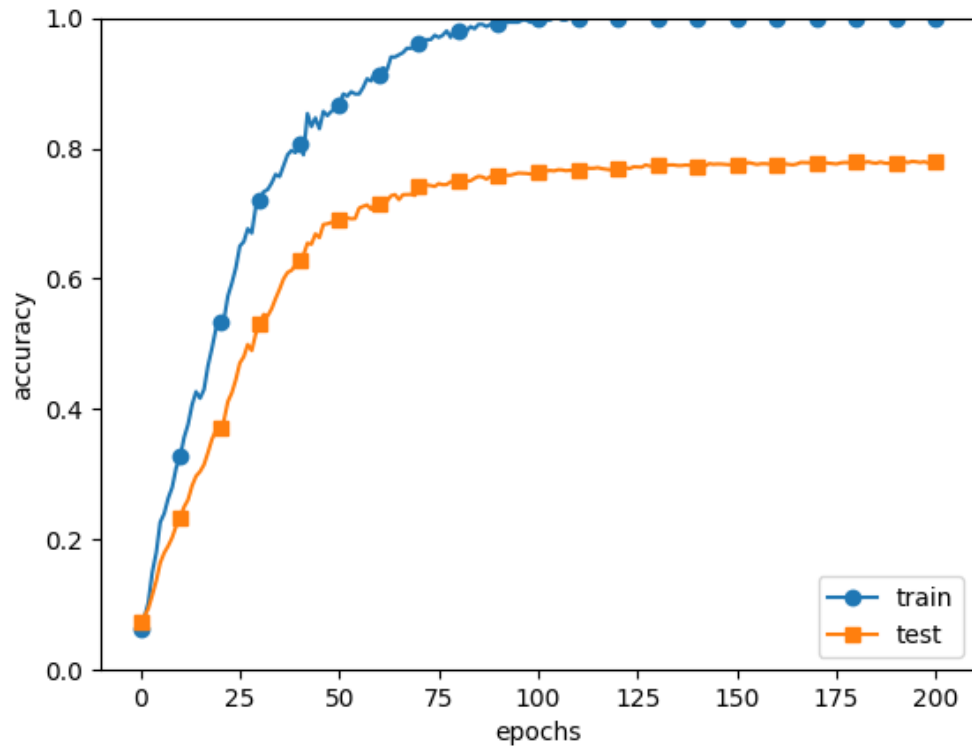
• ex)

$$C(x) + \frac{1}{2}\lambda W^2$$

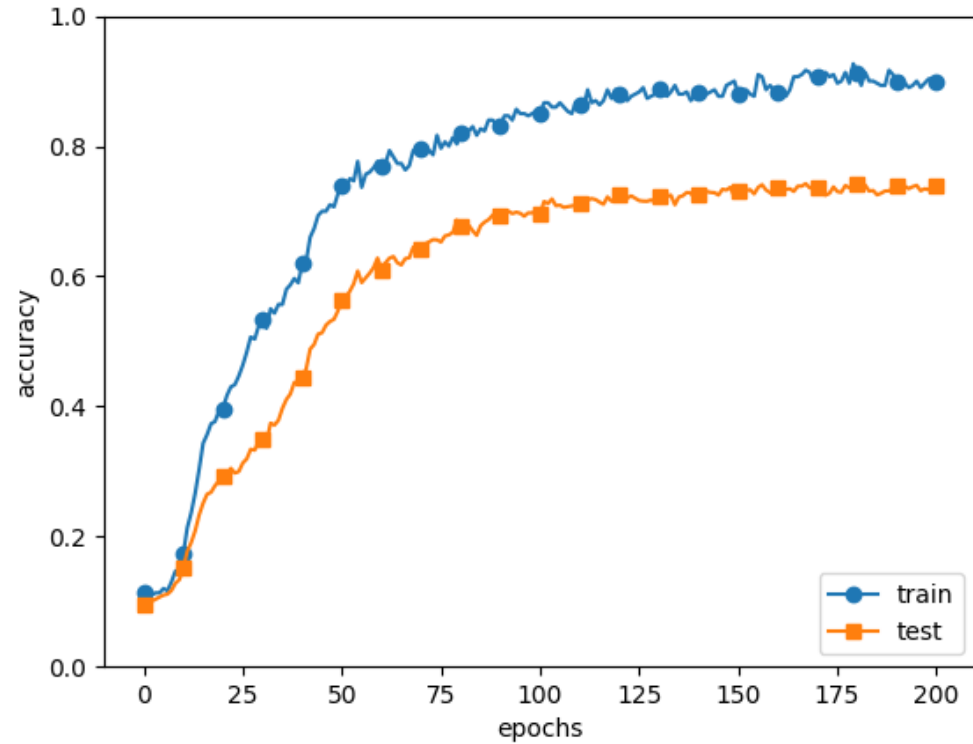
손실 함수

정규화 세기 조절

6-4. 오버피팅 - 가중치 감소

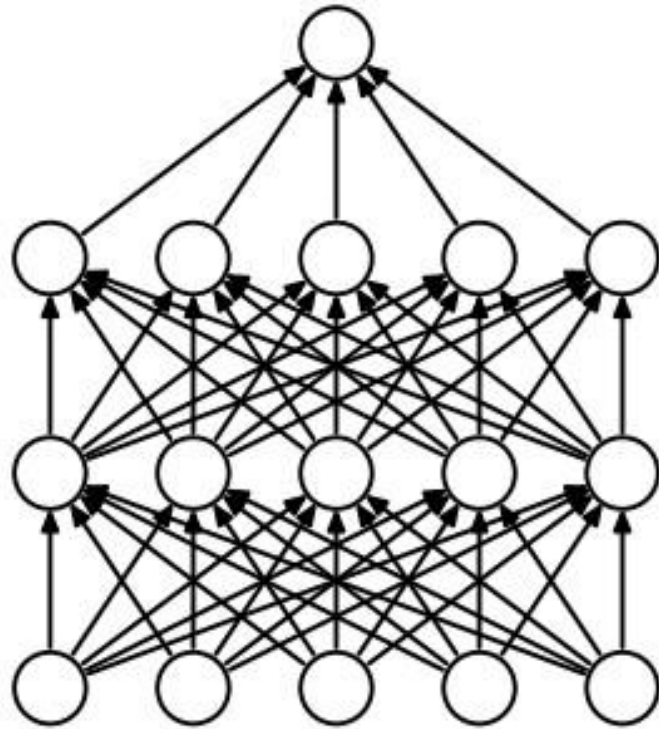


가중치 감소 전

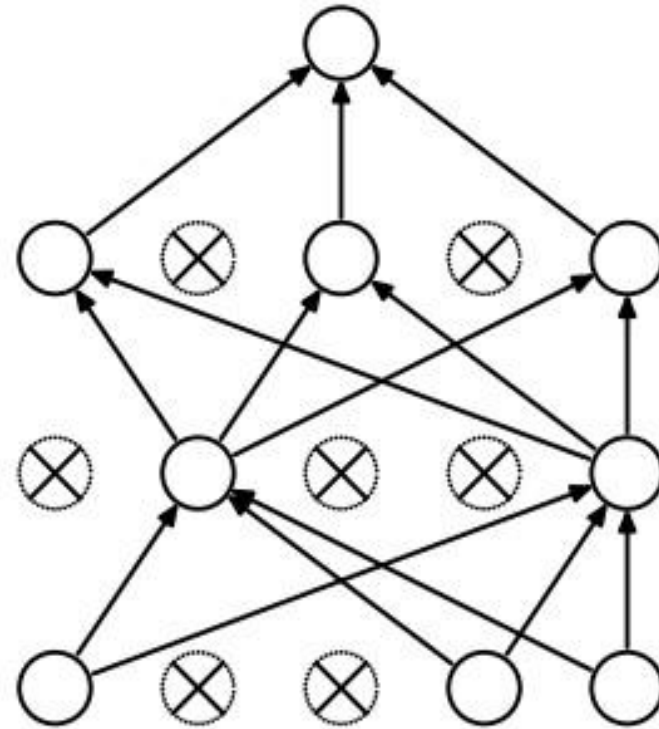


가중치 감소 후

6-4. 오버피팅 - 드롭아웃 아웃

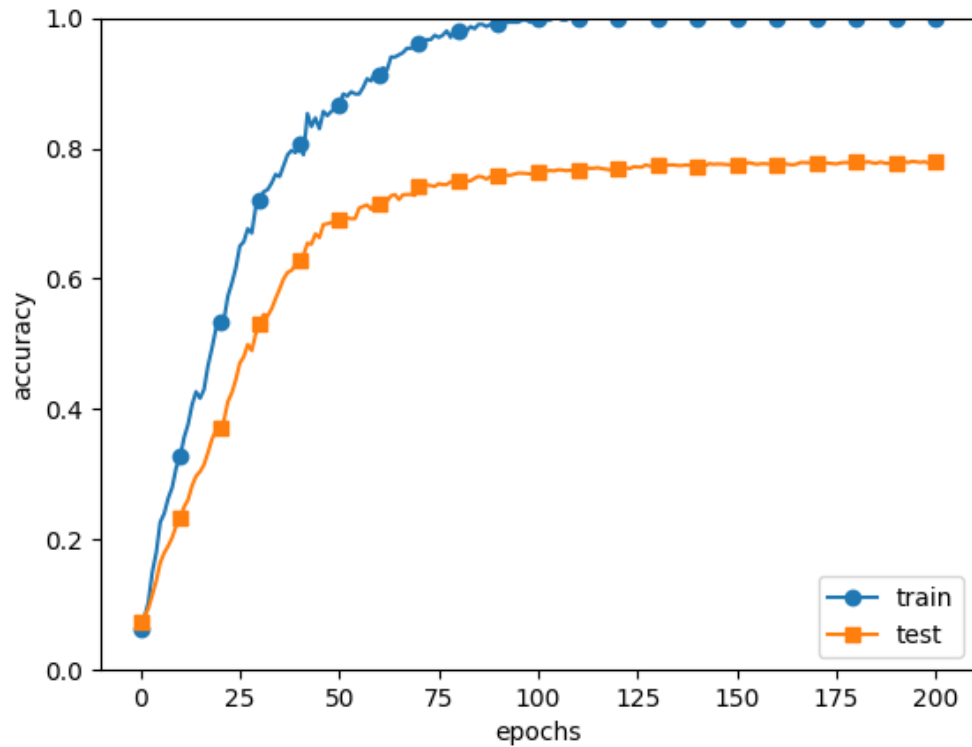


(a) Standard Neural Net

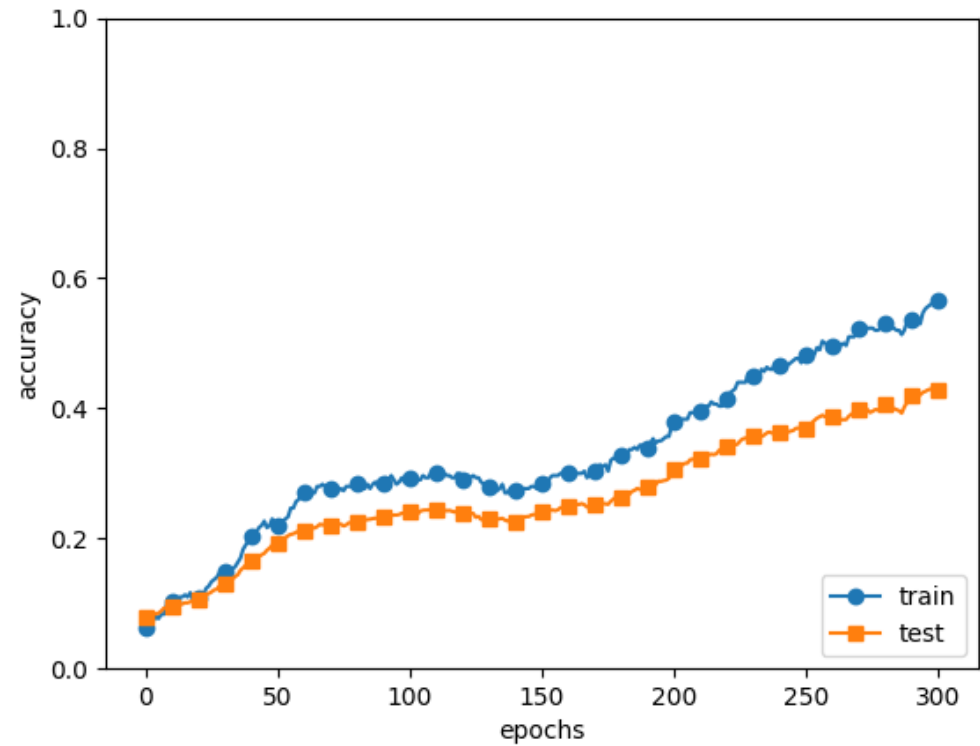


(b) After applying dropout.

6-4. 오버피팅 - 드롭아웃 아웃



드롭아웃 전

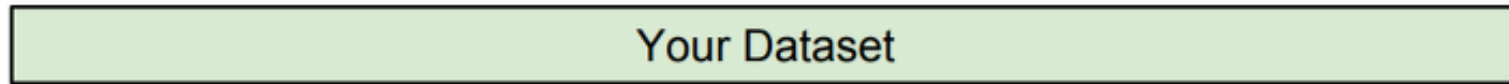


드롭아웃 후

6-5. 하이퍼 파라미터 최적화

Idea #1: Choose hyperparameters that work best on the data

BAD: $K = 1$ always works perfectly on training data



Idea #2: Split data into **train** and **test**, choose hyperparameters that work best on test data

BAD: No idea how algorithm will perform on new data



Idea #3: Split data into **train**, **val**, and **test**; choose hyperparameters on val and evaluate on test

Better!



최적의 하이퍼 파라미터를 찾을 data set

6-5. 하이퍼 파라미터 최적화

1. 하이퍼 파라미터 값의 범위 설정

- 대략적으로 지정하는 것이 효과적 (ex, Log scale)

2. 하이퍼 파라미터 값 무작위 추출

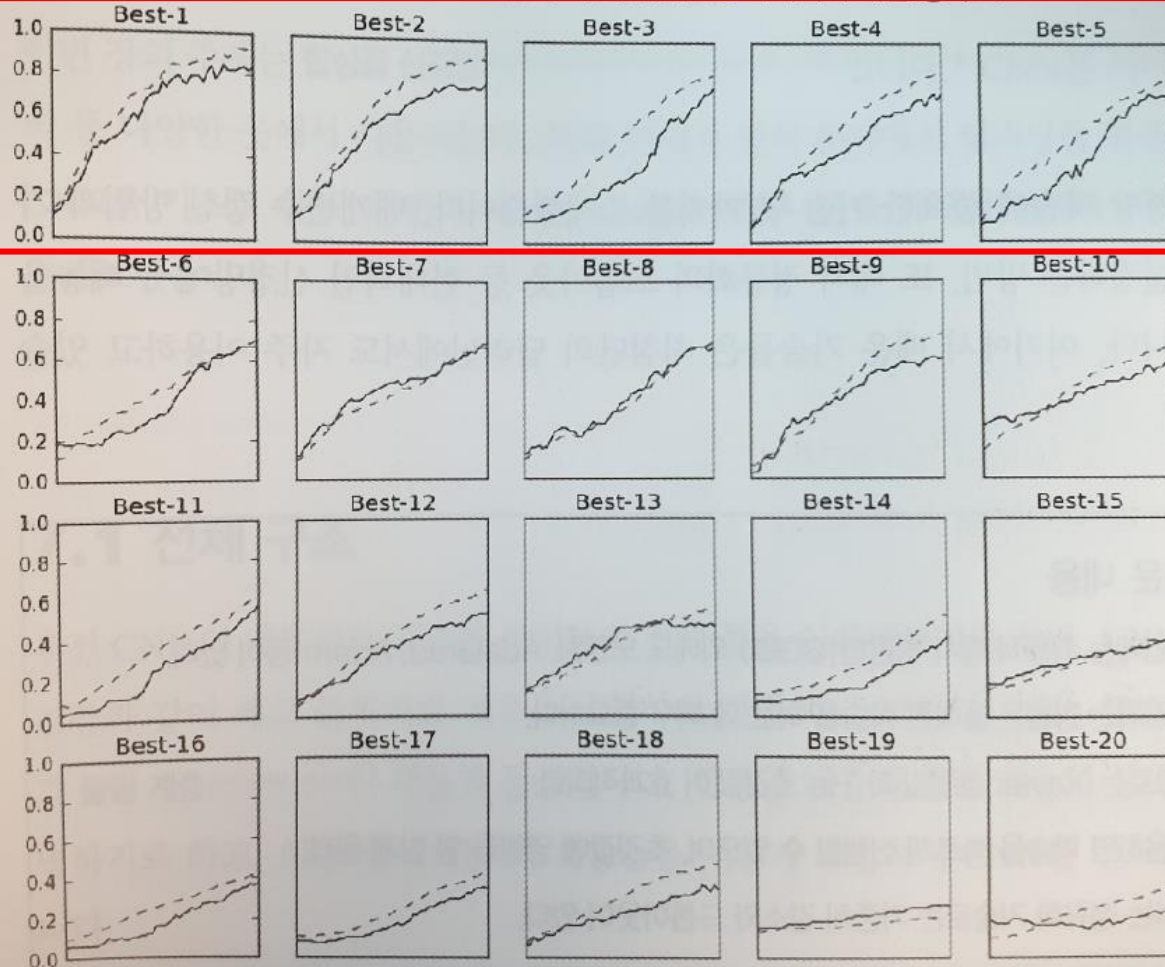
3. 2단계에서 추출한 값으로 학습 후 정확도 평가

- epoch 작게 지정 (빠른 학습)

4. 1~3단계 반복하여 하이퍼 파라미터 범위 좁힘

6-5. 하이퍼 파라미터 최적화

그림 6-24 실선은 검증 데이터에 대한 정확도 점선은 훈련 데이터에 대한 정확도



이 범위 내에서 하이퍼 파라미터 선택

Q&A