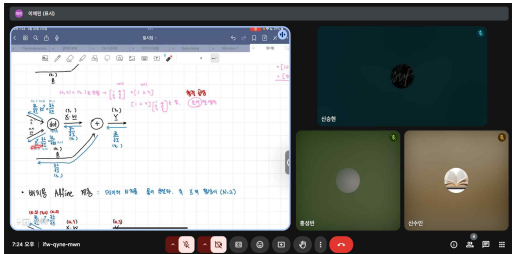


CUAI 밑바닥부터 시작하는 딥러닝 스터디 1팀

2025.05.13

발표자 : 이채린

스터디원 소개 및 만남 인증



스터디원 1 : 신수인

스터디원 2 : 신승현

스터디원 3 : 이채린

스터디원 4 : 전해인

스터디원 5 : 홍성빈

6.1 매개변수 갱신 - SGD

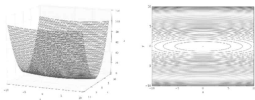
$$\mathbf{W} \leftarrow \mathbf{W} - \eta \frac{\partial L}{\partial \mathbf{W}}$$

\mathbf{W} = 갱신할 가중치 매개변수

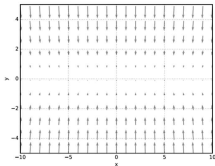
$\frac{\partial L}{\partial \mathbf{W}}$ = \mathbf{w} 에 대한 손실 함수의 기울기

η = 학습률 (0.01 or 0.001)

$f(x,y) = \frac{1}{20}x^2 + y^2$ 의 그래프와 등고선



기울기



최솟값이 되는 장소 $(x, y) = (0, 0)$

6.1 매개변수 갱신 - 모멘텀, AdaGrad, Adam

모멘텀

속도 변수를 갱신해준 뒤,
가중치를 더하여
가중치를 갱신하는 방식

$$\mathbf{v} \leftarrow \alpha \mathbf{v} - \eta \frac{\partial L}{\partial \mathbf{W}}$$

$$\mathbf{W} \leftarrow \mathbf{W} + \mathbf{v}$$

AdaGrad

개별 매개변수에
적응적으로 학습률을
조정하여 학습 진행

$$\mathbf{h} \leftarrow \mathbf{h} + \frac{\partial L}{\partial \mathbf{W}} \odot \frac{\partial L}{\partial \mathbf{W}}$$

$$\mathbf{W} \leftarrow \mathbf{W} - \eta \frac{1}{\sqrt{\mathbf{h}}} \frac{\partial L}{\partial \mathbf{W}}$$

\odot : 행렬의 원소별 곱셈

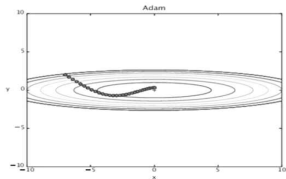
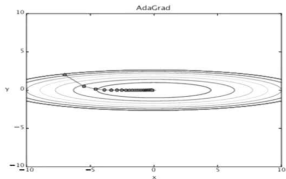
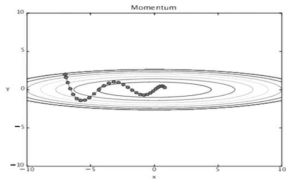
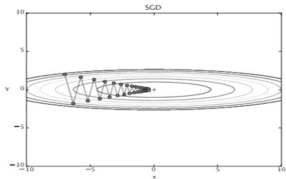
Adam

모멘텀+AdaGrad

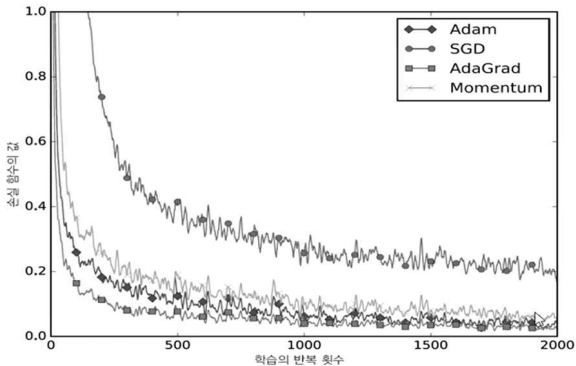
모멘텀처럼
공이 그릇바닥을
구르는 듯한 움직임
+

AdaGrad처럼
적응적으로 매개변수 갱신

6.1 매개변수 갱신

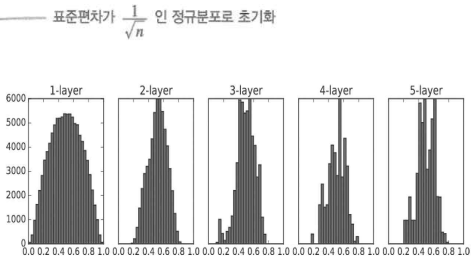
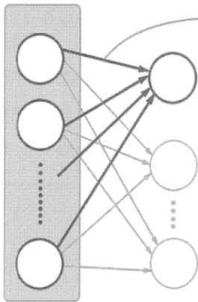


6.1 매개변수 갱신



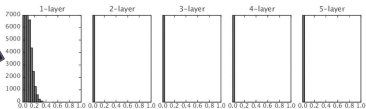
6.2 가중치의 초깃값 - Xavier

n 개의 노드

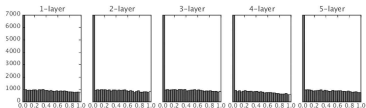
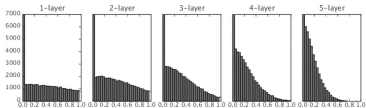


6.2 가중치의 초깃값 - He

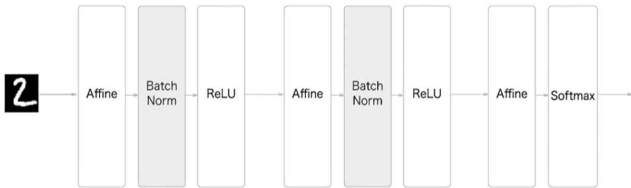
학습이 거의 이뤄지지 않음



기울기 소실 문제

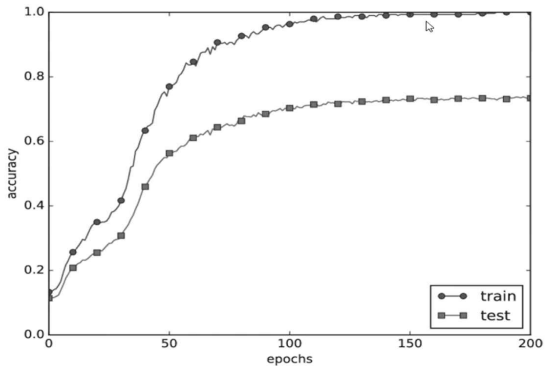


6.3 배치 정규화

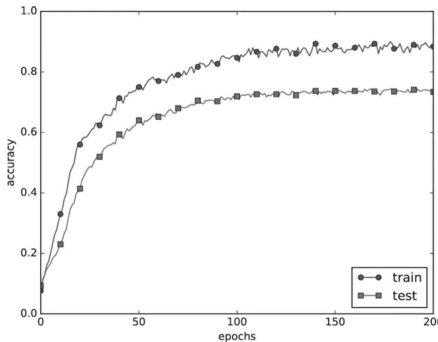


배치 정규화: 학습속도 빠름, 초깃값에 크게 의존하지 않음, 오버피팅 억제

6.4 바른 학습을 위해 - 오버피팅

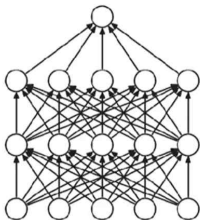


6.4 바른 학습을 위해 - 오버피팅

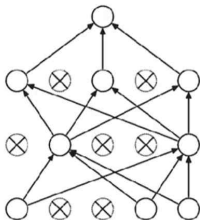


손실함수에
가중치의 제곱norm을 더해
가중치가 커지는 걸 억제

6.4 바른 학습을 위해 - 드롭아웃



(a) 일반 신경망



(b) 드롭아웃을 적용한 신경망

뉴런을 임의로 삭제하면서 학습하는 방법

6.5 적절한 하이퍼파라미터 값 찾기

