**4.1 데이터에서 학습한다!**

**4.1.1 데이터 주도 학습**

기계학습의 중심에는 데이터가 존재, 사람의 개입을 최소화

이미지에서 특징을 추출하고 그 특징의 패턴을 기계학습 기술로 학습

모아진 데이터로부터 규칙을 찾아내는 역할은 기계가 담당

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

신경망은 이미지에 포함된 중요한 특징까지도 기계가 스스로 학습할 것

**4.1.2 훈련 데이터와 시험 데이터**

기계학습 문제는 데이터를 훈련 데이터와 시험 데이터로 나눠 학습과 실험을 수행하는 것이 일반적

훈련 데이터만 사용하여 학습하면서 최적의 매개변수를 찾음 -> 시험 데이터를 사용하여 앞서 훈련한 모델의 실력을 평가

범용 능력을 제대로 평가하기 위해 훈련 데이터와 시험 데이터를 분리한다.

범용 능력은 아직 보지 못한 데이터로도 문제를 올바르게 풀어내는 능력이다.

한 데이터셋에만 지나치게 최적화된 상태를 오버피팅이라고 한다.

**4.2 손실 함수**

**4.2.1 오차제곱합**

폰트, 화이트, 상징, 텍스트이(가) 표시된 사진

자동 생성된 설명

<https://github.com/sinyebin/DeepLearning-from-Scratch/blob/master/4.%EC%8B%A0%EA%B2%BD%EB%A7%9D%ED%95%99%EC%8A%B5/sum_squares_error.ipynb>

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

오차제곱합 기준으로는 첫 번째 추정 결과가 정답에 더 가까울 것으로 판단

**4.2.2 교차 엔트로피 오차**

폰트, 화이트, 텍스트, 그래픽이(가) 표시된 사진

자동 생성된 설명

도표, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

X가 1일 때 y는 0이되고 x가 0에 가까워질수록 y의 값은 점점 작아진다.

정답에 해당하는 출력이 커질수록 0에 다가가다가, 그 출력이 1일 때 0이된다. 반대로 정답일 때의 출력이 작아질수록 오차가 커진다.

<https://github.com/sinyebin/DeepLearning-from-Scratch/blob/master/4.%EC%8B%A0%EA%B2%BD%EB%A7%9D%ED%95%99%EC%8A%B5/cross_entropy_error.ipynb>

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

오차 제곱합의 판단과 일치

**4.2.3 미니배치 학습**

폰트, 텍스트, 화이트, 그래픽이(가) 표시된 사진

자동 생성된 설명

데이터 일부를 추려 전체의 ‘근사치’로 이용 가능

신경망 학습에서도 훈련 데이터로부터 일부만 골라 학습을 수행하는 일부를 미니배치

그 일부만 학습하는 것을 미니배치 학습

<https://github.com/sinyebin/DeepLearning-from-Scratch/blob/master/4.%EC%8B%A0%EA%B2%BD%EB%A7%9D%ED%95%99%EC%8A%B5/4.2.3.ipynb>

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

무작위로 선택한 인덱스를 사용해 미니배치를 뽑아내고, 손실함수를 계산한다.

**4.2.4 (배치용) 교차 엔트로피 오차 구현하기**

def cross\_entropy\_error(y,t):

if y.ndim ==1:

t=t.reshape(1,t.size)

y=y.reshape(1,y.size)

batch\_size=y.shape[0]

return -np.sum(np.log(y[np.arange(batch\_size),t]))/batch\_size

**4.2.5 왜 손실 함수를 설정하는가?**

라인, 도표, 그래프, 평행이(가) 표시된 사진

자동 생성된 설명

계단 함수는 한순간의 변화를 일으키지만, 시그모이드 함수의 미분은 연속적으로 변하여 신경망이 올바르게 학습 가능하다.

**4.3 수치 미분**

**4.3.1 미분**

폰트, 라인, 텍스트, 화이트이(가) 표시된 사진

자동 생성된 설명

미분은 한순간의 변화량을 표시한 것이다.

텍스트, 폰트, 스크린샷, 라인이(가) 표시된 사진

자동 생성된 설명

이 방식은 반올림 오차를 일으킨다.

도표, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

위 그림과 같이 수치미분에는 오차가 포함된다.

이 차분은 x를 중심으로 그 전후의 차분을 계산한다는 의미에서 중심 차분 혹은 중앙 차분이라고 한다.

<https://github.com/sinyebin/DeepLearning-from-Scratch/blob/master/4.%EC%8B%A0%EA%B2%BD%EB%A7%9D%ED%95%99%EC%8A%B5/numerical_diff.ipynb>

**4.3.2 수치 미분의 예**

폰트, 텍스트, 타이포그래피, 화이트이(가) 표시된 사진

자동 생성된 설명

<https://github.com/sinyebin/DeepLearning-from-Scratch/blob/master/4.%EC%8B%A0%EA%B2%BD%EB%A7%9D%ED%95%99%EC%8A%B5/function_1.ipynb>

라인, 도표, 그래프, 평행이(가) 표시된 사진

자동 생성된 설명

라인, 도표, 그래프, 평행이(가) 표시된 사진

자동 생성된 설명

앞의 수치 미분과 결과를 비교하면 그 오차가 매우 적다.

**4.3.3 편미분**

친필, 폰트, 라인, 서예이(가) 표시된 사진

자동 생성된 설명

<https://github.com/sinyebin/DeepLearning-from-Scratch/blob/master/4.%EC%8B%A0%EA%B2%BD%EB%A7%9D%ED%95%99%EC%8A%B5/function_2.ipynb>

스케치, 디자인이(가) 표시된 사진

자동 생성된 설명

변수와 여럿인 함수에 대한 미분을 편미분이라고 하며, 3차원으로 그려짐

**4.4 기울기**

모든 변수의 편미분을 백터로 정리한 것을 기울기

<https://github.com/sinyebin/DeepLearning-from-Scratch/blob/master/4.%EC%8B%A0%EA%B2%BD%EB%A7%9D%ED%95%99%EC%8A%B5/numerical_gradient.ipynb>

기울기가 가리키는 쪽은 각 장소에서 함수의 출력 값을 가장 크게 줄이는 방향이다.

**4.4.1 경사법(경사 하강법)**

경사법: 기울기를 이용해 함수의 최소값을 찾으려는 것, 현 위치에서 기울어진 방향으로 일정거리만큼 이동 후 기울기를 계속 구하고 이러한 함수의 값을 점차 줄이는 것

폰트, 친필, 텍스트, 화이트이(가) 표시된 사진

자동 생성된 설명

학습률: 한번의 학습으로 얼마만큼 학습해야 할지, 매개변수 값을 얼마나 갱신하느냐를 정하는 것

<https://github.com/sinyebin/DeepLearning-from-Scratch/blob/master/4.%EC%8B%A0%EA%B2%BD%EB%A7%9D%ED%95%99%EC%8A%B5/gradient_descent.ipynb>

학습률이 너무 크면 큰값으로 발산하고, 너무 작으면 갱신되지 않은 채 끝남

**4.4.2 신경망에서의 기울기**

텍스트, 폰트, 친필, 도표이(가) 표시된 사진

자동 생성된 설명

여기에서 말하는 기울기는 가중치 매개변수에 대한 손실 함수의 기울기이다.

<https://github.com/sinyebin/DeepLearning-from-Scratch/blob/master/4.%EC%8B%A0%EA%B2%BD%EB%A7%9D%ED%95%99%EC%8A%B5/simpleNet.ipynb>

**4.5 학습 알고리즘 구현하기**

전제: 신경망에는 적응 가능한 가중치와 편향이 있고, 이 가중치와 편향을 훈련 데이터에 적응하도록 조정하는 과정을 ‘학습’이라 한다.

1단계 – 미니배치: 훈련 데이터 중 일부를 무작위로 가져옵니다. 이렇게 선별한 데이터를 미니배치라 하며, 그 미니배치의 손실 함수 값을 줄이는 것이 목표이다.

2단계 – 기울기 산출: 미니배치 손실 함수 값을 줄이기 위해 각 가중치 매개변수의 기울기를 구합니다. 기울기는 손실 함수의 값을 가장 작게 하는 방향을 제시한다.

3단계 - 매개변수 갱신: 가중치 매개변수를 기울기 방향으로 아주 조금 갱신한다.

4단계 – 반복: 1~3 단계를 반복한다.

데이터를 미니배치로 무작위로 선정하기 때문에 확률적 경사 하강법이라고 부른다.