

Chapter4 : 신경망 학습

4.1.1 데이터 주도 학습

- 기계학습은 데이터가 생명이다.
- 데이터에서 답을 찾고 데이터에서 패턴을 발견하고 데이터로 이야기를 만드는 것이 [기계 학습]이다.
- [기계학습]에서는 사람의 개입을 최소화하고 수집한 데이터로부터 패턴을 찾으려 시도합니다. 게다가 신경망과 딥러닝은 기존 기계학습에서 사용하던 방법보다 사람의 개입을 더욱 배제할 수 있게 해주는 중요한 특성을 지녔습니다.
- 이미지에서 [특징]을 추출하고, 그 특징의 패턴을 기계학습 기술로 학습하는 방법이 있습니다.
- [특징]은 입력 데이터에서 본질적인 데이터를 정확하게 추출할 수 있도록 설계된 변환기를 가리킵니다.
- 기계학습에서는 모아진 데이터로부터 규칙을 찾아내는 역할을 [기계]가 담당 => 효율 ↑
- 다만, 이미지를 벡터로 변환할 때 사용하는 특징은 여전히 '사람'이 설계하는 것. 즉, 문제에 적합한 특징을 쓰지 않으면 좀처럼 좋은 결과를 얻을수 없다.
- 특징과 기계학습을 활용한 접근에도 문제에 따라서는 '사람'이 적절한 특징을 생각해내야 한다.
- [분석방법]
 - 1) 데이터 - 사람이 생각한 특징 - 기계학습 (SVM, KNN 등) - 결과
 - 2) 데이터 - 신경망(딥러닝) - 결과
- 신경망은 이미지에 포함된 중요한 특징까지도 '기계'가 스스로 학습
- 딥러닝을 [종단간 기계학습]이라고도 합니다. 여기서 종단간은 '처음부터 끝까지'라는 의미로, 데이터(입력)에서 목표한 결과(출력)를 얻는다는 뜻을 담고 있죠.

4.1.2 훈련 데이터와 시험 데이터

- 기계학습 문제는 데이터를 1) 훈련데이터와 2) 시험 데이터로 나눠 학습과 실험을 수행
- 훈련데이터만 사용하여 학습하면서 [최적의 매개변수]를 찾는다.
- 그런 다음 [시험 데이터]를 사용하여 앞서 훈련한 모델의 실력을 평가하는 것
- [훈련 데이터]와 [시험 데이터]를 나누는 이유 = 범용적으로 사용할 수 있는 모델을 얻기 위해
- 기계학습의 가장 큰 문제점 = [오버피팅]
: 훈련데이터에 너무 맞춰져 있어 모델의 범용성이 없어지는 현상

4.2 손실함수

- 신경망 학습에서는 현재의 상태를 '하나의 지표'로 표현합니다. 그리고 그 지표를 가장 좋게 만들어주는 가중치 매개변수의 값을 탐색하는 것
- 신경망도 '하나의 지표'를 기준으로 최적의 매개변수 값을 탐색합니다. 신경망 학습에서 사용하는 지표는 [손실 함수]라고 합니다.
 - 1) 평균제곱오차, 2) 교차 엔트로피 교차

4.2.1 평균 제곱 오차

- 회귀 모델을 사용할 때 많이 쓰는 손실 함수
- [원 핫 인코딩] : 한 원소만 1로 하고 그 외는 0으로 나타내는 표기법
- 평균 제곱 오차의 값이 작을수록, 모델의 성능이 좋다고 평가한다.

4.2.2 교차 엔트로피 오차

- 평균 제곱 오차와 같이 값이 작을수록, 모델의 성능이 좋다고 평가한다.

4.2.3 미니배치 학습

- 기계학습 문제는 훈련 데이터를 사용해 학습
- 훈련 데이터에 대한 손실 함수의 값을 구하고, 그 값을 최대한 줄여주는 매개변수를 찾는다.
- 신경망 학습에서도 훈련 데이터로부터 일부분 골라 학습을 수행합니다. 이 일부를 [미니배

치]라고 합니다. 가령 60,000만장의 훈련 데이터 중에서 100장을 무작위로 뽑아 그 100장을 사용하여 학습하는 것. 이러한 학습 방법을 [미니배치 학습]이라고 합니다.

4.2.5 왜 손실 함수를 설정하는가?

- 신경망 학습에서는 최적의 매개변수(**가중치와 편향**)를 탐색할 때 손실 함수의 값을 가능한 한 작게 하는 매개변수 값을 찾습니다.
- 매개변수의 손실 함수의 미분이란 '가중치 매개변수의 값을 아주 조금 변화시켰을 때, 손실 함수가 어떻게 변하냐'라는 의미입니다.
- 만약 이 미분 값이 음수면, 그 가중치 매개변수를 양의 방향으로 변화시켜 손실 함수의 값을 줄일 수 있습니다.
- <신경망을 학습할 때 정확도를 지표로 삼아서 안 된다. 정확도를 지표로 하면 매개변수의 미분이 대부분의 장소에서 0이 되기 때문이다.>
- 만약 정확도가 지표였다면 가중치 매개변수의 값을 '조금' 바꾼다고 해도 정확도는 그대로 32%일 것입니다. 하지만 손실함수를 지표로 하면 매개변수의 값이 조금 변해도 그에 반응하여 손실 함수의 값도 변한다.
- 활성화 함수를 [계단 함수]로 사용해도 똑같은 이유로 신경망이 학습되지 않는다. 그 결과, 계단 함수를 이용하면 손실 함수를 지표로 삼는 게 아무 의미가 없게 된다.
- 그래서, 활성화 함수로 [시그모이드 함수]를 쓴다. (반드시 시그모이드를 신경망 학습의 활성화 함수로 쓴다는 것은 아니다.) 시그모이드 함수 함수의 미분은 계단 함수와 달리 어느 장소라도 0이 되지 않기 때문이다. 이는 신경망 학습에서 중요한 성질로, 기울기가 0이 되지 않는 덕분에 신경망이 올바르게 학습할 수 있는 것입니다.

4.4 기울기

4.4.1 경사법(경사 하강법)

- [최적]이란 손실 함수가 최솟값이 될 때의 매

개변수 값입니다.

- 여기서 주의할 점은 각 지점에서 함수의 값을 낮추는 방안을 제시하는 지표가 [기울기]라는 것입니다.
- cf) 경사법은 최솟값을 찾느냐, 최댓값을 찾느냐에 따라 이름이 다릅니다. 전자를 [경사 하강법], 후자는 [경사 상승법]이라고 한다. 다만 손실 함수의 부호를 반전시키면 최솟값을 찾는 문제와 최댓값을 찾는 문제는 같은 것이니 하강이나 상승이라는 본질적으로 중요하지 않습니다. [일반적으로 신경망 분야에서의 경사법은 '경사 하강법'으로 등장할 때] 많습니다.
- [학습률 - learning rate] : 한 번의 학습으로 얼마만큼 학습해야 할지, 즉 매개변수 값을 얼마나 갱신하느냐를 정하는 것이 학습률
- [학습률]이 너무 크면 값이 폭주하고, 작으면 학습이 매우 느리다. 학습률은 [하이퍼파라미터]의 일부다.
- 학습률과 같이 인간이 직접 지정하는 매개변수를 [하이퍼파라미터]라고 한다. 신경망의 가중치 매개변수는 훈련 데이터와 학습 알고리즘에 의해서 '자동'으로 획득되는 매개변수인 반면, 학습률 같은 하이퍼파라미터는 사람이 직접 설정해야 하는 매개변수이다.

4.5 학습 알고리즘 구현하기

[신경망의 학습 절차]

1) 전제

- 신경망에는 적응 가능한 가중치와 편향이 있고, 이 가중치와 편향을 훈련 데이터에 적응하도록 조정하는 과정을 '학습'이라고 한다. 신경망 학습은 다음과 같이 4단계로 수행.

2) 1단계 - 미니배치

: 훈련 데이터 중 일부를 무작위로 가져온다. 이렇게 선별한 데이터를 미니배치라고 하며, 그 미니배치의 손실 함수 값을 줄이는 것을 목표로.

3) 2단계 - 기울기 산출

: 미니배치의 손실 함수 값을 줄이기 위해 각 가중치 매개변수의 기울기를 구합니다. 기울기는 손실 함수의 값을 가장 작게 하는 방향을 제시.

4) 매개변수 갱신

: 가중치 매개변수를 기울기 방향으로 아주 조금 갱신.

5) 반복 : 1~3단계를 반복

- 이때, 데이터를 미니배치로 무작위로 선정하기 때문에 [확률적 경사 하강법]이라고 함.

4.5.3 시험 데이터로 평가하기

- 신경망 학습의 원래 목표는 범용적인 능력을 키우는 것입니다. 이를 위해 다음 구현에서는 학습 도중 정기적으로 훈련 데이터와 시험 데이터를 대상으로 정확도를 기록합니다. 여기에서는 1에폭별로 훈련 데이터와 시험 데이터에 대한 정확도를 기록합니다.
- [에폭]은 하나의 단위입니다. 1에폭은 학습에서 훈련데이터를 모두 소진했을 때의 횟수에 해당합니다. 예를 들어, 10,000개를 100개의 미니배치로 학습할 경우, 확률적 경사 하강법을 100회 반복하면 모든 훈련 데이터를 '소진'한 게 됩니다. 이 경우 100회가 1에폭이 됩니다.
- 정확도를 1에폭마다 계산하는 이유는 for 문 안에서 매번 계산하기에는 시간이 오래 걸리고, 또 그렇게까지 자주 기록할 필요도 없기 때문이죠.
- 또, train data와 test data 사이의 차이가 없음은 학습에서 오버피팅이 일어나지 않았다는 것이다.
-