

chapter5 : 오차역전파법

- 신경망의 가중치 매개변수 기울기는 수치 미분을 사용해서 구했습니다.
- 수치 미분은 단순하고 구현하기 쉽지만 계산 시간일 오래 걸린다는 게 단점입니다.
- 그래서 가중치 매개변수의 기울기를 효율적으로 계산하는 [오차역전파법]을 사용합니다.

5.1 계산 그래프

- [계산 그래프]는 계산 과정을 그래프로 나타낸 것입니다. 여기에서의 그래프는 우리가 잘 아는 그래프 자료구조로, 복수의 1) 노드와 2) 에지로 표현합니다. (노드 사이의 직선을 '에지'라고 합니다.)

5.1.1 계산 그래프로 풀다

- 노드는 원으로 표기하고 원 안에 연산 내용을 적습니다. / 계산 결과를 화살표로.
- 계산을 왼쪽에서 오른쪽으로 진행하는 단계를 [순전파]라고 합니다.
- 반대방향으로 진행하는 단계를 [역전파]라고 합니다.

5.1.2 국소적 계산

- [국소적 계산]은 결국 전체에서 어떤 일이 벌어지든 상관없이 자신과 관계된 정보만을 다음 결과를 통해 출력할 수 있다는 것입니다.
- 각 노드에서의 계산은 국소적 계산이라는 점입니다. 다시 말하점, 각 노드는 자신과 관련한 계산 외에는 아무것도 신경 쓸 게 없습니다.

5.1.3 왜 계산 그래프로 푸는가?

- 다시! 계산 그래프의 이점은? : [국소적 계산]
- 계산 그래프는 중간 계산 결과를 모두 보관할 수 있습니다.
- 실제 계산 그래프를 사용하는 가장 큰 이유는 역전파를 통해 '미분'을 효율적으로 계산할 수 있는 점에 있습니다.

5.2 연쇄법칙

- 한편 역전파는 '국소적인 미분'을 순방향과는 반대인 오른쪽에서 왼쪽으로 전달합니다. 또한, 이 '국소적 미분'을 전달하는 원리는 [연쇄법칙]에 따른 것입니다.

5.2.1 계산 그래프의 역전파

5.2.2 연쇄 법칙이란?

- 합성 함수란 여러 함수로 구성된 함수입니다.
- 합성 함수의 미분은 합성 함수를 구성하는 각 함수의 미분의 곱으로 나타낼 수 있다.

5.3 역전파

5.3.1 덧셈 노드의 역전파

- 즉, 덧셈 노드의 역전파는 1을 곱하기만 할 뿐이므로 입력된 값을 그대로 다음 노드로 보내게 됩니다.

- 이 예에서 상류에서 전해진 미분 값을 $\frac{\alpha L}{\alpha z}$ 이라 했는데, 이는 [그림 5-10]과 같이 최종적으로 L 이라는 값을 출력하는 큰 계산 그래프를 가정하기 때문입니다.

5.3.2 곱셈 노드의 역전파

5.4 단순한 계층 구현하기

5.4.1 곱셈 계층

- forward()는 순전파, backward()은 역전파
- 코드가 다소 길어졌습지만, 하나하나의 명령은 단순합니다. 필요한 계층을 만들어 순전파 메서드인 forward()를 적절한 순서로 호출합니다. 그런 다음 순전파와 반대 순서로 역전파 메서드인 backward()를 호출하면 원하는 미분이 나옵니다.

5.5 활성화 함수 계층 구현하기

5.5.1 ReLU 계층

- 순전파 때의 입력인 x 가 0보다 크면 역전파는 상류의 값을 그대로 하류로 흘립니다.
- 반면, 순전파 때 x 가 0 이하면 역전파 때는 하류로 신호를 보내지 않습니다.

5.5.2 Sigmoid 계층

- sigmoid 계층의 역전파는 순전파의 출력(y)만으로 계산할 수 있습니다.

5.6 Affine / Softmax 계층 구현하기

5.6.1 Affine 계층

- 신경망의 순전파 때 수행하는 행렬의 내적은 기하학에서는 [어파인 변환]이라고 합니다. 그래서 이 책에서는 어파인 변환을 수행하는 처리를 'Affine 계층'이라는 이름으로 구현합니다.

5.6.2 배치용 Affine 계층

- Affine 계층은 입력데이터 X 하나만을 고려하는 것이었습니다.
- 그와 대비되게, 배치용 Affine 계층은 데이터 N 개를 묶어 순전파한다.

5.6.3 Softmax-with-Loss 계층

- 이 소프트맥스 함수는 입력 값을 정규화합니다.
- 신경망 학습에서 수행하는 작업은 1) 학습과 2) 추론이 있습니다. 추론할 때는 일반적으로 Softmax 계층을 사용하지 않습니다. 예컨대 [그림 5-28]의 신경망은 추론할 때는 마지막 Affine 계층의 출력을 인식 결과로 이용합니다. 또한, 신경망에서 정규화하지 않는 출력결과([그림 5-28])에서는 Softmax 앞의 Affine 계층의 출력)를 점수라고 합니다. 즉, 신경망 추론에서 답을 하나만 내는 경우에는 가장 높은 점수만 알면 되니 Softmax 계층은 필요 없다는 것이죠. 반면, 신경망을 학습할 때는 Softmax 계층이 필요합니다.
- 여기에서는 3클래스 분류를 가정하고 이전 계층에서 3개의 입력(점수)을 받습니다. 그림과 같이 Softmax 계층은 입력을 정규화하여 $y(n)$ 을 출력합니다. Cross Entropy Error 계층은 Softmax의 출력과 정답 레이블을 받고, 이들 데이터로부터 손실 L 을 출력합니다.
- 소프트맥스 함수는 손실 함수로 주로 '교차 엔트로피 오차'를 사용합니다.

5.7 오차역전파법 구현하기

5.7.1 신경망 학습의 전체 그림

- 오차역전파법이 등장하는 단계는 두 번째인 '기울기 산출'입니다.

5.7.2 오차역전파법을 적용한 신경망 구현하기

- 신경망의 계층을 OrderedDict에 보관하는 점이 중요합니다. (코드 참고)
- OrderedDict는 순서가 있는 딕셔너리입니다.
- 그래서 순전파 때는 추가한 순서대로 각 계층의 forward() 메서드를 호출하기만 하면 처리가 완료됩니다. 마찬가지로 역전파 때는 계층을 반대 순서로 호출하기만 하면 됩니다.