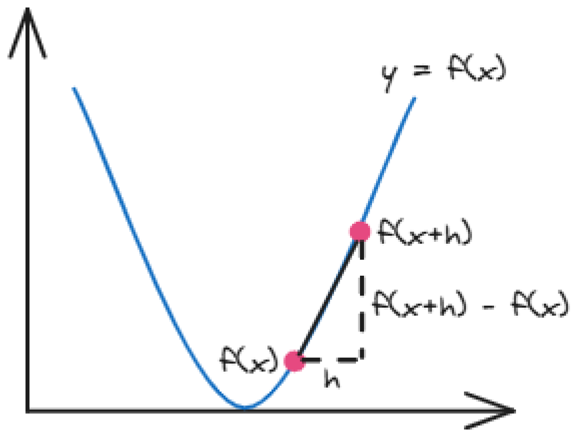
3. 미분 기초

1. 일 변수 함수 미분

- 도함수 (미분)이란 :

  $\lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = f'(x)$



- 평균값 정리

  함수 f(x)가 구간 [a,b]에서 미분 가능하면

  $\frac{f(b)-f(a)}{(b-a)} = f'(c), (a < c < b)$

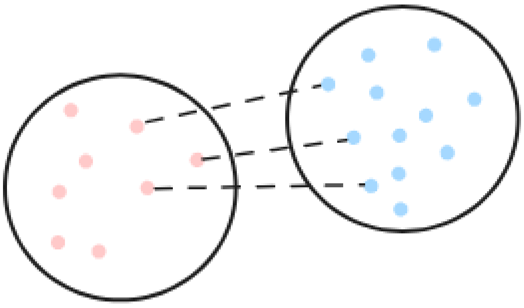
  를 만족시키는 c가 존재한다.

도함수가 머신러닝에 필요한 이유

손실 함수

- 모델이 예측한 답과 실제 답 사이의 차이를 측정.

- 앞서 봤던 스펙트럴 클러스터링을 보자면, 분류가 얼마나 정확하게 됐는지 보기 위해서 각 그룹에 있는 노드 사이의 거리를 측정 했었다.



이 때, 두 노드 사이의 거리를 측정하는데 사용된 거리 척도(함수)를 '손실 함수'라고 한다.

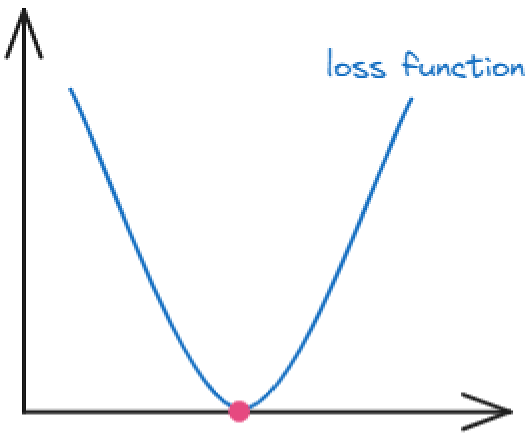
모델 수정 (역전파)

- 모델이 예측한 값과 실제 값 사이의 차이를 측정하는 것을 '손실 함수'라고 했다.
- 모델의 목표는 '손실 함수'를 최소화하는 모델을 찾는 것이다.

머신러닝의 학습 과정

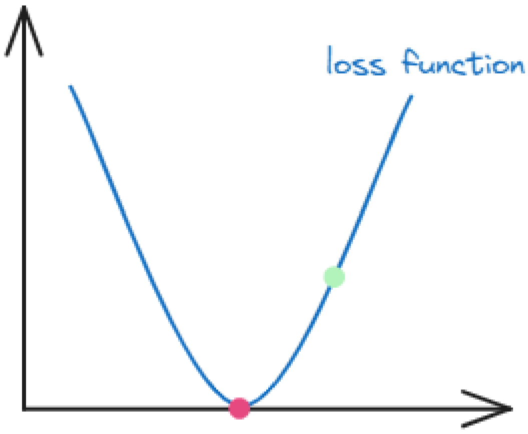
데이터 입력 - 모델 예측 - 실제 답과 비교 - 틀린만큼 모델 수정 - 다시 데이터 입력 ...

위 과정을 반복하게 된다. 이 때, '틀린만큼 모델을 수정'하는 과정에서 도함수가 필요하게 된다.



그래프로 표현하자면 위 그래프가 최소화 되는 지점, 붉은 점의 위치를 찾는 것이 목표이다.

만약 이번 예측의 손실 값이 아래 그래프와 같이 초록 점의 값이 나타났다고 해보자.



여기서 loss를 최소화하기 위해선 모델을 수정해서 loss 값이 빨간 점 쪽으로 이동하도록 해야 한다.

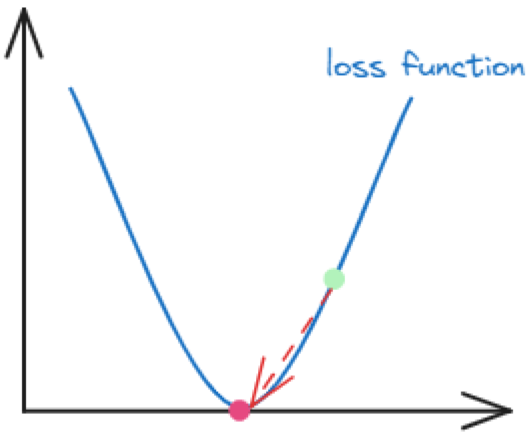
그러기 위해선 그래프에서 어느 방향으로 이동해야 하는 지를 알아내야 한다.

방향을 알아내기 위해선 아래와 같은 수식을 사용하면 된다.

$x' = x - \alpha \frac{df(x)}{dx}$ (x는 모델)

여기서 주목할 것은 예측 값에서 도함수 값을 뺀다는 것.

도함수는 해당 지점에서의 **순간 변화율**을 나타낸다. 즉, 그 뒤로 증가한다면 양수, 감소한다면 음수의 값을 갖는다. 따라서, 도함수의 값을 뺀다는 것은 loss가 감소하는 방향으로 이동하는 것과 같다!

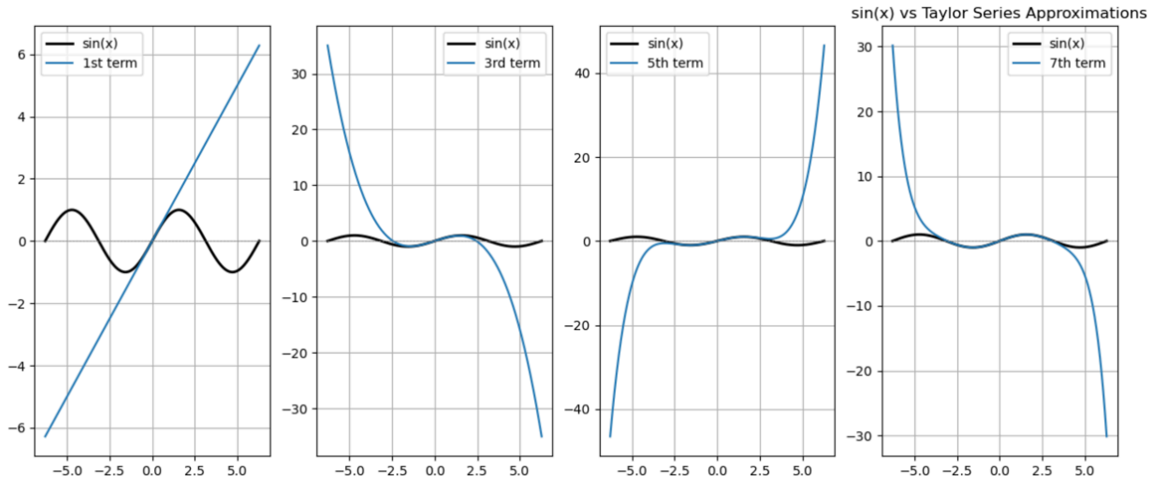


- 여기서 알 수 있듯, 머신러닝에서 미분은 모델의 학습을 위해 필수적인 요소이다!

2. 테일러 급수와 2차 미분 테스트

- 테일러 급수 : 어떤 함수 $y = f(x)$가 $x = x^*$에서 무한 미분 가능하면 함수 $y = f(x)$를 아래 식과 같이 무한한 다항식의 합들로 표시할 수 있다.

$f(x) = f(x^*) + f'(x^*)(x - x^*) + \frac{f''(x^*)}{2!}(x - x^*)^2 + \ldots$
$= \sum_{n=0}^{\infty} \frac{f^{(n)}(x^*)}{n!}(x - x^*)^n$



- 구간 $[a, b]$에서 n회 미분 가능하고 $x^*, x \in [a, b]$, $x \neq x^*$이면 다음 식을 만족하는 $0 < \theta < 1$이 존재한다.

  $f(x) = f(x^*) + f'(x^*)(x - x^*) + \frac{f''(x^*)}{2!}(x - x^*)^2 + \ldots + \frac{f^{(n-1)}(x^*)}{(n-1)!}(x - x^*)^{n-1} + \frac{f^{(n)}(x^* + \theta(x - x^*))}{n!}(x - x^*)^n$

- 2차 미분 테스트 : 테일러 급수를 이용해 함수의 최댓값과 최솟값을 계산한다.

  2차 미분 가능한 함수 f(x)에 대하여,

$$f(x) = f(x^*) + f'(x^*)(x - x^*) + \frac{f''(x^* + \theta(x - x^*))}{2}(x - x^*)^2 \cdots \textbf{(1)}$$

단, $0 < \theta < 1$

여기서 $f'(x^*) = 0$일 때, 위 함수는

$$f(x) = f(x^*) + \frac{f''(x^* + \theta(x - x^*))}{2}(x - x^*)^2 \cdots \textbf{(2)}$$

단, $0 < \theta < 1$

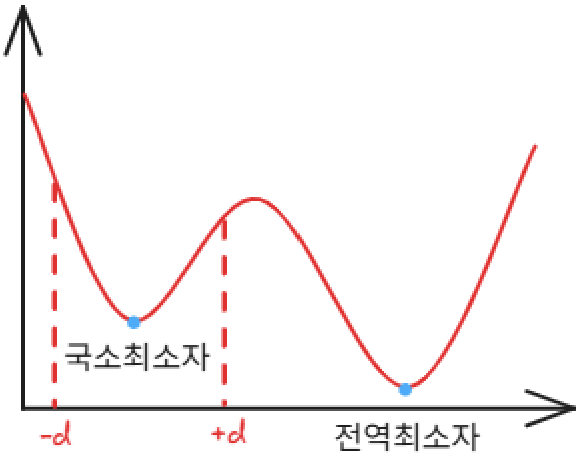
여기서 맨 끝 항 $(x - x^*)^2 >= 0$ 이므로, $f(x)$와 $f(x^*)$의 관계는 $f''(x)$에 의해 결정된다. ($f(x) = f(x^*) + f''(x)$)

따라서 $f''(x^* + \theta(x - x^*)) > 0$일 경우, $f(x^*)$은 $f(x)$의 최솟값이 되며,

$f''(x^* + \theta(x - x^*)) < 0$일 경우, $f(x^*)$은 $f(x)$의 최댓값이 된다.

→ f′(x)=0 : 최대 or 최솟값. → f′′(x) > 0 : 최솟값 / f′′(x) < 0 : 최댓값

- **전역 최소자와 국소 최소자**



- **임계점** : 미분해서 0이 되는 지점

  $f'(x^*)$가 존재하고 $f'(x^*) = 0$이다.

  $f'(c)$

테일러 급수의 활용

- 머신러닝에서 함수의 형태를 근사하고, 최적화 과정을 수치적으로 계산할 수 있게 해주는 이론적 도구.
  - 선형 회귀에서 손실 함수를 근사할 때, 실제로 테일러 급수를 이용해 2차 도함수를 계산하여 손실을 최소화하는 방식을 사용하기도 한다. (newton 방식)
  - 복잡한 함수의 계산을 단순화할 때, 테일러 급수를 이용해 다항식으로 근사하여 계산을 빠르게 하거나 이론 해석을 쉽게 만든다.