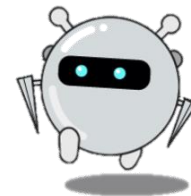


경사하강법



진행자 : 현시은

Overview

- Overview

손실함수 복습

경사하강법

Hyper parameter

경사하강법 한계

우리의 목표

손실함수를 최소화 하는 모델을 찾아내는 것!

손실함수를 최소화 하는 모델의 파라미터를 찾아내는 최적화 기법 중 하나가 경사하강법!

손실함수 복습

Overview

- 손실함수 복습

경사하강법

Hyper parameter

경사하강법 한계

손실함수란?

- 손실함수(Loss function) : 모델의 예측값과 실제 값의 다른 정도를 측정하는 함수이다. 이 함수는 모델이 얼마나 잘 혹은 못하고 있는지를 수치적으로 나타낸다.

(ex) 잔차(**Residual**) : $d_i = (y_i^{(true)} - y_i^{(pred)})$

잔차 제곱합 (**RSS**) :

$$\sum_{n=0}^{N-1} d_i^2$$

- 최적화(Optimization) : 손실함수를 어떻게 가장 작게 할 것인지에 대한 다양한 수학적 방법들을 의미한다.

(ex) 경사하강법, 최소제곱법

Overview

- 손실함수 복습

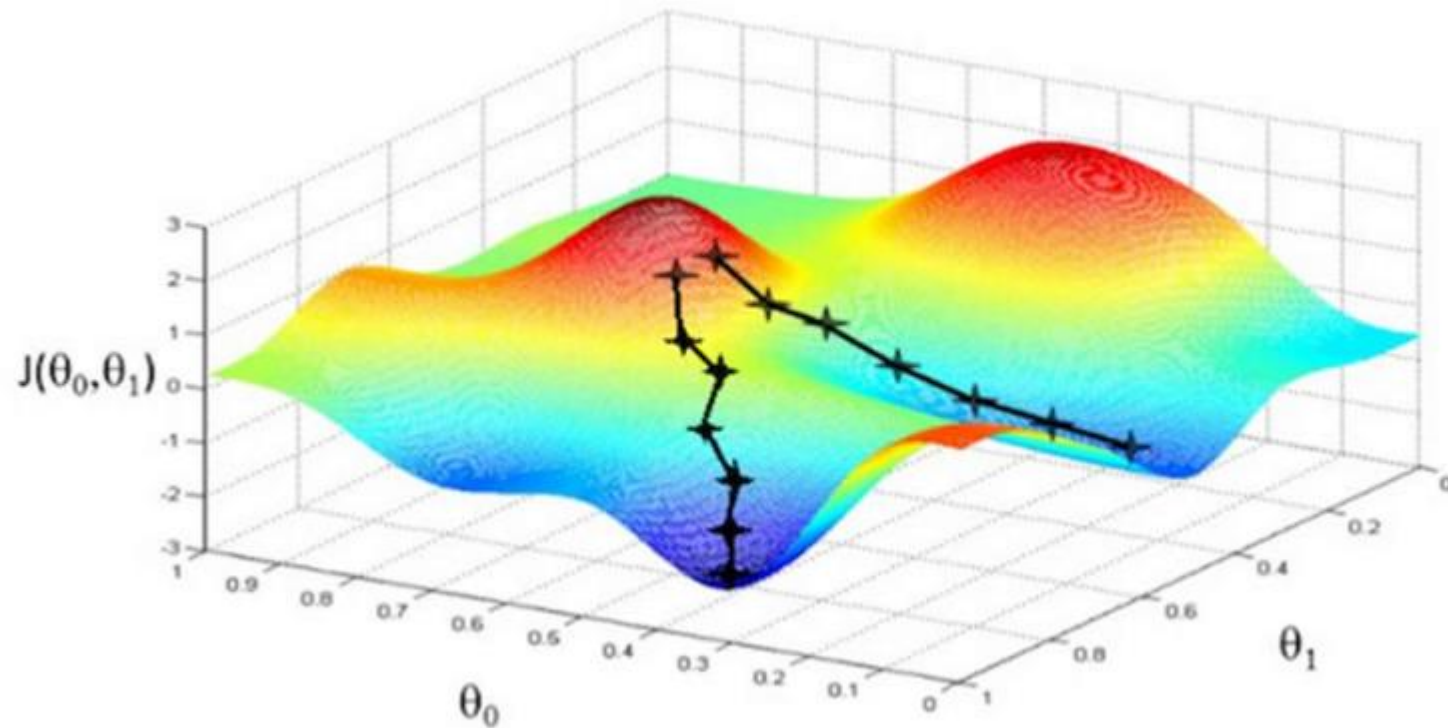
경사하강법

Hyper parameter

경사하강법 한계

최적화(Optimization)

- 딥러닝 분야에서 최적화란, 손실 함수(Loss function) 값을 최소화하는 모델 파라미터 값을 구하는 것



Overview

- 손실함수 복습

경사하강법

Hyper parameter

경사하강법 한계

일반적인 손실함수의 표기

- 모델 파라미터 : $w_1, w_2, w_3, \dots, w_M$
- Data points : $(x_n, y_n)_{n=0}^{N-1}$

$$\mathcal{L}(w_1, \dots, w_M; (x_n, y_n)_{n=0}^{N-1})$$

→ 특정 파라미터들을 사용한 모델에 대해, 그것이 예측한 값이 얼마나 잘못 되었는지를 수치로 표현한 것이 손실함수!

경사하강법

Overview

손실함수 복습

- 경사하강법

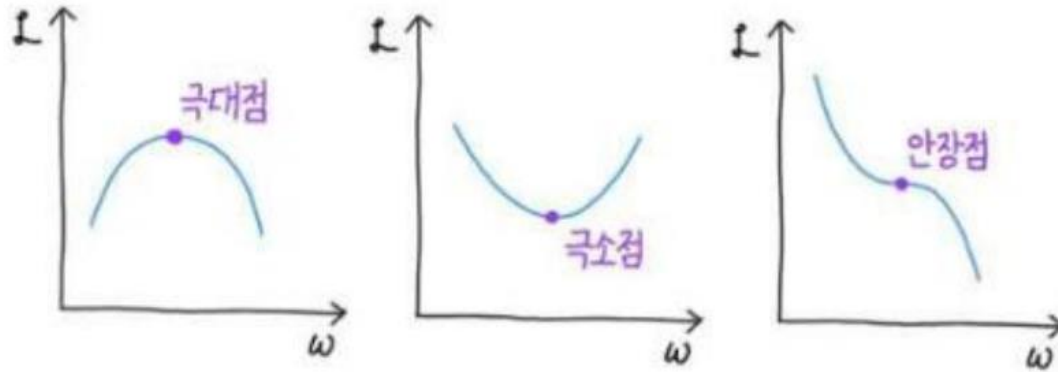
Hyper parameter

경사하강법 한계

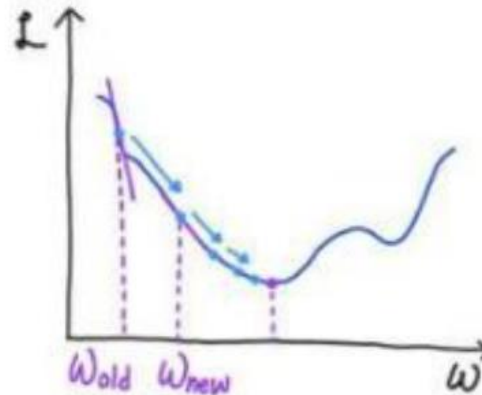
손실함수와 경사하강법

<손실함수가 w 에 대한 일변수함수임을 가정>

- 다양한 손실함수의 개형



- 경사하강법 : 손실함수를 따라 내려가면서 극소를 만족시키는 w 를 찾는 과정



Overview

손실함수 복습

• 경사하강법

Hyper parameter

경사하강법 한계

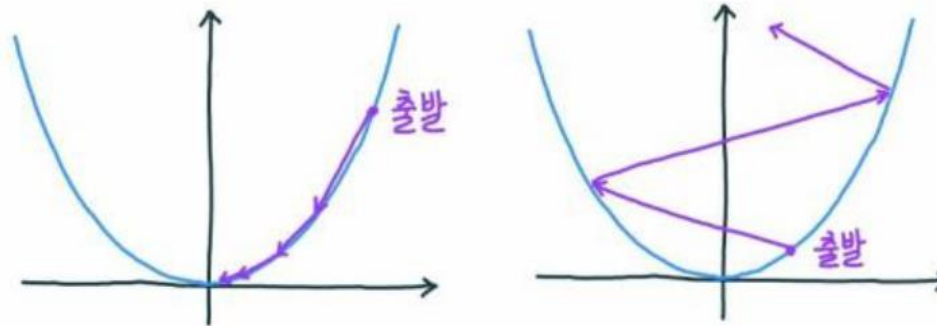
손실함수와 경사하강법

- $\mathcal{L}(w)$ 의 최소를 찾아야 함.

- $\frac{d\mathcal{L}}{dw} < 0 \rightarrow \text{increase } w, \quad \frac{d\mathcal{L}}{dw} > 0 \rightarrow \text{decrease } w$

$$\therefore w_{new} = w_{old} - \delta \cdot \left. \frac{d\mathcal{L}}{dw} \right|_{w=w_{old}}$$

- δ : 학습률(Learning rate) : 한 번의 **Step**에서 가중치 값을 변화시키는 양에 대한 척도



(적절한 학습률의 경우) / 학습률이 너무 큰 경우)

Overview

손실함수 복습

- 경사하강법

Hyper parameter

경사하강법 한계

손실함수와 경사하강법

Big Learning Rate



Just right



Too small



Overview

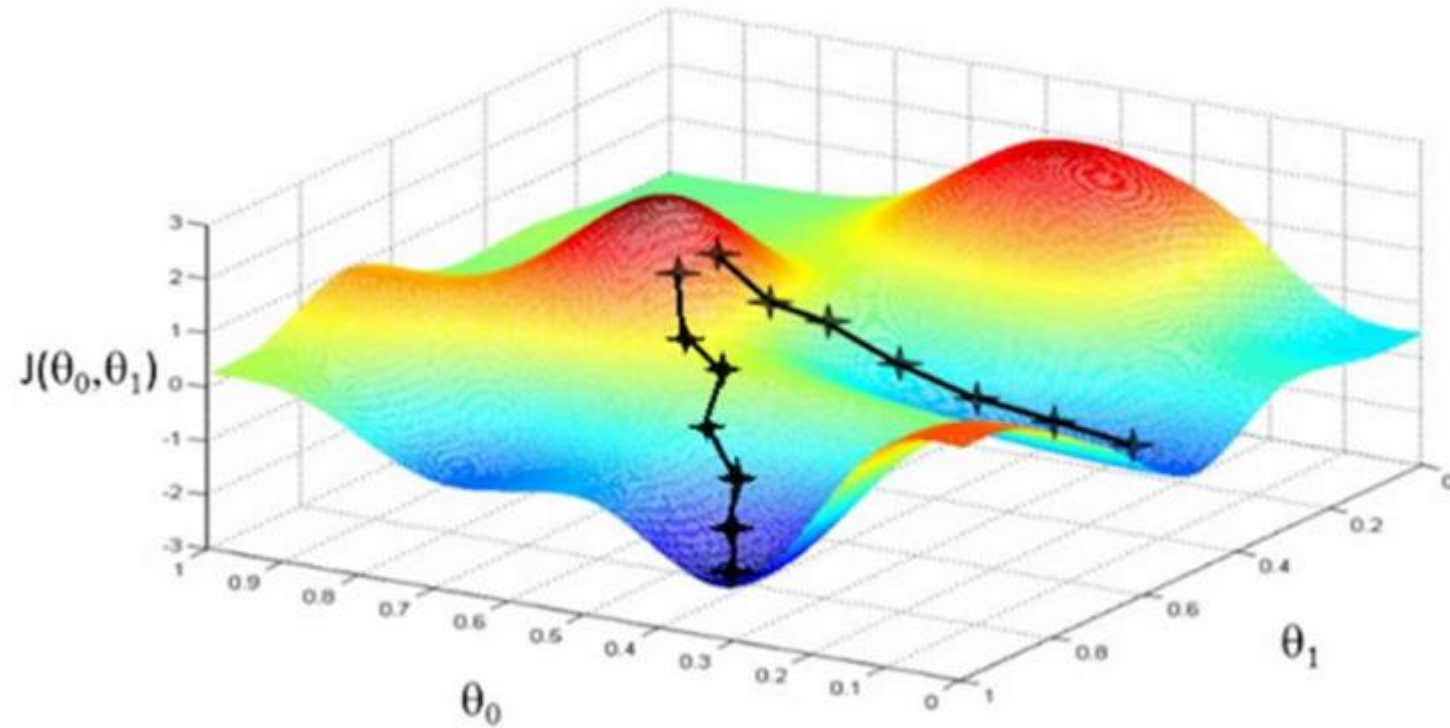
손실함수 복습

- **경사하강법**

Hyper parameter

경사하강법 한계

손실함수와 경사하강법



Overview

손실함수 복습

• 경사하강법

Hyper parameter

경사하강법 한계

왜 경사하강법을 쓰는가?

- 우리의 목표는 손실함수의 최솟값을 찾는 것!
- 손실함수의 최솟값을 찾기 위해서는 미분 계수가 0이 되는 지점을 찾아야 함 즉, $\mathcal{L}'(w) = 0$ 또는 $\nabla \mathcal{L}(\mathbf{w}) = 0$ 을 만족하는 모델 파라미터를 찾아야 함.
- 그러나 실제 상황, 특히 딥러닝에서의 손실함수는 굉장히 복잡하기에 $\mathcal{L}'(w) = 0$ 또는 $\nabla \mathcal{L}(\mathbf{w}) = 0$ 을 만족하는 w 또는 \mathbf{w} 를 찾는 것은 매우 어렵다.
- 따라서 우리는 손실함수를 점점 작아지게 하는 전략을 택하는 것!
- 경사하강법은 방대한 데이터에 대해서 적용하기에도 쉽고, 계산량도 훨씬 적다.

Overview

손실함수 복습

- **경사하강법**

Hyper parameter

경사하강법 한계

경사하강법 (Gradient Descent, GD)

- 안개가 가득 낀 산의 중턱에서, 산을 하산하는 것이 목표이다.
- 산의 내리막 방향으로 조금씩 내려가다보면 언젠가는 산을 하산할 수 있다.

Overview

손실함수 복습

- 경사하강법

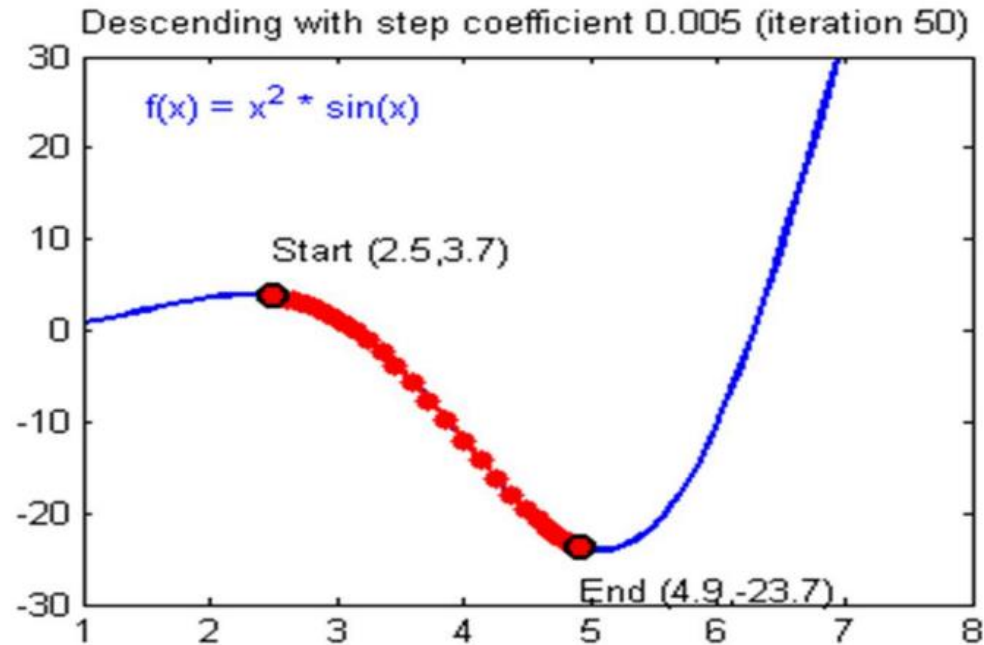
Hyper parameter

경사하강법 한계

경사하강법의 종료 조건

- 경사하강법의 종료 조건

$\frac{d\mathcal{L}}{dw} \big|_{w=w_{old}} \cong 0$ (Thresholding) 등 다양한 방법이 있음



Overview

손실함수 복습

- 경사하강법

Hyper parameter

경사하강법 한계

다변수함수에서의 경사하강법

- 모델 파라미터 : $w_1, w_2, w_3, \dots, w_M$

- $\mathbf{w}_{new} = \mathbf{w}_{old} - \delta \cdot \nabla \mathcal{L}(\mathbf{w})|_{\mathbf{w}=\mathbf{w}_{old}}$

$$\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{pmatrix}, \quad \nabla \mathcal{L}(\mathbf{w}) \Big|_{\mathbf{w}=\mathbf{w}_{old}} = \begin{pmatrix} \frac{\partial \mathcal{L}}{\partial w_1} \\ \frac{\partial \mathcal{L}}{\partial w_2} \\ \vdots \\ \frac{\partial \mathcal{L}}{\partial w_M} \end{pmatrix}_{old}$$



Hyper parameter

Overview

손실함수 복습

경사하강법

- Hyper parameter

경사하강법 한계

경사하강법의 Hyper parameter

- 경사하강법의 Hyperparameter
 - : 경사하강법 함수 내에서 사용되는 변수와 방법
 - $w_{old,0}$, δ , *Criteria for stopping step*, 반복횟수

Overview

손실함수 복습

경사하강법

● Hyper parameter

경사하강법 한계

경사하강법의 Hyper parameter

1. $\mathbf{w}_{old,0}$: Gaussian Random Distribution으로 Randomly choose
2. δ (학습률, Learning Rate)
 - 10의 거듭제곱들을 각각 학습률로 사용하여 경사하강법 진행
3. Criteria for stopping step; “큰 변화가 없으면 종료한다”
 - ① Gradient 값을 그래프에 나타낸 뒤 사용자의 판단하에 종료
 - ② Auto-stopping option : n번 반복 후 값의 변화가 없을 시 종료
 - ③ Thresholding : 특정 값 이하로 gradient가 감소하면 종료
4. 반복횟수

Overview

손실함수 복습

경사하강법

● Hyper parameter

경사하강법 한계

경사하강법의 Hyper parameter

적절한 Hyperparameter 의 설정방법

1. 대략적으로 하이퍼파라미터 값의 범위를 설정한다.
2. 설정된 범위에서 하이퍼파라미터의 값을 무작위로 추출한다.
3. 1단계에서 샘플링한 값을 사용하여 학습하고, 정확도를 평가한다.
4. 1,2 단계를 반복하며, 그 정확도를 보고 범위를 점점 좁힌다.

경사하강법 한계

Overview

손실함수 복습

경사하강법

Hyper parameter

- **경사하강법 한계**

경사하강법의 한계

(1) 데이터가 많아지면 계산량이 너무 많아진다.

(2) Local Minimum 문제.

(3) Plateau 문제

Overview

손실함수 복습

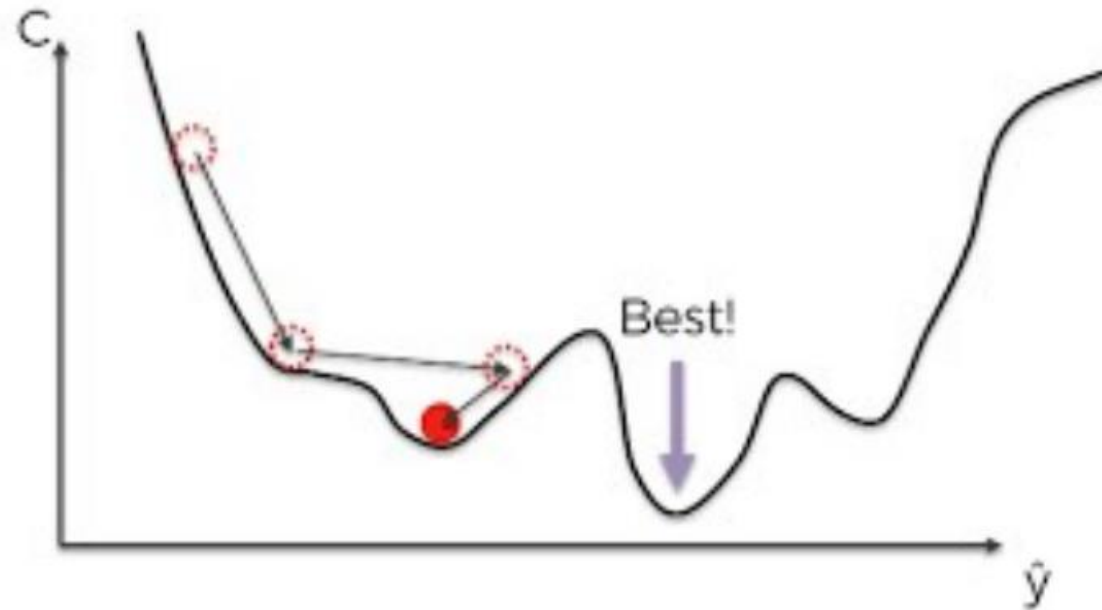
경사하강법

Hyper parameter

• 경사하강법 한계

경사하강법의 한계

- Local minimum
 - Global minimum 이 아닌 Local minimum에 빠지게 된다!



Overview

손실함수 복습

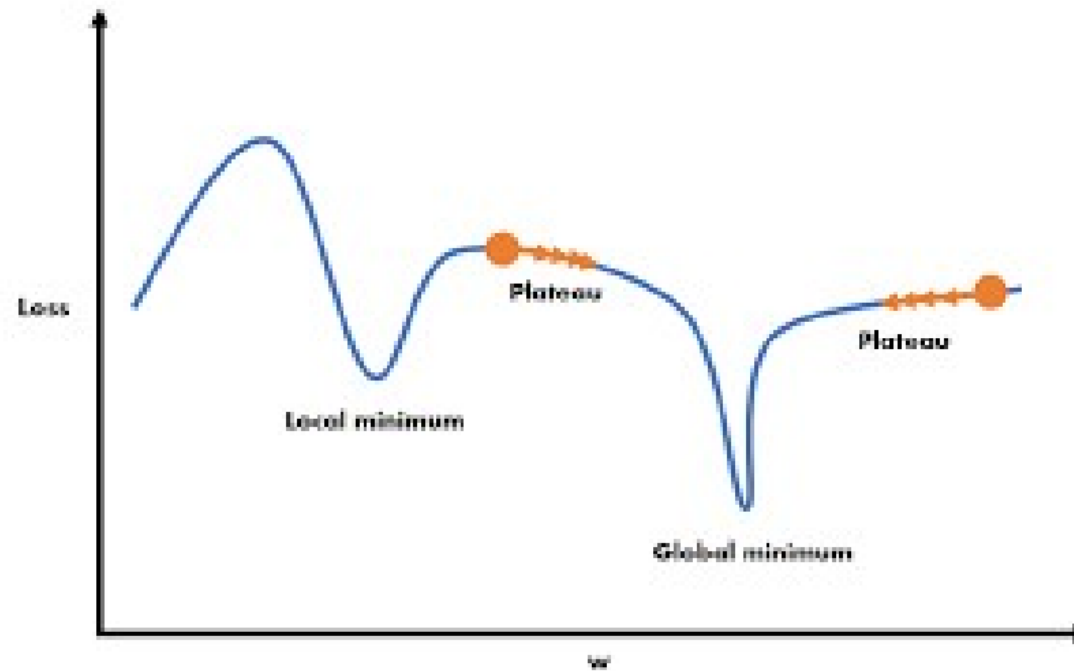
경사하강법

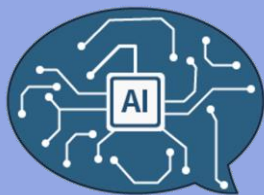
Hyper parameter

• 경사하강법 한계

경사하강법의 한계

- Plateau
 - 평탄한 지역에서는 학습 속도가 매우 느려지며, 느려지다 못해 정지해버릴 수도 있다.





감사합니다

