

The Mathematical Principles of Deep Learning

김세원

1. 골짜기로 내려가기 위해 경사를 따라 내려가는 방법

Gradient Descent와 Steepest Descent(Convex optimization)

Simply stepping in the direction of steepest descent is often not the best strategy for finding a minimum.

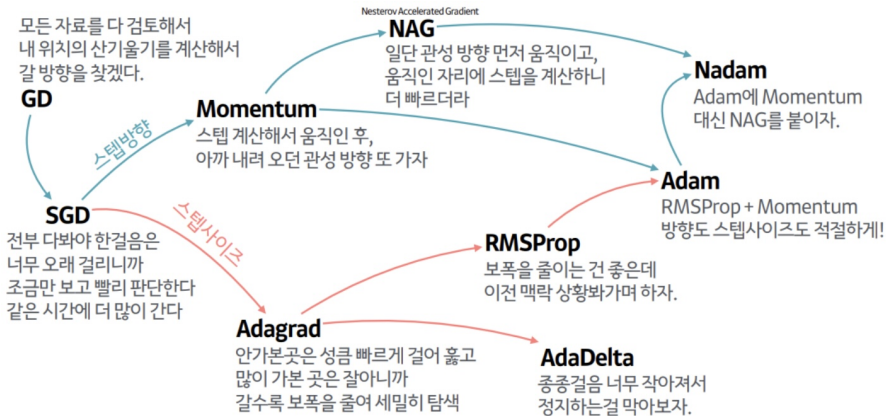


Figure: Optimizer for DL

2. 딥러닝의 최적화

Problem Awareness

- ▶ 데이터 축적 파이프라인 설계에 비해 과소평가된 수학



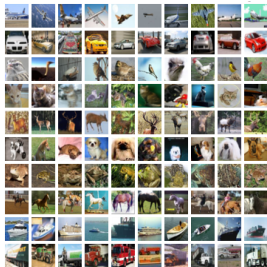
Objective

- ▶ 1. 각 optimizer의 수학(최적화)적인 증명과 배경 분석
- ▶ 2. Benchmark dataset에 대해 다양한 최적화 기법(optimizer) 구현(python)
- ▶ 3. 성능 비교 및 좋은 성능에 대한 수학적인 근거와 단서 탐색

3. Benchmark Data

알고리즘의 성능을 비교하고 평가하기 위해 사용되는 표준화된 데이터셋

Cifar-10



Imagenet



MNIST



10개의 클래스로
구분된 32×32 사물
사진을 모은 데이터셋
(학습용: 50,000개 /
테스트용: 10,000개)

Amazon Mechanical
Turk 서비스를 통해
사람이 일일이 분류한
대규모 데이터셋 (1000
만 개 이상)

0부터 9까지의 28×28
손글씨 사진을 모은
데이터셋 (학습용:
60,000개 / 테스트용:
10,000)

4. 참조

- Stephen Boyd. Convex Optimization.
- Mykel J. Kochenderfer. Algorithm for Optimization.
- Oilseok. Machine Learning.
- Aurelien Jeroen. Hands-on Machine Learning.
- Gilbert Strang. Linear Algebra and Learning from data.
- Diederik P. Kingma(2015). ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION.
- Pierre Foret(2021). SHARPNESS-AWARE MINIMIZATION FOR EFFICIENTLYIMPROVING GENERALIZATION.
- Jingzhao Zhang(2020). WHY GRADIENT CLIPPING ACCELERATES TRAINING:A THEORETICAL JUSTIFICATION FOR ADAPTIVITY.
- Taero Kim(2023). Sufficient Invariant Learning for Distribution Shift - Jonathon Shlens(2014). arXiv:1404.2000v1 [cs.IT] 8 Apr 2014Notes on Kullback-Leibler Divergence and Likelihood Theory.