ML Optimization Overview

Gradient Descent

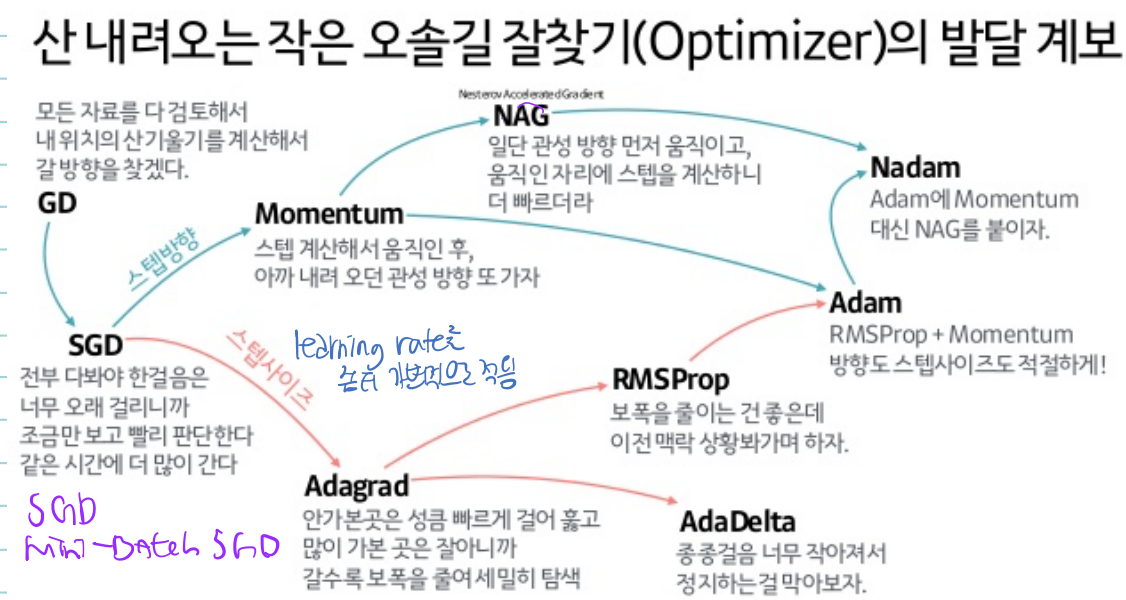
The whole story of optimization problem in ML starts with issues associated with Gradient Descent algorithm.
Gradient Descent is absolutely the most representative optimization tool we use in ML field.
Let's begin with two major weaknesses of Gradient Descent.

problem 1. Gradient Descent is a first-order algorithm.

- Does not use information about the curvature of the surface.
- Long valley issues

→ To deal with this, we adjust momentum or stepsizes.

- Conjugate gradient methods : take previous directions into account
- Second-order methods
  - Newton methods → Use the Hessian to provide information about the curvature.
  - Quasi-newton methods → use cheaper computational methods to approximate the Hessian.



problem 2. Handle non-differentiable functions.

- Subgradient methods can be used

Data size

- Batch gradient descent
- Stochastic gradient descent

Mathematical approaches

- Duality
- Convex optimization