Optimization Techniques Report

Mahanth Yalla
January 29, 2025

Abstract

Experimented on different optimization Techniques and findings are documented in this document.

---

1 GRADIENT DESCENT

Gradient descent is a way to minimize an objective function $J(\theta)$ parameterized by a model's parameters $\theta \in \mathbb{R}^d$ by updating the parameters in the opposite direction of the gradient of the objective function $\nabla_\theta J(\theta)$ with respect to the parameters. The learning rate η determines the size of the steps we take to reach a (local) minimum. In other words, we follow the direction of the slope of the surface created by the objective function downhill until we reach a valley. [1]

2 VARIATIONS OF GRADIENT DESCENT

- Stochastic (SGD)
- Mini-batch SGD
- Momentum
- Nesterov Accelerated (NAG)
- Adagrad
- Adadelta
- RMSprop
- Adam
- AdaMax
- Nadam
- AMSGrad
- AdamW
- Yogi
- RAdam
- Lookahead

REFERENCES

[1] Sebastian Ruder. "An overview of gradient descent optimization algorithms". In: *CoRR* abs/1609.04747 (2016). arXiv: 1609.04747. URL: http://arxiv.org/abs/1609.04747.