

Hypergradient Descentにおける 学習率学習率の調整

研究背景・目的

- 勾配降下法において、学習率の調整は重要
- **Hypergradient Descent**を用いると、学習率も勾配を使って調整できる
- 適切な学習率 α の調整が可能となる「学習率の学習率 β 」を求めたい

研究内容

目的関数 $f(\theta)$ が凸かつ L -平滑な単純なモデルの場合、 $\beta = \frac{1}{L\|\nabla_{\theta}f(\theta_{t-2})\|^2}$ とすれば、適切な α の調整ができることを実験により確認

要素技術

- **Hypergradient Descent** (Baydin et al. 2017)

$$\begin{aligned}\alpha_t &= \alpha_{t-1} - \beta \cdot \nabla_{\alpha} f(\theta_{t-1}) \\ \theta_t &= \theta_{t-1} - \alpha_t \cdot \nabla_{\theta} f(\theta_{t-1})\end{aligned}$$

- 凸関数

任意の \mathbf{u}, \mathbf{v} と任意の $\alpha \in [0, 1]$ に対して

$$f(\alpha\mathbf{u} + (1 - \alpha)\mathbf{v}) \leq \alpha f(\mathbf{u}) + (1 - \alpha)f(\mathbf{v})$$

- L -平滑な関数

任意の \mathbf{u}, \mathbf{v} に対して

$$\|\nabla f(\mathbf{u}) - \nabla f(\mathbf{v})\| \leq L\|\mathbf{u} - \mathbf{v}\|$$