Towards Robust Learning to Optimize with Theoretical Guarantee

Qingyu Song†, Wei Lin†, Juncheng Wang‡, Hong Xu†
†The Chinese University of Hong Kong, ‡Hong Kong Baptist University
What is learning to optimize (L2O)?

• Optimization Problem

\[ \min_{x} f(x), \quad x \in \mathbb{R}^n, f: \mathbb{R}^n \rightarrow \mathbb{R} \]

• Benefits
  • Better optimality (potential).
  • Better convergence/efficiency [1].

How does L2O work?

- Workflow of L2O (Inference)

Given initial point $x_i^0$

$x_i^0 \in \mathcal{N}(0,1), f(x_i^0) \rightarrow x_i^*$

A distribution
A class of objective
L2O’s Failure in Out-of-Distribution (OOD) Scenarios

1. $x$ OOD
   $x^0 \in N(10,1), f(x)$

   In-Distribution (InD)
   $x^0 \in N(0,1), f(x)$

2. $f$ OOD
   $x^0 \in N(0,1), f = f(x - 10)$

   $x^* \in D'$
   $x^* \neq x^*$
   $x^* \in D$
   $x^* \neq x^* + 10$
   $x^* \in D''$
Theoretical Convergence Analysis

• Convergence of Single-Iteration (Smooth Case)

\[
\begin{align*}
F'(x_k + s_k) - F'(x_{k-1} + s_{k-1}) &
\leq - \frac{\| \nabla f'(x_{k-1} + s_{k-1}) \|^2}{2L} \\
&+ L \| \text{diag}(J_{1,k-1}s') \nabla f'(x_{k-1} + s_{k-1}) \|^2 \\
&+ L \| \frac{\nabla f'(x_{k-1} + s_{k-1}) - \nabla f(x_{k-1})}{2L} - J_{2,k-1}s' \|^2.
\end{align*}
\]

Convergence of Gradient-Descent

Deterioration w.r.t. OOD
Theoretical Convergence Rate Analysis

• Convergence Rate (Smooth Case)

$$\min_{k=1,\ldots,K} F'(x_k + s_k) - F'(x^* + s^*) \leq \frac{L}{2} \| x_0 - x^* + s_0 - s^* \|^2 - \frac{L}{2} \| x_K - x^* + s_K - s^* \|^2$$

Convergence Rate of Gradient-Descent

Deterioration w.r.t. OOD
Convergence Improvement

• Upper Bound Relaxation

\[
\begin{align*}
F'(x_k + s_k) - F'(x_{k-1} + s_{k-1}) & \leq - \frac{\|\nabla f'(x_{k-1} + s_{k-1})\|^2}{2L} \\
& + \frac{\|\nabla f'(x_{k-1} + s_{k-1}) - \nabla f(x)\|^2}{2L} \\
& + (LC_1^2 n \|\nabla f'(x_{k-1} + s_{k-1})\|^2 + 2LC_2^2 n) \|s\|^2.
\end{align*}
\]

• Improve upper bound: Magnitude reduction.
  • Our approach: Input Feature Simplification.

OOD vector, NN’s input feature
A New L2O Model with Gradient-Only Input

• New Model Formulation Based on [1]

\[ x_k = x_{k-1} - R_k \nabla f(x_{k-1}) - R_k g_k - Q_k v_{k-1} - b_{1,k}, \]
\[ v_k = (I - B_k)G_k + B_k G_{k-1} - b_{2,k}, \]
\[ G_k := R_k^{-1}(x_{k-1} - x_k - Q_k v_{k-1} - b_{1,k}), \]

• Learn \( R, Q, B \). Details at [1].

Empirical Outperformance
Empirical Outperformance

Figure 14. Logistic Regression: Real-World Ionosphere Dataset.

Figure 15. Logistic Regression: Real-World Spambase Dataset.

Figure 17. Logistic Regression: OOD by Trigger 2.
Thank You!