Second-Order Optimization: L-BFGS

Limited-Memory Broyden-Fletcher-Goldfarb-Shanno

🔍 Core Concept

Optimization using second-order information (Hessian) with efficient inverse Hessian approximation

⚡ Performance Trade-off

Iterations	Per-iteration Cost
Fewer ✓	Higher ✗

⚠ Limitations

- Unsuitable for large-scale deep learning due to memory constraints
- Not suitable for mini-batch optimization

First-Order vs Second-Order

First-Order

- Uses gradient only
- More iterations
- Lower cost per iteration

Second-Order

- Uses Hessian
- Fewer iterations
- Higher cost per iteration

💡 Suitable Applications

Small Models: Optimization of small-scale models

Scientific Computing: Scientific computing applications

Traditional ML: Logistic Regression, etc.

⚖ Characteristics

- ✓ Fewer iterations than first-order methods
- ✓ Deterministic method
- ✗ High computational cost per iteration ($O(n^2)$)

❌ Requires full batch evaluation

Specific Cases: When full batch evaluation is feasible