# Newton's method. Cubic Regularized Newton Methods for Logistic Regression

Dmitrii Topchii

December 17, 2024

# Motivation

- Logistic regression: widely used model for binary classification with convex loss.
- First-order methods (e.g., Gradient Descent) are simple but can be slow.
- Newton's method uses second-order info for faster local convergence, but can be costly.
- Cubic regularization: adds a third-order term to improve global convergence rates and potentially faster overall progress.

# Problem Statement

Minimize the logistic regression objective:

$$f(w) = \frac{1}{n} \sum_{i=1}^{n} \log\bigl(1 + \exp(-y_i x_i^T w)\bigr) + \frac{\lambda}{2} \|w\|^2.$$

- $x_i \in \mathbb{R}^d$, $y_i \in \{0, 1\}$, $\lambda > 0$.
- $f(w)$ is convex in $w$.

# Methods Overview

Examined methods:

- **GD**: $w_{k+1} = w_k - \eta \nabla f(w_k)$
- **Nesterov**: Accelerated gradient with momentum.
- **Adam**: Adaptive first-order method.
- **Newton**: $w_{k+1} = w_k - [\nabla^2 f(w_k)]^{-1} \nabla f(w_k)$
- **L-BFGS**: Quasi-Newton approximation of Hessian inverse.
- **Cubic Reg. Newton (CRN)**: Solve

$$m_k(h) = f(w_k) + \nabla f(w_k)^T h + \frac{1}{2} h^T \nabla^2 f(w_k) h + \frac{L}{6} \|h\|^3.$$

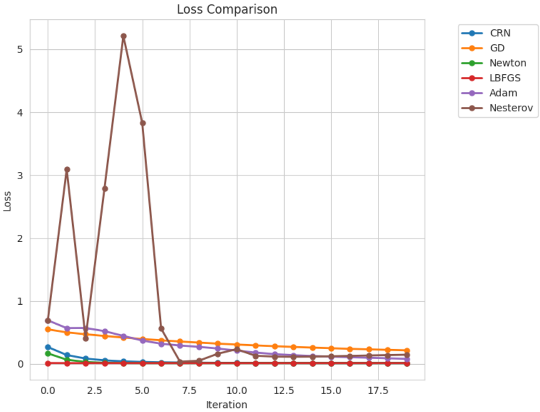
# Cubic Regularized Newton

- Incorporates a third-order term to ensure global convergence.
- Step $h_k$ chosen to approximately minimize $m_k(h)$.
- Potentially better theoretical complexity.

# Experimental Setup

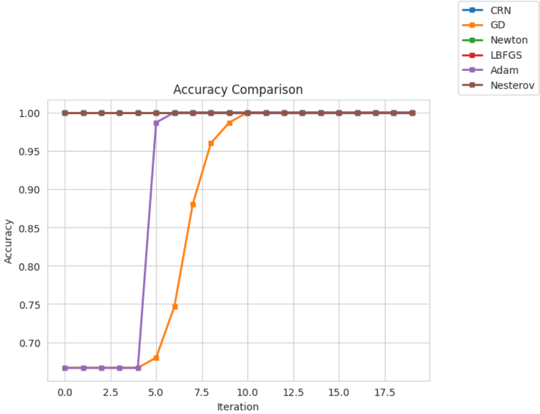
- Datasets: Iris (binary classification).
- Metrics: Loss $f(w)$, accuracy, runtime per iteration.
- Sensitivity analysis for CRN parameter $L$.

# Convergence Results: Loss



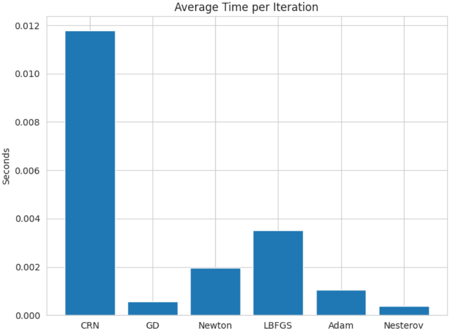

- CRN and Newton achieve very low loss.
- First-order methods require more steps or show initial fluctuations.

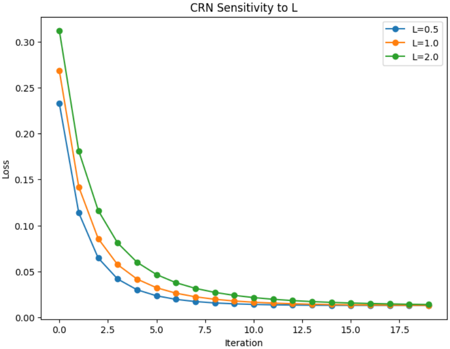
# Accuracy Comparison




- Most methods eventually reach perfect accuracy.
- Differences mainly in the speed of convergence.

# Runtime and Sensitivity




- CRN is more expensive per iteration.
- Sensitivity to $L$: smaller $L$ can yield smoother convergence.

# Summary Table

| Method | Final Loss | Acc. | Time/Iter | Total Time |
|---|---|---|---|---|
| CRN | 0.0130 | 1.0 | 0.0118 | 5.4052 |
| GD | 0.2162 | 1.0 | 0.0006 | 0.0162 |
| Newton | 0.0129 | 1.0 | 0.0019 | 0.0426 |
| L-BFGS | 0.0129 | 1.0 | 0.0035 | 0.0751 |
| Adam | 0.0814 | 1.0 | 0.0010 | 0.0226 |
| Nesterov | 0.1464 | 1.0 | 0.0004 | 0.0086 |

# Conclusion

- Cubic Regularized Newton achieves low loss and perfect accuracy with fewer iterations.
- Higher per-iteration cost compared to simpler methods.
- Good for small-scale problems and where high precision is needed.
- Trade-off: choose method based on problem size and resource constraints.