Optimization in Machine Learning

Univariate optimization Brent's method

× 0 0 × 0 × ×



Learning goals

- Quadratic interpolation
- Brent's procedure

Similar to golden ratio procedure but select x^{new} differently: x^{new} as minimum of a parabola fitted through

 $(x^{\text{left}}, f^{\text{left}}), (x^{\text{best}}, f^{\text{best}}), (x^{\text{right}}, f^{\text{right}}).$



× × ×

Similar to golden ratio procedure but select x^{new} differently: x^{new} as minimum of a parabola fitted through

 $(x^{\text{left}}, f^{\text{left}}), (x^{\text{best}}, f^{\text{best}}), (x^{\text{right}}, f^{\text{right}}).$



× × 0 × ×

Similar to golden ratio procedure but select x^{new} differently: x^{new} as minimum of a parabola fitted through

 $(x^{\text{left}}, f^{\text{left}}), (x^{\text{best}}, f^{\text{best}}), (x^{\text{right}}, f^{\text{right}}).$



× ° × ×

Similar to golden ratio procedure but select x^{new} differently: x^{new} as minimum of a parabola fitted through

 $(x^{\text{left}}, f^{\text{left}}), (x^{\text{best}}, f^{\text{best}}), (x^{\text{right}}, f^{\text{right}}).$



× × 0 × ×

Similar to golden ratio procedure but select x^{new} differently: x^{new} as minimum of a parabola fitted through

 $(x^{\text{left}}, f^{\text{left}}), (x^{\text{best}}, f^{\text{best}}), (x^{\text{right}}, f^{\text{right}}).$



× 0 × ×

QUADRATIC INTERPOLATION COMMENTS

- Quadratic interpolation **not robust**. The following may happen:
 - Algorithm suggests the same x^{new} in each step,
 - x^{new} outside of search interval,
 - Parabola degenerates to line and no real minimum exists
- Algorithm must then abort, finding a global minimum is not guaranteed.

× × ×

BRENT'S METHOD

- Brent proposed an algorithm (1973) that alternates between golden ratio search and quadratic interpolation as follows:
 - Quadratic interpolation step acceptable if: (i) x^{new} falls within $[x^{\text{left}}, x^{\text{right}}]$ (ii) x^{new} sufficiently far away from x^{best} (Heuristic: Less than half of movement of step before last)
 - Otherwise: Proposal via golden ratio
- Benefit: Fast convergence (quadratic interpolation), unstable steps (e.g. parabola degenerated) stabilized by golden ratio search
- Convergence guaranteed if the function f has a local minimum
- Used in R-function optimize()

× 0 0 × 0 × × ×

EXAMPLE: MLE POISSON

• Poisson density:
$$f(k \mid \lambda) := \mathbb{P}(x = k) = \frac{\lambda^{k} \cdot \exp(-\lambda)}{k!}$$

• Negative log-likelihood for *n* observations:



× × ×

GR and Brent converge to minimum at $x^* \approx 1$. But: GR needs ≈ 45 it., Brent only needs ≈ 15 it. for same tolerance.