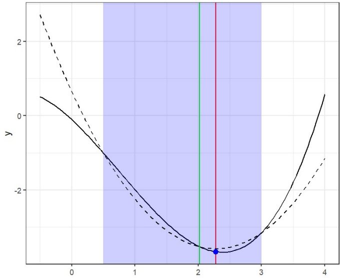
Optimization in Machine Learning

Univariate optimization
Brent’s method



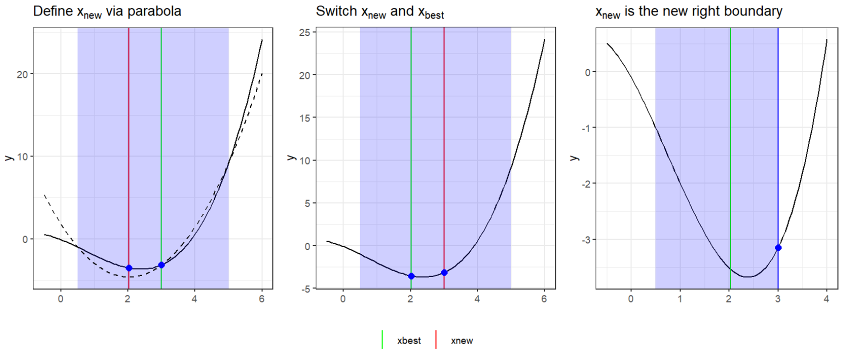
Learning goals

- Quadratic interpolation
- Brent’s procedure

QUADRATIC INTERPOLATION

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

$$(x_{\text{left}}, f_{\text{left}}), (x_{\text{best}}, f_{\text{best}}), (x_{\text{right}}, f_{\text{right}})$$

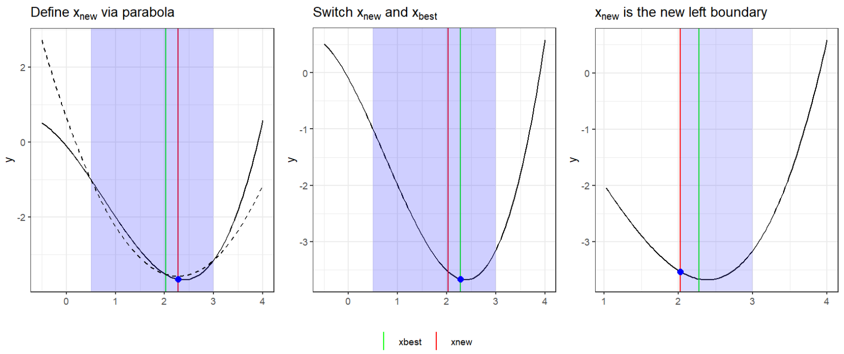


Left: Fit parabola (dashed) and propose minimum (red) as new point. Middle: Switch / not switch with x_{best}. Right: New interval.

QUADRATIC INTERPOLATION

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

$$(x_{\text{left}}, f_{\text{left}}), (x_{\text{best}}, f_{\text{best}}), (x_{\text{right}}, f_{\text{right}})$$

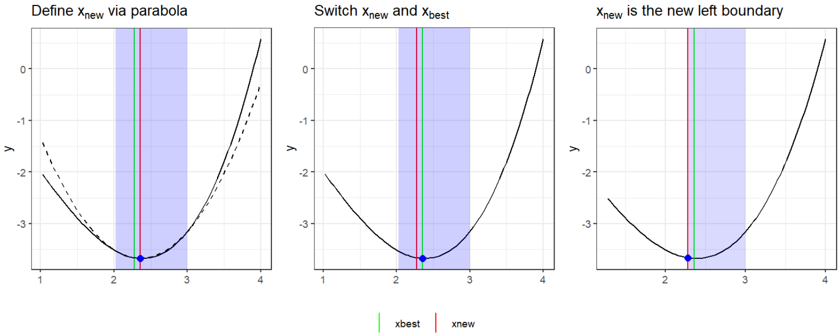


Left: Fit parabola (dashed) and propose minimum (red) as new point. Middle: Switch / not switch with x_{best}. Right: New interval.

QUADRATIC INTERPOLATION

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

$$(x_{\text{left}}, f_{\text{left}}), (x_{\text{best}}, f_{\text{best}}), (x_{\text{right}}, f_{\text{right}})$$

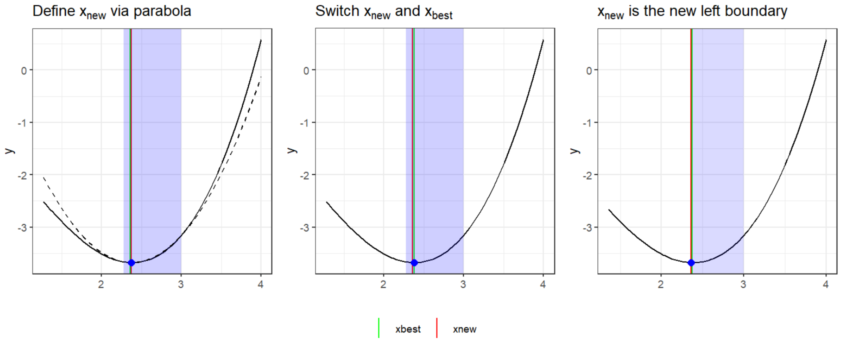


Left: Fit parabola (dashed) and propose minimum (red) as new point. Middle: Switch / not switch with x_{best}. Right: New interval.

QUADRATIC INTERPOLATION

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

$$(x_{\text{left}}, f_{\text{left}}), (x_{\text{best}}, f_{\text{best}}), (x_{\text{right}}, f_{\text{right}})$$

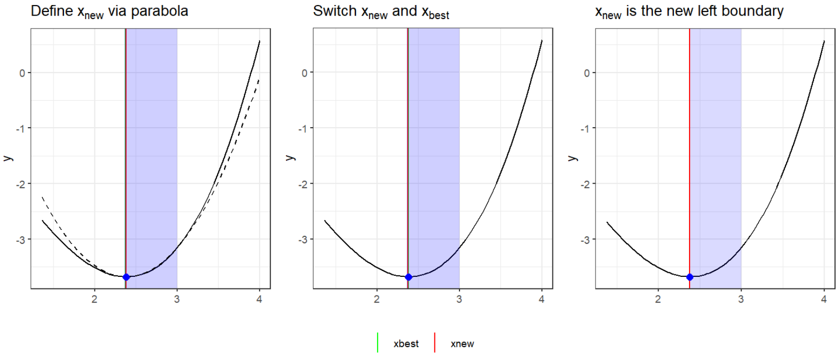


Left: Fit parabola (dashed) and propose minimum (red) as new point. Middle: Switch / not switch with x_{best}. Right: New interval.

QUADRATIC INTERPOLATION

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

$$(x_{\text{left}}, f_{\text{left}}), (x_{\text{best}}, f_{\text{best}}), (x_{\text{right}}, f_{\text{right}})$$



Left: Fit parabola (dashed) and propose minimum (red) as new point.
Middle: Switch / not switch with x_{best}. Right: New interval.

QUADRATIC INTERPOLATION COMMENTS

- Quadratic interpolation **not robust**. The following may happen:
  - Algorithm suggests 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

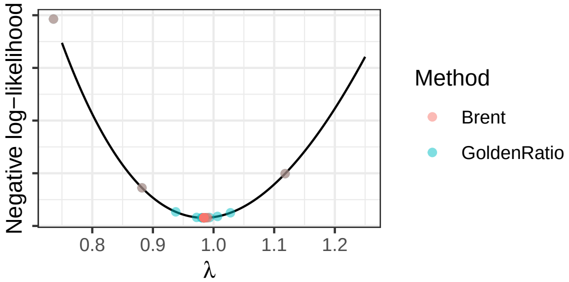
BRENT'S METHOD

- Brent's algorithm ▸ Brent 1973 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 the movement of step before last)
  - Otherwise: proposal via golden ratio
- Benefit: fast convergence (quadratic interpolation), unstable steps stabilized by golden ratio search
- Convergence guaranteed if function f has local min
- Used in R-function `optimize()`

EXAMPLE: MLE POISSON

- Poisson density: $f(k|\lambda) := \mathbb{P}(x = k) = \frac{\lambda^k \cdot \exp(-\lambda)}{k!}$
- Negative log-likelihood for n samples:

$$-\ell(\lambda, \mathcal{D}) = -\log \prod_{i=1}^{n} f(x^{(i)}|\lambda) = -\sum_{i=1}^{n} \log f(x^{(i)}|\lambda)$$



GR and Brent converge to global min at $x^* \approx 1$

But GR needs $\approx$ 45 iters, Brent only $\approx$ 15 for same tolerance