Parameter Estimation for Gaussian Processes

Yiannis Andrianakis and Peter Challenor

July 8, 2009

Outline

- Parameter estimation
  - ML
  - REML
  - Toolkit approach (ThreadCoreGP)
- Design points
  - Input dimension
  - Correlation length
- Method Comparison
- Future work

Outline

- Parameter estimation
  - ML
  - REML
  - Toolkit approach (ThreadCoreGP)
- Design points
  - Input dimension
  - Correlation length
- Method Comparison
- Future work

Outline

- Parameter estimation
  - ML
  - REML
  - Toolkit approach (ThreadCoreGP)
- Design points
  - Input dimension
  - Correlation length
- Method Comparison
- Future work

Outline

- Parameter estimation
  - ML
  - REML
  - Toolkit approach (ThreadCoreGP)
- Design points
  - Input dimension
  - Correlation length
- Method Comparison
- Future work

Outline

- Parameter estimation
  - ML
  - REML
  - Toolkit approach (ThreadCoreGP)
- Design points
  - Input dimension
  - Correlation length
- Method Comparison
- Future work

Outline

- Parameter estimation
  - ML
  - REML
  - Toolkit approach (ThreadCoreGP)
- Design points
  - Input dimension
  - Correlation length
- Method Comparison
- Future work

Outline

- Parameter estimation
  - ML
  - REML
  - Toolkit approach (ThreadCoreGP)
- Design points
  - Input dimension
  - Correlation length
- Method Comparison
- Future work

Outline

- Parameter estimation
  - ML
  - REML
  - Toolkit approach (ThreadCoreGP)
- Design points
  - Input dimension
  - Correlation length
- Method Comparison
- Future work

Outline

- Parameter estimation
  - ML
  - REML
  - Toolkit approach (ThreadCoreGP)
- Design points
  - Input dimension
  - Correlation length
- Method Comparison
- Future work

Outline

- Parameter estimation
  - ML
  - REML
  - Toolkit approach (ThreadCoreGP)
- Design points
  - Input dimension
  - Correlation length
- Method Comparison
- Future work

Problem formulation

- Gaussian Process model of the data

$$y|\beta, \sigma^2, \delta \sim \mathcal{N}(H\beta, \sigma^2 A)$$

β : Regression Coefficients
σ^2: Scale Factor
δ : Correlation lengths
$H = [h(\mathbf{x}_1), h(\mathbf{x}_2), \cdots, h(\mathbf{x}_n)]^{\mathrm{T}}$

$$[A]_{(k,l)} = c(\mathbf{x}_k, \mathbf{x}_l)$$

$$c(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^{p} \exp\left\{ -\frac{(x_i - x_i')^2}{\delta_i^2} \right\}$$

- We want to estimate $\theta = [\beta, \sigma^2, \delta]$ using $[y_1, \cdots, y_n]$ and $[\mathbf{x}_1, \cdots, \mathbf{x}_n]$

Problem formulation

- Gaussian Process model of the data

$$y|\beta, \sigma^2, \delta \sim \mathcal{N}(H\beta, \sigma^2 A)$$

β : Regression Coefficients
σ^2: Scale Factor
δ : Correlation lengths
$H = [h(\mathbf{x}_1), h(\mathbf{x}_2), \cdots, h(\mathbf{x}_n)]^{\mathrm{T}}$

$$[A]_{(k,l)} = c(\mathbf{x}_k, \mathbf{x}_l)$$

$$c(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^{p} \exp\left\{ -\frac{(x_i - x_i')^2}{\delta_i^2} \right\}$$

- We want to estimate $\theta = [\beta, \sigma^2, \delta]$ using $[y_1, \cdots, y_n]$ and $[\mathbf{x}_1, \cdots, \mathbf{x}_n]$

Maximum Likelihood

- Likelihood

$$p(y|\beta, \sigma^2, \delta) = \frac{|A|^{-1/2}}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(y - H\beta)^{\mathrm{T}} A^{-1}(y - H\beta)\right\}$$

- Parameters

$$\hat{\beta}_{ML} = (H^{\mathrm{T}} A^{-1} H)^{-1} H^{\mathrm{T}} A^{-1} y \qquad \hat{\sigma}^2_{ML} = \frac{(y - H\hat{\beta}_{ML})^{\mathrm{T}} A^{-1}(y - H\hat{\beta}_{ML})}{n}$$

$$\hat{\delta}_{ML} = \arg\max_{\delta} \left[p(y|\beta_{ML}, \sigma^2_{ML}, \delta)\right]$$

Maximum Likelihood

- Likelihood

$$p(y|\beta,\sigma^2,\delta) = \frac{|A|^{-1/2}}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(y - H\beta)^{\mathrm{T}} A^{-1}(y - H\beta)\right\}$$

- Parameters

$$\hat{\beta}_{ML} = (H^{\mathrm{T}} A^{-1} H)^{-1} H^{\mathrm{T}} A^{-1} y \qquad \hat{\sigma}^2_{ML} = \frac{(y-H\hat{\beta}_{ML})^{\mathrm{T}} A^{-1}(y-H\hat{\beta}_{ML})}{n}$$

$$\hat{\delta}_{ML} = \arg\max_{\delta} \left[p(y|\beta_{ML},\sigma^2_{ML},\delta)\right]$$

Maximum Likelihood

- Likelihood

$$p(y|\beta,\sigma^2,\delta) = \frac{|A|^{-1/2}}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(y - H\beta)^{\mathrm{T}} A^{-1}(y - H\beta)\right\}$$

- Parameters

$$\hat{\beta}_{ML} = (H^{\mathrm{T}} A^{-1} H)^{-1} H^{\mathrm{T}} A^{-1} y \qquad \hat{\sigma}^2_{ML} = \frac{(y - H\hat{\beta}_{ML})^{\mathrm{T}} A^{-1} (y - H\hat{\beta}_{ML})}{n}$$

$$\hat{\delta}_{ML} = \arg\max_{\delta} \left[p(y|\beta_{ML}, \sigma^2_{ML}, \delta)\right]$$

- Likelihood

$$p(y|\sigma^2,\delta) = \int p(y|\sigma^2,\delta,\beta)\mathrm{d}\beta$$

$$\propto \frac{|A|^{-1/2}|H'A^{-1}H|^{-1/2}}{(2\pi\sigma^2)^{\frac{n-q}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(y-H\hat{\beta})'A^{-1}(y-H\hat{\beta})\right\}$$

- Parameters

$$\hat{\beta} = (H^{\mathrm{T}}A^{-1}H)^{-1}H^{\mathrm{T}}A^{-1}y \qquad \hat{\sigma}^2_{RL} = \frac{(y-H\hat{\beta})^{\mathrm{T}}A^{-1}(y-H\hat{\beta})}{n-q}$$

$$\hat{\delta}_{RL} = \arg\max_{\delta}\left[p(y|\sigma^2_{RL},\delta)\right]$$

- Likelihood

$$p(y|\sigma^2,\delta) = \int p(y|\sigma^2,\delta,\beta)\mathrm{d}\beta$$

$$\propto \frac{|A|^{-1/2}|H'A^{-1}H|^{-1/2}}{(2\pi\sigma^2)^{\frac{n-q}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(y-H\hat{\beta})'A^{-1}(y-H\hat{\beta})\right\}$$

- Parameters

$$\hat{\beta} = (H^{\mathrm{T}}A^{-1}H)^{-1}H^{\mathrm{T}}A^{-1}y \qquad \hat{\sigma}^2_{RL} = \frac{(y-H\hat{\beta})^{\mathrm{T}}A^{-1}(y-H\hat{\beta})}{n-q}$$

$$\hat{\delta}_{RL} = \arg\max_{\delta}\left[p(y|\sigma^2_{RL},\delta)\right]$$

- Likelihood

$$p(y|\sigma^2,\delta) = \int p(y|\sigma^2,\delta,\beta)\mathrm{d}\beta$$

$$\propto \frac{|A|^{-1/2}|H'A^{-1}H|^{-1/2}}{(2\pi\sigma^2)^{\frac{n-q}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(y-H\hat{\beta})'A^{-1}(y-H\hat{\beta})\right\}$$

- Parameters

$$\hat{\beta} = (H^{\mathrm{T}}A^{-1}H)^{-1}H^{\mathrm{T}}A^{-1}y \qquad \hat{\sigma}^2_{RL} = \frac{(y-H\hat{\beta})^{\mathrm{T}}A^{-1}(y-H\hat{\beta})}{n-q}$$

$$\hat{\delta}_{RL} = \arg\max_{\delta}\left[p(y|\sigma^2_{RL},\delta)\right]$$

- Likelihood

$$p(y|\sigma^2,\delta) = \int p(y|\sigma^2,\delta,\beta)\mathrm{d}\beta$$

$$\propto \frac{|A|^{-1/2}|H'A^{-1}H|^{-1/2}}{(2\pi\sigma^2)^{\frac{n-q}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(y - H\hat{\beta})'A^{-1}(y - H\hat{\beta})\right\}$$

- Parameters

$$\hat{\beta} = (H^{\mathrm{T}}A^{-1}H)^{-1}H^{\mathrm{T}}A^{-1}y \qquad \hat{\sigma}^2_{RL} = \frac{(y-H\hat{\beta})^{\mathrm{T}}A^{-1}(y-H\hat{\beta})}{n-q}$$

$$\hat{\delta}_{RL} = \arg\max_{\delta}\left[p(y|\sigma^2_{RL},\delta)\right]$$

- Likelihood

$$p(y|\sigma^2,\delta) = \int p(y|\sigma^2,\delta,\beta)\mathrm{d}\beta$$

$$\propto \frac{|A|^{-1/2}|H'A^{-1}H|^{-1/2}}{(2\pi\sigma^2)^{\frac{n-q}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(y-H\hat{\beta})'A^{-1}(y-H\hat{\beta})\right\}$$

- Parameters

$$\hat{\beta} = (H^{\mathrm{T}}A^{-1}H)^{-1}H^{\mathrm{T}}A^{-1}y \qquad \hat{\sigma}^2_{RL} = \frac{(y-H\hat{\beta})^{\mathrm{T}}A^{-1}(y-H\hat{\beta})}{n-q}$$

$$\hat{\delta}_{RL} = \arg\max_{\delta}\left[p(y|\sigma^2_{RL},\delta)\right]$$

- Likelihood

$$p(y|\sigma^2,\delta) = \int p(y|\sigma^2,\delta,\beta)\mathrm{d}\beta$$

$$\propto \frac{|A|^{-1/2}|H'A^{-1}H|^{-1/2}}{(2\pi\sigma^2)^{\frac{n-q}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(y - H\hat{\beta})'A^{-1}(y - H\hat{\beta})\right\}$$

- Parameters

$$\hat{\beta} = (H^{\mathrm{T}}A^{-1}H)^{-1}H^{\mathrm{T}}A^{-1}y \qquad \hat{\sigma}^2_{RL} = \frac{(y-H\hat{\beta})^{\mathrm{T}}A^{-1}(y-H\hat{\beta})}{n-q}$$

$$\hat{\delta}_{RL} = \arg\max_{\delta}\left[p(y|\sigma^2_{RL},\delta)\right]$$

The Toolkit Approach
Accounting for uncertainty in σ^2

- Likelihood

$$p(y|\delta) = \int p(y|\sigma^2, \delta)\sigma^{-2}\mathrm{d}\sigma^2$$

$$\propto \left[(y - H\hat{\beta})' A^{-1}(y - H\hat{\beta})\right]^{-\frac{n-q}{2}} |A|^{-1/2}|H'A^{-1}H|^{-1/2}$$

- Parameters

$$\hat{\beta} = (H^{\mathrm{T}}A^{-1}H)^{-1}H^{\mathrm{T}}A^{-1}y \qquad \hat{\sigma}^2 = \frac{(y-H\hat{\beta})^{\mathrm{T}}A^{-1}(y-H\hat{\beta})}{n-q}$$

$$\hat{\delta}_{TL} = \arg\max_{\delta} [p(y|\delta)]$$

The Toolkit Approach

- Likelihood

$$p(y|\delta) = \int p(y|\sigma^2, \delta)\sigma^{-2}\mathrm{d}\sigma^2$$

$$\propto \left[(y - H\hat{\beta})'A^{-1}(y - H\hat{\beta})\right]^{-\frac{n-q}{2}} |A|^{-1/2}|H'A^{-1}H|^{-1/2}$$

- Parameters

$$\hat{\beta} = (H^{\mathrm{T}}A^{-1}H)^{-1}H^{\mathrm{T}}A^{-1}y \qquad \hat{\sigma}^2 = \frac{(y-H\hat{\beta})^{\mathrm{T}}A^{-1}(y-H\hat{\beta})}{n-q}$$

$$\hat{\delta}_{TL} = \arg\max_{\delta} [p(y|\delta)]$$

- Likelihood

$$p(y|\delta) = \int p(y|\sigma^2, \delta)\sigma^{-2}\mathrm{d}\sigma^2$$

$$\propto \left[(y - H\hat{\beta})' A^{-1} (y - H\hat{\beta})\right]^{-\frac{n-q}{2}} |A|^{-1/2} |H' A^{-1} H|^{-1/2}$$

- Parameters

$$\hat{\beta} = (H^{\mathrm{T}} A^{-1} H)^{-1} H^{\mathrm{T}} A^{-1} y \qquad \hat{\sigma}^2 = \frac{(y - H\hat{\beta})^{\mathrm{T}} A^{-1} (y - H\hat{\beta})}{n - q}$$

$$\hat{\delta}_{TL} = \arg\max_{\delta} \left[p(y|\delta)\right]$$

- Likelihood

$$p(y|\delta) = \int p(y|\sigma^2, \delta)\sigma^{-2}\mathrm{d}\sigma^2$$

$$\propto \left[(y - H\hat{\beta})' A^{-1}(y - H\hat{\beta})\right]^{-\frac{n-q}{2}} |A|^{-1/2}|H' A^{-1} H|^{-1/2}$$

- Parameters

$$\hat{\beta} = (H^{\mathrm{T}} A^{-1} H)^{-1} H^{\mathrm{T}} A^{-1} y \qquad \hat{\sigma}^2 = \frac{(y-H\hat{\beta})^{\mathrm{T}} A^{-1}(y-H\hat{\beta})}{n-q}$$

$$\hat{\delta}_{TL} = \arg\max_{\delta}\,[p(y|\delta)]$$

Relationship between REML and the Toolkit Approach

- Likelihood

$$p_{\mathrm{RL}}(y|\sigma^2_{\mathrm{RL}}, \delta) \propto p_{\mathrm{TL}}(y|\delta)$$
$$\propto \left[(y - H\hat{\beta})' A^{-1} (y - H\hat{\beta})\right]^{-\frac{n-q}{2}} |A|^{-1/2} |H' A^{-1} H|^{-1/2}$$

- Prediction

REML	Toolkit
$\eta(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), \sigma^2 u(\mathbf{x}, \mathbf{x}'))$	$\frac{\eta(\mathbf{x}) - m(\mathbf{x})}{\sqrt{\sigma^2 u(\mathbf{x}, \mathbf{x}')}} \sim t_{n-q}$

- Both methods have the same predictive mean $m(\mathbf{x})$
- The variance of the Toolkit method is $\frac{n-q}{n-q-2}$ times the variance of REML

Relationship between REML and the Toolkit Approach

- Likelihood

$$p_{\rm RL}(y|\sigma^2_{\rm RL},\delta) \propto p_{\rm TL}(y|\delta)$$
$$\propto \left[(y - H\hat{\beta})' A^{-1}(y - H\hat{\beta})\right]^{-\frac{n-q}{2}} |A|^{-1/2}|H'A^{-1}H|^{-1/2}$$

- Prediction

REML	Toolkit
$\eta(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), \sigma^2 u(\mathbf{x},\mathbf{x}'))$	$\frac{\eta(\mathbf{x})-m(\mathbf{x})}{\sqrt{\sigma^2 u(\mathbf{x},\mathbf{x}')}} \sim t_{n-q}$

- Both methods have the same predictive mean $m(\mathbf{x})$
- The variance of the Toolkit method is $\frac{n-q}{n-q-2}$ times the variance of REML

Relationship between REML and the Toolkit Approach

- Likelihood

$$p_{\mathrm{RL}}(y|\sigma^2_{\mathrm{RL}}, \delta) \propto p_{\mathrm{TL}}(y|\delta)$$

$$\propto \left[(y - H\hat{\beta})' A^{-1}(y - H\hat{\beta})\right]^{-\frac{n-q}{2}} |A|^{-1/2}|H'A^{-1}H|^{-1/2}$$

- Prediction

REML	Toolkit
$\eta(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), \sigma^2 u(\mathbf{x}, \mathbf{x}'))$	$\frac{\eta(\mathbf{x})-m(\mathbf{x})}{\sqrt{\sigma^2 u(\mathbf{x},\mathbf{x}')}} \sim t_{n-q}$

- Both methods have the same predictive mean $m(\mathbf{x})$
- The variance of the Toolkit method is $\frac{n-q}{n-q-2}$ times the variance of REML

Relationship between REML and the Toolkit Approach

- Likelihood

$$p_{\mathrm{RL}}(y|\sigma^2_{\mathrm{RL}}, \delta) \propto p_{\mathrm{TL}}(y|\delta)$$
$$\propto \left[(y - H\hat{\beta})' A^{-1} (y - H\hat{\beta})\right]^{-\frac{n-q}{2}} |A|^{-1/2} |H' A^{-1} H|^{-1/2}$$

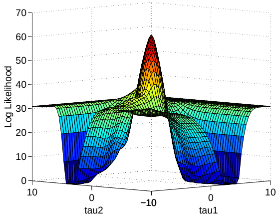
- Prediction

REML	Toolkit
$\eta(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), \sigma^2 u(\mathbf{x}, \mathbf{x}'))$	$\frac{\eta(\mathbf{x}) - m(\mathbf{x})}{\sqrt{\sigma^2 u(\mathbf{x}, \mathbf{x}')}} \sim t_{n-q}$

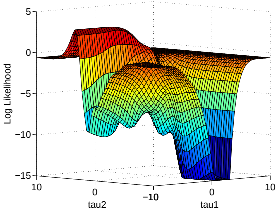
- Both methods have the same predictive mean $m(\mathbf{x})$
- The variance of the Toolkit method is $\frac{n-q}{n-q-2}$ times the variance of REML

The effect of *n* on parameter estimation

$$\ln(p_{\mathrm{RL}}(y|\sigma_{\mathrm{RL}},\tau)), \qquad \delta^{-2} = e^{\tau}$$



(a) $n = 20$

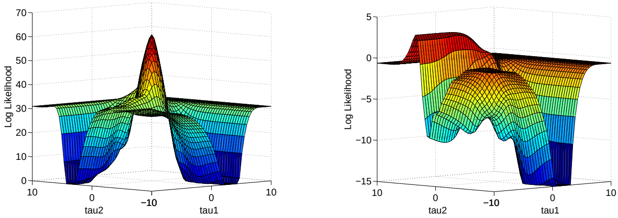


(b) $n = 10$

- The data was drawn from a zero mean GP with $\delta = \mathbf{1}$, $(\tau = \mathbf{0})$

The effect of n on parameter estimation

$$\ln(p_{\mathrm{RL}}(y|\sigma_{\mathrm{RL}},\tau)), \qquad \delta^{-2} = e^{\tau}$$



(c) $n = 20$ (d) $n = 10$

- The data was drawn from a zero mean GP with $\delta = \mathbf{1}$, $(\tau = \mathbf{0})$

Experiment

- Input dimension
- Correlation length
- Design points per input
- Design drawn from LHC in $[0, 1]$
- # of realisations for which $\max(\delta) < \delta_{\text{thr}}$
- Assume that when $\max(\delta) > \delta_{\text{thr}}$, data are explained with less inputs

Experiment

- Input dimension
- Correlation length
- Design points per input
- Design drawn from LHC in [0, 1]

- # of realisations for which $\max(\delta) < \delta_{\text{thr}}$
- Assume that when $\max(\delta) > \delta_{\text{thr}}$, data are explained with less inputs

Experiment

- Input dimension
- Correlation length
- Design points per input
- Design drawn from LHC in $[0, 1]$
- # of realisations for which $\max(\delta) < \delta_{\text{thr}}$
- Assume that when $\max(\delta) > \delta_{\text{thr}}$, data are explained with less inputs

Experiment

- Input dimension
- Correlation length
- Design points per input
- Design drawn from LHC in $[0, 1]$
- # of realisations for which $\max(\delta) < \delta_{\text{thr}}$
- Assume that when $\max(\delta) > \delta_{\text{thr}}$, data are explained with less inputs

Experiment

- Input dimension
- Correlation length
- Design points per input
- Design drawn from LHC in $[0, 1]$
- # of realisations for which $\max(\delta) < \delta_{\text{thr}}$
- Assume that when $\max(\delta) > \delta_{\text{thr}}$, data are explained with less inputs

Experiment

- Input dimension
- Correlation length
- Design points per input
- Design drawn from LHC in $[0, 1]$

- # of realisations for which $\max(\delta) < \delta_{\text{thr}}$
- Assume that when $\max(\delta) > \delta_{\text{thr}}$, data are explained with less inputs

Experiment

- Input dimension
- Correlation length
- Design points per input
- Design drawn from LHC in $[0, 1]$

- # of realisations for which $\max(\delta) < \delta_{\text{thr}}$
- Assume that when $\max(\delta) > \delta_{\text{thr}}$, data are explained with less inputs

Results - $\delta = 1$

REML					
n/p	$p = 2$	$p = 3$	$p = 5$	$p = 8$	$p = 10$
5	93	81	46	42	39
10	**100**	**100**	99	92	81
15	100	100	**100**	99	97
20	100	100	100	**100**	**100**
25	100	100	100	100	100
30	100	100	100	100	100

ML					
n/p	$p = 2$	$p = 3$	$p = 5$	$p = 8$	$p = 10$
5	94	77	23	12	11
10	**100**	**100**	99	69	39
15	100	100	**100**	98	88
20	100	100	100	**100**	99
25	100	100	100	100	**100**
30	100	100	100	100	100

Results - $\delta = 1$

REML					
n/p	$p=2$	$p=3$	$p=5$	$p=8$	$p=10$
5	93	81	46	42	39
10	**100**	**100**	99	92	81
15	100	100	**100**	99	97
20	100	100	100	**100**	**100**
25	100	100	100	100	100
30	100	100	100	100	100

ML					
n/p	$p=2$	$p=3$	$p=5$	$p=8$	$p=10$
5	94	77	23	12	11
10	**100**	**100**	99	69	39
15	100	100	**100**	98	88
20	100	100	100	**100**	99
25	100	100	100	100	**100**
30	100	100	100	100	100

Results - $\delta = 0.3$

REML					
n/p	$p = 2$	$p = 3$	$p = 5$	$p = 8$	$p = 10$
5	92	75	8	0	0
10	**100**	99	59	1	0
20	100	**100**	93	12	0
30	100	100	99	28	1
50	100	100	**100**	58	7

ML					
n/p	$p = 2$	$p = 3$	$p = 5$	$p = 8$	$p = 10$
5	82	57	6	0	0
10	**100**	95	38	0	0
20	100	**100**	89	2	0
30	100	100	99	17	1
50	100	100	**100**	50	6

Results - $\delta = 0.3$

REML					
n/p	$p = 2$	$p = 3$	$p = 5$	$p = 8$	$p = 10$
5	92	75	8	0	0
10	**100**	99	59	1	0
20	100	**100**	93	12	0
30	100	100	99	28	1
50	100	100	**100**	58	7

ML					
n/p	$p = 2$	$p = 3$	$p = 5$	$p = 8$	$p = 10$
5	82	57	6	0	0
10	**100**	95	38	0	0
20	100	**100**	89	2	0
30	100	100	99	17	1
50	100	100	**100**	50	6

(Normalised) Mahalanobis distance ($\delta = 1$)

REML					
n/p	$p = 2$	$p = 3$	$p = 5$	$p = 8$	$p = 10$
5	5.98	5.58	5.06	2.81	2.53
10	1.50	1.86	2.72	2.07	1.93
15	**0.73**	0.97	1.62	1.42	1.43
20	0.38	**0.62**	1.21	1.15	1.20
25	0.37	0.58	**0.97**	**0.92**	1.01
30	0.27	0.43	0.69	0.84	**0.87**

ML					
n/p	$p = 2$	$p = 3$	$p = 5$	$p = 8$	$p = 10$
5	8.16	5.42	5.32	6.17	5.60
10	3.53	1.84	2.04	2.93	3.16
15	1.70	1.58	1.21	1.98	2.29
20	**0.89**	1.37	1.01	1.48	1.84
25	0.75	1.24	**0.88**	1.20	1.45
30	0.58	**1.00**	0.71	1.03	1.27

- $n/2$ points were used for prediction
- The 'normalised' Mahalanobis distance is $D' = (D - \mu)/(\sigma)$ where D is the Mahalanobis distance, μ its theoretical mean and σ its standard deviation

(Normalised) Mahalanobis distance ($\delta = 1$)

REML					
n/p	$p = 2$	$p = 3$	$p = 5$	$p = 8$	$p = 10$
5	5.98	5.58	5.06	2.81	2.53
10	1.50	1.86	2.72	2.07	1.93
15	**0.73**	0.97	1.62	1.42	1.43
20	0.38	**0.62**	1.21	1.15	1.20
25	0.37	0.58	**0.97**	**0.92**	1.01
30	0.27	0.43	0.69	0.84	**0.87**

ML					
n/p	$p = 2$	$p = 3$	$p = 5$	$p = 8$	$p = 10$
5	8.16	5.42	5.32	6.17	5.60
10	3.53	1.84	2.04	2.93	3.16
15	1.70	1.58	1.21	1.98	2.29
20	**0.89**	1.37	1.01	1.48	1.84
25	0.75	1.24	**0.88**	1.20	1.45
30	0.58	**1.00**	0.71	1.03	1.27

- $n/2$ points were used for prediction
- The 'normalised' Mahalanobis distance is $D' = (D - \mu)/(\sigma)$ where D is the Mahalanobis distance, μ its theoretical mean and σ its standard deviation

Comparison between ML and REML

- REML was able to explain the data using all the inputs more often
- Mahalanobis distances that were generally closer to the theoretical mean for REML
- REML estimates of σ^2 are unbiased
- REML estimates of δ were higher than those of ML
- For $n >>$ the results of the two methods were virtually identical

Comparison between ML and REML

- REML was able to explain the data using all the inputs more often
- Mahalanobis distances that were generally closer to the theoretical mean for REML
- REML estimates of σ^2 are unbiased
- REML estimates of δ were higher than those of ML
- For $n >>$ the results of the two methods were virtually identical

Comparison between ML and REML

- REML was able to explain the data using all the inputs more often
- Mahalanobis distances that were generally closer to the theoretical mean for REML
- REML estimates of σ^2 are unbiased
- REML estimates of δ were higher than those of ML
- For $n >>$ the results of the two methods were virtually identical

Comparison between ML and REML

- REML was able to explain the data using all the inputs more often
- Mahalanobis distances that were generally closer to the theoretical mean for REML
- REML estimates of σ^2 are unbiased
- REML estimates of δ were higher than those of ML
- For $n >>$ the results of the two methods were virtually identical

Comparison between ML and REML

- REML was able to explain the data using all the inputs more often
- Mahalanobis distances that were generally closer to the theoretical mean for REML
- REML estimates of σ^2 are unbiased
- REML estimates of δ were higher than those of ML
- For $n >>$ the results of the two methods were virtually identical

Comparison between ML and REML

- REML was able to explain the data using all the inputs more often
- Mahalanobis distances that were generally closer to the theoretical mean for REML
- REML estimates of σ^2 are unbiased
- REML estimates of δ were higher than those of ML
- For $n >>$ the results of the two methods were virtually identical

REML dissection!

ML

$$L_{\text{ML}} \propto L^1_{\text{ML}} + L^2_{\text{ML}}$$

$$L^1_{\text{ML}} = -\frac{1}{2}\ln|A|$$

$$L^2_{\text{ML}} = -\frac{n}{2}\ln(y - H\hat{\beta})^{\text{T}}A^{-1}(y - H\hat{\beta})$$

REML

$$L_{\text{RL}} \propto L^1_{\text{RL}} + L^2_{\text{RL}} + L^3_{\text{RL}}$$

$$L^1_{\text{RL}} = -\frac{1}{2}\ln|A|$$

$$L^2_{\text{RL}} = -\frac{n-q}{2}\ln(y - H\hat{\beta})^{\text{T}}A^{-1}(y - H\hat{\beta})$$

$$L^3_{\text{RL}} = -\frac{1}{2}\ln|H^{\text{T}}A^{-1}H|$$

Complexity Penalty

Data fit

REML only

REML dissection!

ML

$$L_{\mathrm{ML}} \propto L^1_{\mathrm{ML}} + L^2_{\mathrm{ML}}$$

$$L^1_{\mathrm{ML}} = -\frac{1}{2}\ln|A|$$

$$L^2_{\mathrm{ML}} = -\frac{n}{2}\ln(y - H\hat{\beta})^{\mathrm{T}} A^{-1}(y - H\hat{\beta})$$

REML

$$L_{\mathrm{RL}} \propto L^1_{\mathrm{RL}} + L^2_{\mathrm{RL}} + L^3_{\mathrm{RL}}$$

$$L^1_{\mathrm{RL}} = -\frac{1}{2}\ln|A|$$

$$L^2_{\mathrm{RL}} = -\frac{n-q}{2}\ln(y - H\hat{\beta})^{\mathrm{T}} A^{-1}(y - H\hat{\beta})$$

$$L^3_{\mathrm{RL}} = -\frac{1}{2}\ln|H^{\mathrm{T}} A^{-1} H|$$

Complexity Penalty

Data fit

REML only

REML dissection!

ML

$$L_{\mathrm{ML}} \propto L^1_{\mathrm{ML}} + L^2_{\mathrm{ML}}$$
$$L^1_{\mathrm{ML}} = -\frac{1}{2} \ln |A|$$
$$L^2_{\mathrm{ML}} = -\frac{n}{2} \ln (y - H\hat{\beta})^{\mathrm{T}} A^{-1} (y - H\hat{\beta})$$

REML

$$L_{\mathrm{RL}} \propto L^1_{\mathrm{RL}} + L^2_{\mathrm{RL}} + L^3_{\mathrm{RL}}$$
$$L^1_{\mathrm{RL}} = -\frac{1}{2} \ln |A|$$
$$L^2_{\mathrm{RL}} = -\frac{n-q}{2} \ln (y - H\hat{\beta})^{\mathrm{T}} A^{-1} (y - H\hat{\beta})$$
$$L^3_{\mathrm{RL}} = -\frac{1}{2} \ln |H^{\mathrm{T}} A^{-1} H|$$

Complexity Penalty   Data fit   REML only

REML dissection!

ML

$$L_{\mathrm{ML}} \propto L^1_{\mathrm{ML}} + L^2_{\mathrm{ML}}$$

$$L^1_{\mathrm{ML}} = -\frac{1}{2} \ln |A|$$

$$L^2_{\mathrm{ML}} = -\frac{n}{2} \ln (y - H\hat{\beta})^{\mathrm{T}} A^{-1} (y - H\hat{\beta})$$

REML

$$L_{\mathrm{RL}} \propto L^1_{\mathrm{RL}} + L^2_{\mathrm{RL}} + L^3_{\mathrm{RL}}$$

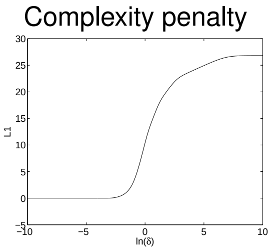
$$L^1_{\mathrm{RL}} = -\frac{1}{2} \ln |A|$$

$$L^2_{\mathrm{RL}} = -\frac{n-q}{2} \ln (y - H\hat{\beta})^{\mathrm{T}} A^{-1} (y - H\hat{\beta})$$

$$L^3_{\mathrm{RL}} = -\frac{1}{2} \ln |H^{\mathrm{T}} A^{-1} H|$$

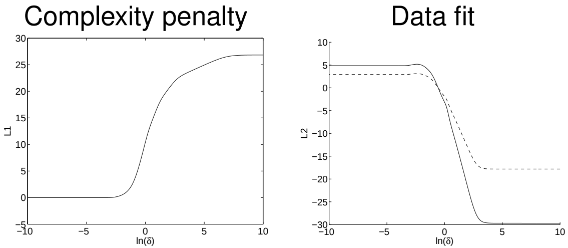
Complexity Penalty | Data fit | REML only

L1 and L2 terms



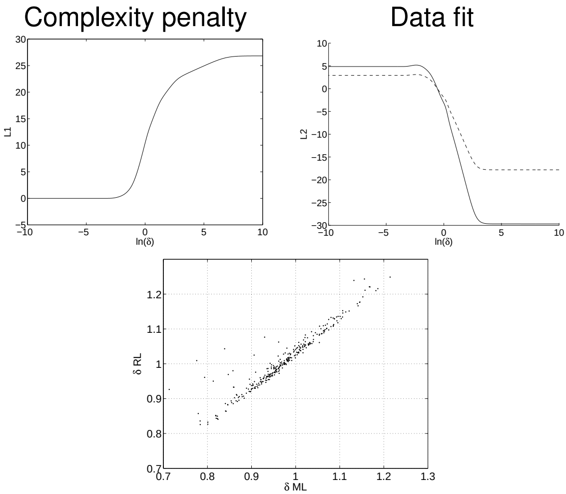
$$L^1_{\mathrm{ML}} = L^1_{\mathrm{RL}} = -\frac{1}{2}\ln|A|$$

L1 and L2 terms

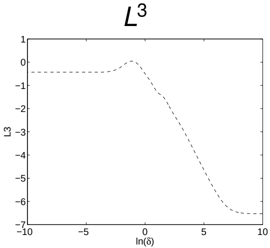


$$L^2_{\mathrm{RL}} = -\frac{n-q}{2}\ln(y - H\hat{\beta})^{\mathrm{T}} A^{-1}(y - H\hat{\beta})$$

L1 and L2 terms

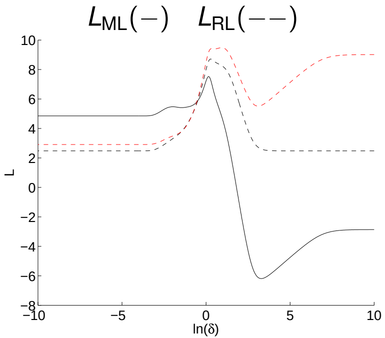


L3 term



$$L^3_{\mathrm{RL}} = -\frac{1}{2} \ln \left| H^{\mathrm{T}} A^{-1} H \right|$$

L3 term



$L_{RL} - L_{RL}^3$

Conclusions

Design points

- 10 points per dimension are enough only for $p < 3$ and large δ
- For relatively smooth functions 20-30 points per dimension is a safer choice for $p \leq 10$
- More points are needed for higher dimensions
- More points are needed for more 'wiggly' functions

ML - REML

- Data explanation using more inputs
- Better Mahalanobis distances
- Unbiased σ^2
- $\delta_{\mathrm{RL}} > \delta_{\mathrm{ML}}$
- For $n >>$ the results of the two methods were virtually identical

Conclusions

Design points

- 10 points per dimension are enough only for $p < 3$ and large δ
- For relatively smooth functions 20-30 points per dimension is a safer choice for $p \leq 10$
- More points are needed for higher dimensions
- More points are needed for more 'wiggly' functions

ML - REML

- Data explanation using more inputs
- Better Mahalanobis distances
- Unbiased σ^2
- $\delta_{\text{RL}} > \delta_{\text{ML}}$
- For $n >>$ the results of the two methods were virtually identical

Conclusions

Design points

- 10 points per dimension are enough only for $p < 3$ and large δ
- For relatively smooth functions 20-30 points per dimension is a safer choice for $p \leq 10$
- More points are needed for higher dimensions
- More points are needed for more 'wiggly' functions

ML - REML

- Data explanation using more inputs
- Better Mahalanobis distances
- Unbiased σ^2
- $\delta_{\text{RL}} > \delta_{\text{ML}}$
- For $n >>$ the results of the two methods were virtually identical

Conclusions

Design points

- 10 points per dimension are enough only for $p < 3$ and large δ
- For relatively smooth functions 20-30 points per dimension is a safer choice for $p \leq 10$
- More points are needed for higher dimensions
- More points are needed for more 'wiggly' functions

ML - REML

- Data explanation using more inputs
- Better Mahalanobis distances
- Unbiased σ^2
- $\delta_{RL} > \delta_{ML}$
- For $n >>$ the results of the two methods were virtually identical

Conclusions

Design points

- 10 points per dimension are enough only for $p < 3$ and large δ
- For relatively smooth functions 20-30 points per dimension is a safer choice for $p \leq 10$
- More points are needed for higher dimensions
- More points are needed for more 'wiggly' functions

ML - REML

- Data explanation using more inputs
- Better Mahalanobis distances
- Unbiased σ^2
- $\delta_{RL} > \delta_{ML}$
- For $n >>$ the results of the two methods were virtually identical

Conclusions

Design points

- 10 points per dimension are enough only for $p < 3$ and large δ
- For relatively smooth functions 20-30 points per dimension is a safer choice for $p \leq 10$
- More points are needed for higher dimensions
- More points are needed for more 'wiggly' functions

ML - REML

- Data explanation using more inputs
- Better Mahalanobis distances
- Unbiased σ^2
- $\delta_{RL} > \delta_{ML}$
- For $n >>$ the results of the two methods were virtually identical

Conclusions

Design points

- 10 points per dimension are enough only for $p < 3$ and large δ
- For relatively smooth functions 20-30 points per dimension is a safer choice for $p \leq 10$
- More points are needed for higher dimensions
- More points are needed for more 'wiggly' functions

ML - REML

- Data explanation using more inputs
- Better Mahalanobis distances
- Unbiased σ^2
- $\delta_{RL} > \delta_{ML}$
- For $n >>$ the results of the two methods were virtually identical

Conclusions

Design points

- 10 points per dimension are enough only for $p < 3$ and large δ
- For relatively smooth functions 20-30 points per dimension is a safer choice for $p \leq 10$
- More points are needed for higher dimensions
- More points are needed for more 'wiggly' functions

ML - REML

- Data explanation using more inputs
- Better Mahalanobis distances
- Unbiased σ^2
- $\delta_{RL} > \delta_{ML}$
- For $n >>$ the results of the two methods were virtually identical

Conclusions

Design points

- 10 points per dimension are enough only for $p < 3$ and large δ
- For relatively smooth functions 20-30 points per dimension is a safer choice for $p \leq 10$
- More points are needed for higher dimensions
- More points are needed for more 'wiggly' functions

ML - REML

- Data explanation using more inputs
- Better Mahalanobis distances
- Unbiased σ^2
- $\delta_{RL} > \delta_{ML}$
- For $n >>$ the results of the two methods were virtually identical

Conclusions

Design points

- 10 points per dimension are enough only for $p < 3$ and large δ
- For relatively smooth functions 20-30 points per dimension is a safer choice for $p \leq 10$
- More points are needed for higher dimensions
- More points are needed for more 'wiggly' functions

ML - REML

- Data explanation using more inputs
- Better Mahalanobis distances
- Unbiased σ^2
- $\delta_{RL} > \delta_{ML}$
- For $n >>$ the results of the two methods were virtually identical

Conclusions

Design points

- 10 points per dimension are enough only for $p < 3$ and large δ
- For relatively smooth functions 20-30 points per dimension is a safer choice for $p \leq 10$
- More points are needed for higher dimensions
- More points are needed for more 'wiggly' functions

ML - REML

- Data explanation using more inputs
- Better Mahalanobis distances
- Unbiased σ^2
- $\delta_{\text{RL}} > \delta_{\text{ML}}$
- For $n >>$ the results of the two methods were virtually identical

Future Work

- Priors Informative-non informative, proper posterior distributions
- Derivatives - speeding up the optimisation
- Uncertainty in δ MCMC - Normal approximation of the posterior
- Specialised design for parameter estimation

Future Work

- Priors Informative-non informative, proper posterior distributions
- Derivatives - speeding up the optimisation
- Uncertainty in δ MCMC - Normal approximation of the posterior
- Specialised design for parameter estimation

Future Work

- Priors Informative-non informative, proper posterior distributions
- Derivatives - speeding up the optimisation
- Uncertainty in δ MCMC - Normal approximation of the posterior
- Specialised design for parameter estimation

Future Work

- Priors Informative-non informative, proper posterior distributions
- Derivatives - speeding up the optimisation
- Uncertainty in δ MCMC - Normal approximation of the posterior
- Specialised design for parameter estimation

Future Work

- Priors Informative-non informative, proper posterior distributions
- Derivatives - speeding up the optimisation
- Uncertainty in δ MCMC - Normal approximation of the posterior
- Specialised design for parameter estimation