Home Page

Title Page

Contents

Page 6 of 30

Go Back

Full Screen

Close

Quit

- Given N obs $\mathbf{t} = \{t_1 \ldots t_N\} \equiv$ input $\mathbf{X} = \{\mathbf{x}_1 \ldots \mathbf{x}_N\}$
- Assume i.i.d.: independence: probs multiplied, identically distr: all from same model $\mathcal{N}(t|y(\cdot), \sigma^2)$
- $\boxed{p(\mathbf{t}|\mathbf{w}, \sigma^2) = \prod_{i=1}^{N} \mathcal{N}(t_i|y(\mathbf{x}_i, \mathbf{w}), \sigma^2)}$ $y(\cdot) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i)$
- LHS: Likelihood: prob of getting the data, given model params: Reasonable to maximise it!
- Maximise log-likelihood: inc fn, mults $\rightarrow$ additions
- $\log p(\mathbf{t}|\mathbf{w}, \sigma^2) = \sum_{i=1}^{N} \log \mathcal{N}(t_i|y(\mathbf{x}_i, \mathbf{w}), \sigma^2)$
- log-likelihood $= \sum_{i=1}^{N} \log(\frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(t_i - \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i))^2}{2\sigma^2})$
- $= -\frac{N}{2} \log 2\pi - \frac{N}{2} \log \sigma^2 - \frac{1}{\sigma^2} \frac{1}{2} \sum_{i=1}^{N} (t_i - \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i))^2$
- Params σ^2, $\mathbf{w}$: $\frac{\partial \text{ log-likelihood}}{\partial \text{parameter}} = 0$

Home Page

Title Page

Contents

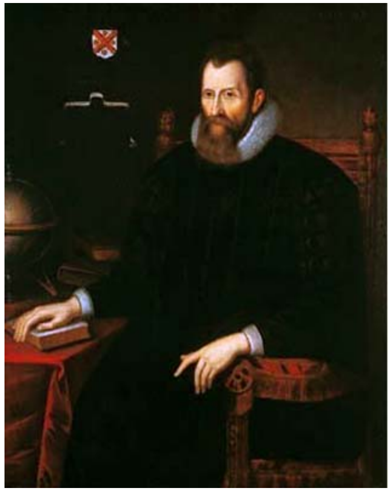
Page 7 of 30

Go Back

Full Screen

Close

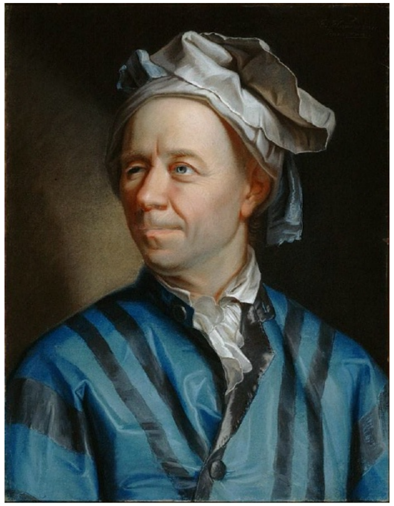
Quit



Logarithms

John Napier
[1550-1617]

Leonhard Euler
[1707-1783]



https://upload.wikimedia.org/wikipedia/commons/e/e3/John_Napier.jpg

https://upload.wikimedia.org/wikipedia/commons/d/d7/Leonhard_Euler.jpg

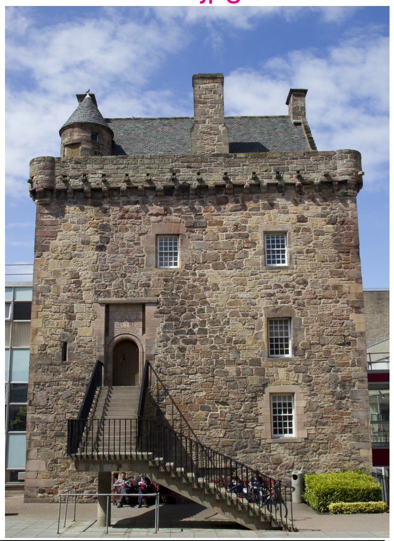
Edinburgh Napier
UNIVERSITY

https://upload.wikimedia.org/wikipedia/en/e/e4/

Edinburgh_Napier_University_logo.png

https://upload.wikimedia.org/wikipedia/commons/1/1f/

Merchiston_Castle



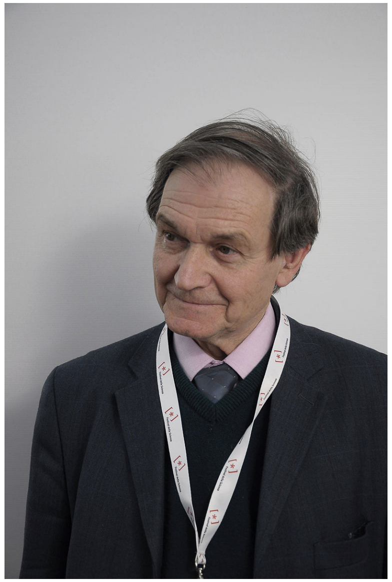
- param = σ^2:
  $$\frac{\partial \text{ log-lh}}{\partial \sigma^2} = -\frac{N}{2}\frac{1}{\sigma^2} - \frac{-1}{\sigma^4}\frac{1}{2}\sum_{i=1}^{N}(t_i - \mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i))^2 = 0$$
- $\boxed{\sigma^2_{ML} = \frac{1}{N}\sum_{i=1}^{N}(t_i - \mathbf{w}^T_{ML}\boldsymbol{\phi}(\mathbf{x}_i))^2}$; $y(\mathbf{x}_i, \mathbf{w}) = \mathbf{w}^T_{ML}\boldsymbol{\phi}(\mathbf{x}_i)$
- Phy Sig: The ML variance is the sample variance of the target values around the regression fn
- param = $\mathbf{w}$: log-lh $=$ $\qquad (\sum a_i^2 = \mathbf{a}^T\mathbf{a}; \frac{\partial \mathbf{a}^T\mathbf{a}}{\partial \mathbf{a}} = 2\mathbf{a}^T)$
  $$-\frac{N}{2}\log 2\pi - \frac{N}{2}\log\sigma^2 - \frac{1}{\sigma^2}\frac{1}{2}\sum_{i=1}^{N}(t_i - \boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w})^2$$
- $\frac{\partial \text{log-lh}}{\partial \mathbf{w}^T} = -\frac{1}{\sigma^2}\frac{1}{2}2(-1)\sum_{i=1}^{N}(t_i - \mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_i))\boldsymbol{\phi}^T(\mathbf{x}_i) = 0$:
- $\implies \sum_{i=1}^{N} t_i\boldsymbol{\phi}^T(\mathbf{x}_i) = \mathbf{w}^T\sum_{i=1}^{N}\boldsymbol{\phi}(\mathbf{x}_i)\boldsymbol{\phi}^T(\mathbf{x}_i) \qquad$ ($\mathbf{w}^T$ out)
- Break the sums up as inner products

- $\sum_{i=1}^{N} t_i \boldsymbol{\phi}^T(\mathbf{x}_i) = \mathbf{w}^T \sum_{i=1}^{N} \boldsymbol{\phi}(\mathbf{x}_i)\boldsymbol{\phi}^T(\mathbf{x}_i) \implies$
- $[t_1 \ldots t_N] \begin{bmatrix} \boldsymbol{\phi}^T(\mathbf{x}_1) \\ \cdots \\ \boldsymbol{\phi}^T(\mathbf{x}_N) \end{bmatrix} = \mathbf{w}^T [\boldsymbol{\phi}(\mathbf{x}_1) \ldots \boldsymbol{\phi}(\mathbf{x}_N)] \begin{bmatrix} \boldsymbol{\phi}^T(\mathbf{x}_1) \\ \cdots \\ \boldsymbol{\phi}^T(\mathbf{x}_N) \end{bmatrix}$
- $\implies \mathbf{t}^T \boldsymbol{\Phi} = \mathbf{w}^T(\boldsymbol{\Phi}^T\boldsymbol{\Phi})$. 'Design Matrix': bases $\forall \mathbf{x}_i$
- $\boldsymbol{\Phi}_{N\times M} = \begin{bmatrix} \boldsymbol{\phi}^T(\mathbf{x}_1) \\ \cdots \\ \boldsymbol{\phi}^T(\mathbf{x}_N) \end{bmatrix} = \begin{bmatrix} \phi_0(\mathbf{x}_1) \ldots \phi_{M-1}(\mathbf{x}_1) \\ \cdots \\ \phi_0(\mathbf{x}_N) \ldots \phi_{M-1}(\mathbf{x}_N) \end{bmatrix}_{N\times M}$
- Transpose both sides: $(\mathbf{t}^T\boldsymbol{\Phi})^T = (\mathbf{w}^T(\boldsymbol{\Phi}^T\boldsymbol{\Phi}))^T \implies$
- $\boldsymbol{\Phi}^T\mathbf{t} = \boldsymbol{\Phi}^T\boldsymbol{\Phi}\mathbf{w} \implies \boxed{\mathbf{w}_{ML} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\mathbf{t} = \boldsymbol{\Phi}^\dagger\mathbf{t}}$
- $\boldsymbol{\Phi}^\dagger$: Moore-Penrose Pseudo-Inverse

Moore-Penrose Pseudo-Inverse

E. H. Moore
(1920)
[1862-1932]

A. Bjerhammar
(1951)
[1917-2011]



R. Penrose
(1955)
[1931-]

https://upload.wikimedia.org/wikipedia/commons/9/96/Moore_Eliakim_2.jpeg

https://thumbnail.myheritageimages.com/021/731/40021731/500/500151_62403790fd28b4g41eg208_W_189x256.jpg

https://upload.wikimedia.org/wikipedia/commons/thumb/d/d5/Roger_Penrose_at_Festival_della_Scienza_Oct_29_2011.jpg/800px-Roger_Penrose_at_Festival_della_Scienza_Oct_29_2011.jpg

Home Page
Title Page
Contents
Page 11 of 30
Go Back
Full Screen
Close
Quit

Moore-Penrose Pseudo-Inverse

- $\mathbf{Bx} = \mathbf{c}$, $\mathbf{B}$: non-square. $\mathbf{x} = \mathbf{B}^{-1}\mathbf{c}$?
- $\mathbf{B}^T\mathbf{Bx} = \mathbf{B}^T\mathbf{c}$: $\boxed{\mathbf{x} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{c}}$
  Moore-Penrose pseudo-inverse, do not compute inverse: the SVD does it algorithmically
- Another derivation: $f(\mathbf{x}) = ||\mathbf{Bx} - \mathbf{c}||^2$, $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = ?$
- Method 1: $f(\mathbf{x}) = ||\mathbf{Bx} - \mathbf{c}||^2 = (\mathbf{Bx} - \mathbf{c})^T(\mathbf{Bx} - \mathbf{c})$,
  $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{B}^T(\mathbf{Bx} - \mathbf{c}) = \mathbf{0}$ (vector calculus rules)
  $2\mathbf{B}^T\mathbf{Bx} = 2\mathbf{B}^T\mathbf{c}$, get the pseudo-inverse above
- Method 2: $f(\mathbf{x}) = (\mathbf{Bx} - \mathbf{c})^T(\mathbf{Bx} - \mathbf{c})$
  $= \mathbf{x}^T\mathbf{B}^T\mathbf{Bx} - \mathbf{x}^T\mathbf{B}^T\mathbf{c} - \mathbf{c}^T\mathbf{Bx} + \mathbf{c}^T\mathbf{c}$. Middle terms scalars, equal! ($(\mathbf{x}^T\mathbf{B}^T\mathbf{c})^T = \mathbf{c}^T\mathbf{Bx}$. $f(\mathbf{x}) = \mathbf{x}^T\mathbf{B}^T\mathbf{Bx} - 2\mathbf{c}^T\mathbf{Bx} + \mathbf{c}^T\mathbf{c}$. Differentiate!