Home Page
Title Page
Contents
Page 18 of 32
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

Maximising the log-likelihood

$\log p(X|\ \mu,\ \sigma^2) = \log \prod_{i=1}^{N} \mathcal{N}(x_i|\ \mu,\ \sigma^2)$, to maximise

$= \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\ \exp - \frac{(x_i-\mu)^2}{2\sigma^2}\right)$

$= -\left(\frac{N}{2}\right)\log 2\pi - \left(\frac{N}{2}\right)\log \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2$

How to maximise the log-likelihood?

Variational Calculus: $\frac{\partial \text{ log-likelihood}}{\partial \text{parameter}} = 0$

put it back in the log-likelihood, check for max/min

Better than 2nd deriv test, higher orders...

Parameters? $\mu,\ \sigma^2$

1. $\frac{\partial \text{ log-lh}}{\partial \mu} = 0 \implies -\frac{1}{2\sigma^2}(2)(-1)\sum_{i=1}^{N}(x_i-\mu) = 0$
$\implies \sum_{i=1}^{N} x_i = \mu N \implies \mu_{ML} = \sum_{i=1}^{N} x_i / N$

2. $\frac{\partial \text{ log-lh}}{\partial \sigma^2} = 0 \implies -\frac{N}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{1}{\sigma^4}\sum_{i=1}^{N}(x_i-\mu_{ML})^2 \implies$
$\frac{\sum_{i=1}^{N}(x_i-\mu_{ML})^2}{\sigma^2} = N \implies \sigma^2 = \sum_{i=1}^{N}(x_i-\mu_{ML})^2/N$

ML mean: sample mean, ML var: sample var

Param Est: K D-D Gaussians

- Given: $\mathbf{X} = \{\mathbf{x}_1, \ldots \mathbf{x}_N\}$, a set of N observations
- Model: K D-D Gaussians, means $\boldsymbol{\mu}_j$, Covs $\boldsymbol{\Sigma}_j$, Mixture coeffs π_j. 3 sets of parameters: $\boldsymbol{\mu}$,$\boldsymbol{\Sigma}$,$\boldsymbol{\pi}$
- Assumptions: Data points i.i.d. Independent: allows marginal prob multiplication without considering conditional dependence terms. Identically distributed: all from same model
- Method: $p(\mathbf{X}|\ \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$: Likelihood, to maximise
- Reasonable? Find params which maximise the likelihood of getting these points, given our model
- $p(\mathbf{X}|\ \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^{N} \sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_i|\ \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, to max
- $\equiv$ Maximise log-likelihood. Why? (John Napier)

Maximising the log-likelihood

$$\log p(\mathbf{X}|\ \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^{N} \log \sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_i|\ \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$$

$$\mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \triangleq \frac{1}{\sqrt{(2\pi)^{\mathcal{D}}|\boldsymbol{\Sigma}_j|}} \exp -\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_j)$$

How to maximise the log-likelihood?

Variational Calculus: $\frac{\partial \text{ log-likelihood}}{\partial \text{parameter}} = 0$

$$\sum_{i=1}^{N} \frac{(+1)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \frac{\partial \pi_j \mathcal{N}(\cdot)}{\partial \text{parameter}} = 0$$

- parameter #1: $\boldsymbol{\mu}_j$: $\frac{\partial \text{log-lh}}{\partial \boldsymbol{\mu}_j} = 0 \implies$

$$\sum_{i=1}^{N} \frac{(+1)\pi_j \exp -\frac{1}{2}[(\mathbf{x}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_j)]}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \frac{2(-1)}{2} \boldsymbol{\Sigma}_j^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_j) = 0$$

$$\implies \frac{-2}{2} \sum_{i=1}^{N} \left[\frac{\pi_j \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}\right] \boldsymbol{\Sigma}_j^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_j) = 0$$

$$\sum_{i=1}^{N} \gamma_j(\mathbf{x}_i).\mathbf{x}_i = \boldsymbol{\mu}_j \sum_{i=1}^{N} \gamma_j(\mathbf{x}_i) \implies \boldsymbol{\mu}_j = \frac{\sum_{i=1}^{N} \gamma_j(\mathbf{x}_i).\mathbf{x}_i}{\sum_{i=1}^{N} \gamma_j(\mathbf{x}_i)}$$

$$\implies \boxed{\boldsymbol{\mu}_j = \frac{\sum_{i=1}^{N} \gamma_j(\mathbf{x}_i).\mathbf{x}_i}{N_j}}$$ prob (resp)-weighted mean

Home Page

Title Page

Contents

Page 22 of 32

Go Back

Full Screen

Close

Quit

- parameter #2: $\boldsymbol{\Sigma}_j = \frac{\sum_{i=1}^{N} \gamma_j(\mathbf{x}_i).(\mathbf{x}_i - \boldsymbol{\mu}_j)(\mathbf{x}_i - \boldsymbol{\mu}_j)^T}{N_j}$
  probability (responsibility)-weighted Covariance
- parameter #3: π_j: $\frac{\partial \text{log-lh}}{\partial \pi_j} = \sum_{i=1}^{N} \frac{(-1)\mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)} = 0$
  To modify the objective function, regularisation

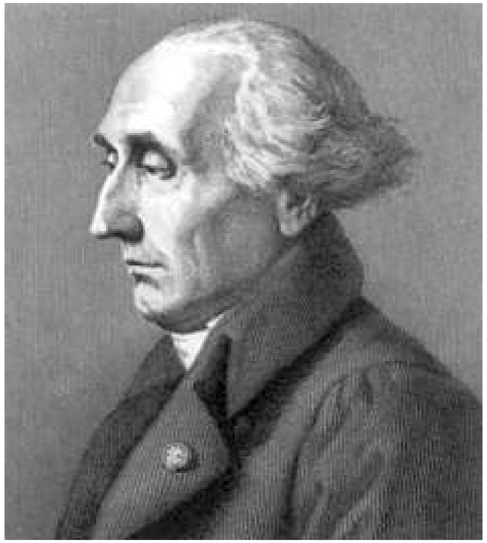
$$E = \log p(\mathbf{X}|\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Sigma}) + \lambda\left(\sum_{j=1}^{K} \pi_j - 1\right)$$

1. $\frac{\partial E}{\partial \boldsymbol{\mu}_j} = 0$: Doesn't affect the previous estimate
2. $\frac{\partial E}{\partial \boldsymbol{\Sigma}_j} = 0$: Doesn't affect the previous estimate
3. $\frac{\partial E}{\partial \pi_j} = 0$: Hope to get some non-trivial solution
4. $\frac{\partial E}{\partial \lambda} = 0$: $\sum_{j=1}^{K} \pi_j - 1 = 0$, the constr to be imposed

For (3) above, $\sum_{i=1}^{N} \frac{(+1)\mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)} + \lambda = 0$. Mult by π_j :

$\sum_{i=1}^{N} \frac{\pi_j \mathcal{N}(\cdot)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\cdot)} = -\lambda \pi_j \implies -\sum_{i=1}^{N} \gamma_j(\mathbf{x}_i) = \lambda \pi_j \implies$
$-\sum_i(\sum_j \gamma_j(\mathbf{x}_i)) = \lambda(\sum_j \pi_j) \implies \lambda(1) = -\sum_i(1) \implies$
$\lambda = -N$. Put back: $\pi_j \lambda = -N_j \implies \pi_j = N_j/N$

Intertwined Histories



J.-L. Lagrange
[1736-1813]

A. Lavoisier
[1743-1794]

J.-B. J. Fourier
[1768-1830]

https://upload.wikimedia.org/wikipedia/commons/1/19/Lagrange_portrait.jpg

https://upload.wikimedia.org/wikipedia/commons/4/44/Lavoisier-statue.jpg

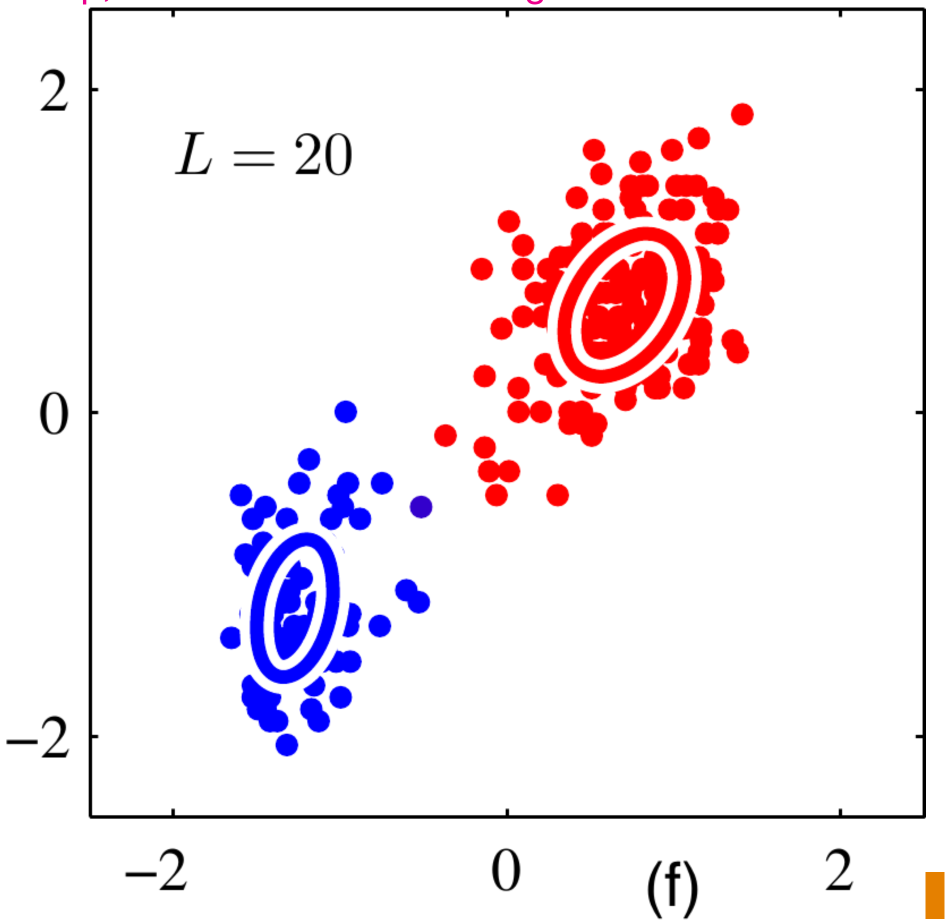
https://upload.wikimedia.org/wikipedia/commons/a/aa/Fourier2.jpg

EM for Gaussian Mixtures

- Given: $\mathbf{X} = \{\mathbf{x}_1, \ldots \mathbf{x}_N\}$, a set of N observations
- Model: $p(\mathbf{X}|\ \boldsymbol{\pi},\ \boldsymbol{\mu},\ \boldsymbol{\Sigma}) = \prod_{i=1}^{N} \sum_{j=1}^{N} \pi_j \mathcal{N}(\mathbf{x}_i|\ \boldsymbol{\mu}_j,\ \boldsymbol{\Sigma}_j)$
- Parameters: 3 sets of parameters: $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}$
- Initialisation: $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}$: Estimate initial value of log-lh
- E-step (Expectation): $\gamma_j(\mathbf{x}_i) = \frac{\pi_j \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}$: Eval responsibilities using current param values
- M-step (Maximisation): Re-estimate params using current $\gamma_j(\mathbf{x}_i)$'s. $N_j = \sum_{i=1}^{N} \gamma_j(\mathbf{x}_i)$ Estimation:
  - $\boldsymbol{\mu}_j^{new} = (1/N_j) \sum_{i=1}^{N} \gamma_j(\mathbf{x}_i).\mathbf{x}_i$
  - $\boldsymbol{\Sigma}_j^{new} = (1/N_j) \sum_{i=1}^{N} \gamma_j(\mathbf{x}_i).(\mathbf{x}_i - \boldsymbol{\mu}_j^{new})(\mathbf{x}_i - \boldsymbol{\mu}_j^{new})^T$
  - $\pi_j^{new} = N_j/N$
- Eval log-lh: $\log p(\mathbf{X}|\ \boldsymbol{\pi},\ \boldsymbol{\mu},\ \boldsymbol{\Sigma}) = \sum_{i=1}^{N} \log \sum_{j=1}^{K} \pi_j \mathcal{N}(\cdot)$; Check for convg of log-lh or params. Else, E-step

EM for Gaussian Mixtures

[C. M. Bishop, Pattern Recognition and Machine Learning]



- Initialisation: Use K-Means to initialise: $\boldsymbol{\mu}$ (sample means), $\boldsymbol{\Sigma}$ (sample Covs), $\boldsymbol{\pi}$ (rel. proportions)

Home Page

Title Page

Contents

Page 26 of 32

Go Back

Full Screen

Close

Quit

Stauffer-Grimson BG Subtraction

- The colour/grey value at a pixel is given by a dis-trn: modelled as a mixture of adaptive Gaussians.

$$p(\mathbf{x}) \triangleq \sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$$ $\mathbf{x}$: $[r\ g\ b]^T$/grey value

https://upload.wikimedia.org/wikipedia/en/4/4b/Fifty-Gray-poster.jpg

  - K: 3-5, empirically
  - mixture: multiple entities can appear
  - adaptive: with lighting conditions
  - Given $\{\mathbf{x}^1, \ldots \mathbf{x}^t, \ldots \mathbf{x}^\tau\}$: history of colour/grey values at a pixel $t \in \{1, \tau\}$
- Heuristic at a pixel: to decide which Gaussians most likely to contribute to the background
- Pixels not matching BG Gaussians: foreground
- FG pixels grouped: 2-D ConnComp/K-Means

Home Page

Title Page

Contents

Page 27 of 32

Go Back

Full Screen

Close

Quit

Model Adaptation

- An on-line K-Means approx: to update $\mathcal{N}(\cdot)$s
  - Why? Stationary pixel process (time-invariant)? EM for time window (costly) Lighting & scene changes reduce past dependence
  - Either: matches one of the K Gaussians $\sim 2.5\sigma$
  - Or: doesn't match any Gaussians $\sim 2.5\sigma$
- If match with one of the K Gaussians $\sim 2.5\sigma$
  - Update the 3 sets of parameters $\boldsymbol{\mu}_j$, $\boldsymbol{\Sigma}_j$, π_j:
  - $\boldsymbol{\mu}_j$, $\boldsymbol{\Sigma}_j$: linear combo of old & new evidence:
    - $\boldsymbol{\mu}_j^{\tau+1} = (1-\rho)\boldsymbol{\mu}_j^{\tau} + \rho \mathbf{x}^{\tau+1}$
    - $\boldsymbol{\Sigma}_j^{\tau+1} =$ $(1-\rho)\boldsymbol{\Sigma}_j^{\tau} + \rho(\mathbf{x}^{\tau+1} - \boldsymbol{\mu}_j^{\tau+1})^T(\mathbf{x}^{\tau+1} - \boldsymbol{\mu}_j^{\tau+1})$
    - Assume diagonal Cov, ind & same variances

- – $* \rho \stackrel{\triangle}{=} \alpha \mathcal{N}(\mathbf{x}^{\tau+1} | \boldsymbol{\mu}_j^{\tau}, \boldsymbol{\Sigma}_j^{\tau})$
    - $*$ α: learning rate. $\alpha = 0 \implies$ no learning, $\rho = 0$
    - $*$ $\boldsymbol{\mu}_j^{\tau+1} = \boldsymbol{\mu}_j^{\tau}$; $\boldsymbol{\Sigma}_j^{\tau+1} = \boldsymbol{\Sigma}_j^{\tau}$
  - – Prior weights of *all* Gaussians adjusted:
  - – $\pi_j^{\tau+1} = (1-\alpha)\pi_j^{\tau} + \alpha \delta_{j,top\ match}^{\tau+1}$
    - $*$ $\delta_{j,top\ match}^{\tau+1} = 1$: matching Gaussian, 0 ow
  - – Renormalise all weights $\pi_j^{\tau+1}$ (only if > 1 best. Grey levels: 2 best equidistant on each side)
    - $*$ $\alpha = 0 \implies \pi_j^{\tau+1} = \pi_j^{\tau}$: no learning
    - $*$ $\alpha = 1 \implies \pi_j^{\tau+1} = 1$: matching Gauss'n, 0 ow
- If $\mathbf{x}^{\tau+1}$ doesn't match any $\sim 2.5\sigma \implies$ something new coming up at this pixel, needs to be put in
  - – Least prob ($\sim \pi/\sigma$) replaced with a new one
  - – New one: $\boldsymbol{\mu}_j^{\tau+1} = \mathbf{x}^{\tau+1}$, $\boldsymbol{\Sigma}_j$ = high, π_j = low

Home Page

Title Page

Contents

Page 29 of 32

Go Back

Full Screen

Close

Quit

Background Model Estimation

- Heuristic: The Gaussian with the most supporting prior evidence and least variance $\equiv$ background
- Order the Gaussians by π_j/σ_j (greater prob, less variability: less change, suggests background)
- First B distributions: background
- $B = argmin_b \sum_{j=1}^{b} \pi_j > Thresh$, (scene dependent)
  - $Thresh$: scene dependent (rel % of image: bg)
  - Low $Thresh$: unimodal bg, use most prob model
  - High $Thresh$: multimodal: repetitive bg motion (leaves, flag): > 1 colour in bg model accepted

The Algo In Perspective

- (+): Each pixel (2.5σ) treated diff, adaptively
- (+): Objects allowed to be a part of the bg, without changing the existing bg model
- (-): Cannot handle sudden lighting changes
- (-): Initialisation of the Gaussians is important
- (-): Many parameters need to be selected smartly

- The $\sim 2.5\sigma$ threshold: per pixel, per distributions
  - Diff regions, diff lighting: objects in shaded regions not much noise as in lighted regions
  - Surveillance cam: diff lighting in FOV, diff times

Results with Diff Parameters

- Reversing vehicle (also entering the shadows), loitering loafers, trees swaying in the wind