Home Page

Title Page

Contents

Page 19 of 32

Go Back

Full Screen

Close

Quit

Maximising the log-likelihood

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

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

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

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

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

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

How to maximise the log-likelihood?

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

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

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

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

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

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

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

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

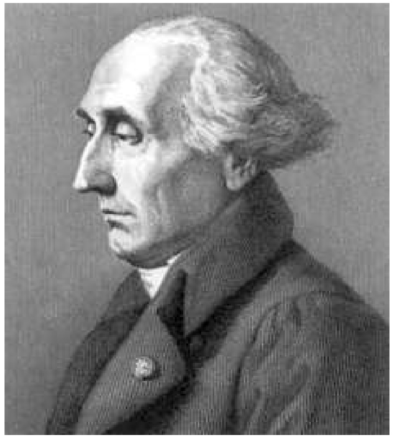
$$E = \log p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \mathbf{\Sigma}) + \lambda\left(\sum_{j=1}^{K} \pi_j - 1\right)$$

1. $\frac{\partial E}{\partial \boldsymbol{\mu}_j} = 0$: Doesn't affect the previous estimate
2. $\frac{\partial E}{\partial \mathbf{\Sigma}_j} = 0$: 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, $\sum_{i=1}^{N} \frac{(+1)\mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j, \mathbf{\Sigma}_j)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j, \mathbf{\Sigma}_j)} + \lambda = 0$. Mult by π_j :

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

