Exercise session 3

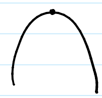
keskiviikko 4. marraskuuta 2020 14.03

Exercise 1.

a) $\frac{\partial}{\partial \theta} \ln (L(\theta ; y))$ $\qquad \hat{\theta}$

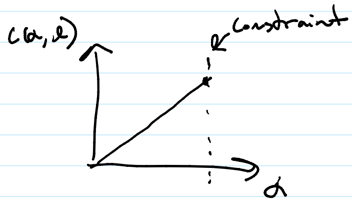
likelihood

$\frac{\partial^2}{\partial \theta^2} < 0$



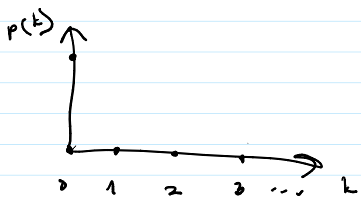
b) $\frac{\partial}{\partial \ell} l(\alpha, \ell ; y) = 0 \qquad \hat{\ell}$

$\frac{\partial}{\partial \alpha} \Rightarrow c$



$\alpha < y_i$

c)



n_0 := "number of $y_i = 0$" $\qquad n$:= "number of samples"

$n - n_0$

$$L(\theta ; y) = \prod_{i=1}^{n_0} f(0 ; \theta, \ell) \prod_{j=1}^{n-n_0} f(y_j ; \theta, \ell)$$

non-zero

Exercise 2

$$\frac{\partial^2}{\partial c^2} l(c)$$

$$Var[\hat{\theta}] \geq \frac{1}{-\mathbb{E}[\frac{\partial}{\partial c^2} l(c)]}$$

Exercise 3.

Google: exponential integration tricks (Hested)

$$\mathbb{E}[X] = \int_0^\infty x f(x)\, dx$$

$$\mathbb{E}[g(x)] = \int_0^\infty g(x) f(x)\, dx$$

$$\boxed{\int_0^\infty \lambda x^n e^{-\lambda x} dx = \mathbb{E}[X^n] = \frac{n!}{\lambda^n} \qquad f(x;\lambda) = \lambda e^{-\lambda x}}$$

$$\text{Var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2$$

$$\text{Var}\left[\sum_{i=1}^{N} x_i\right] = N \,\text{Var}[x]$$

i.i.d

Exercise 4:

$$\hat{\theta}(k) = [X^T w(k) X]^{-1} X^T W(k) y$$

update step

$$W(k) = \begin{bmatrix} w_{11} & 0 & 0 \\ 0 & w_{22} & \\ 0 & & \ddots \end{bmatrix}$$

$$w_{ii} = \psi(\tilde{y}_i) \qquad \tilde{y}_i = y_i - \vec{x}\,\hat{\theta}(k-1)$$

$$\tilde{y} = y - X\hat{\theta}(k-1)$$

diag(w)

$$\psi(3\sigma) = 0$$

Exercise 5.

$a_1, b_1, a_2, b_2 \rightarrow$ choose to generate data

$x(n) \rightarrow$ choose any signal e.g. sinewave or line

Latent variables $\rightarrow$ Problem for ML method

$$r_k(n) = a_k x(n) + b_k - y(n)$$

Probability to sample from model k

$$p(y(n);\theta) = \sum_{k=1}^{2} \pi_k \, p\left(a_k x(n) + b_k - y(n);\ \sigma^2\right)$$

$$p(y(n);\theta) = \sum_{k=1}^{K} \pi_k \, p(a_k x(n) + b_k - y(n) ; \sigma^2)$$

$\{\pi_k, a_k, b_k\}_{k=1}^{2}$

$$= \sum_{k=1}^{2} \pi_k \, p(r_k(n) ; \sigma^2)$$

$$L(\theta \mid y, Z) = \prod_{n=1}^{N} \prod_{k=1}^{K} \left[\pi_k \, p(r_k(n) ; \sigma^2) \right]^{z_k(n)}$$

Matrix where element $[Z]_{nk}$ is 1 if $y(n)$ was sampled from model k and otherwise zero

Do the same steps as in exercise session 2.

$$E_Z[L(\theta \mid y, Z)] = \underbrace{\sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_k(n) \left[\log \pi_k + \log p(r_k(n) ; \sigma^2) \right]}$$

Prove that is WLS objective withe respect to a_k, b_k

$\pi_1 = \pi_2$ can be assumed known since 64 samples obtained from both models

Calculating weights:

$w_k(n)$ is the probability that sample $y(n)$ was drawn from model k

$\Rightarrow p(a_k, b_k \mid r_k(n))$ or using the latent variable notation $p(z_k(n) = 1 \mid r_k(n))$

$\Rightarrow$ Bayes rule

EM-algorithm:

Initialize $a_k, b_k \ \forall k$ (Guess or sample from prior distribution)

E-step) Calculate weights

M-step) Minimise the WLS objective

Initialization via weights is also possible if applicable prior distribution is available (then M-step after initialization)

How to generate the data set in exercise 5?

- Since both models generate 64 samples
  $\Rightarrow$ y used in simulation has size 128.
  - E.g. create y_1 and y_2 where $y_k \in \mathbb{R}^{128 \times 1}$ and sample for each $y(n)$ from either $y_1(n)$ or $y_2(n)$ such that (at least approximately) equal number of samples are obtained from each model)

- You may want to select a_k, b_k and $x(n)$ such that the signals overlap at some point and you can also see the additive noise (to make things interesting)