# Examples of MLE estimation.

## example 1 : One dimensional Gaussian Case

$$p_{x|y=i} = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{ - \frac{(x - \mu_i)^2}{2\sigma_i^2} \right\}$$

$x, \mu_i, \sigma_i \in \mathbb{R}$

Consider we have $n$-examples & obtaining the log likelihood.

$$= \sum_{j=1}^{n} \ln p_{x|y=i}(x_j)$$

$$= \sum_{j=1}^{n} \left[ -\ln \sigma_i - 0.5 \ln 2\pi - \frac{(x_j - \mu_i)^2}{2\sigma_i^2} \right]$$

$$= -n \ln \sigma_i - 0.5n \ln 2\pi - \sum_{j=1}^{n} \frac{(x_j - \mu_i)^2}{2\sigma_i^2}$$

To maximize the log likelihood, we equate the partial derivative to zero

$$\nabla \sigma_i = \frac{-n}{\sigma_i} + \frac{1}{\sigma_i^3} \sum_{j=1}^{n} (x_j - \mu_i)^2 = 0$$

$$(\sigma_i^2)^* = \frac{1}{n} \sum_{j=1}^{n} (x_j - \mu_i)^2$$

$$\nabla \mu_i = \frac{1}{\sigma_i^2} \sum_{j=1}^{n} (x_j - \mu) = 0$$

$$\mu_i^* = \frac{1}{n} \sum_{j=1}^{n} x_j$$

Consider the multidimensional Gaussian density

$$p_{X|Y=i}(x) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp\left\{ \frac{-1}{2} (x-\mu_i)^T \Sigma^{-1} (x-\mu_i) \right\}$$

$$x, \mu_i \in \mathbb{R}^d$$

$\Sigma \in \mathbb{R}^{d \times d}$ : Co-Variance matrix.

Considering the "n" samples & taking log likelihood.

$$= \sum_{j=1}^{n} \ln p_{X|Y=i}(x)$$

$$= \sum_{j=1}^{n} \ln \left[ \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp \left\{ -\frac{1}{2} (x_j - \mu_i)^T \Sigma_i^{-1} (x_j - \mu_i) \right\} \right]$$

$$= \sum_{j=1}^{n} \left[ -\frac{1}{2} \ln \left( (2\pi)^d |\Sigma| \right) - \frac{1}{2} (x_j - \mu)^T \Sigma^{-1} (x_j - \mu) \right]$$

now need to take the partial derivatives wrt parameters and equate to zero.

$$\nabla \mu = \sum_{j=1}^{n} \Sigma^{-1} (x_j - \mu) = 0$$

$$\mu^* = \frac{1}{n} \sum_{j=1}^{n} x_j$$

now we need to take the derivative wrt $\Sigma$ out log likelihood.

$$= \sum_{j=1}^{n} -\frac{1}{2} \ln \left( (2\pi)^d |\Sigma| \right) - \frac{1}{2} (x_j - \mu)^T \Sigma^{-1} (x_j - \mu)$$

$$= -\frac{n}{2} d \ln 2\pi - \frac{n}{2} \ln(|\Sigma|) - \sum_{j=1}^{n} \frac{1}{2} (x_j - \mu)^T \Sigma^{-1} (x_j - \mu)$$

$$\Sigma^{-1} = \Lambda$$

$$z_j = (x_j - \mu)$$

$$= \frac{-nd}{2} \ln 2\pi - \frac{n}{2} \ln(|\Lambda|) - \frac{1}{2} \sum_{j=1}^{n} z_j^T \Lambda z_j$$

we need derivatives of the form

$$\frac{\partial}{\partial A} z^T A z \quad \& \quad \frac{\partial}{\partial A} \ln(|A|)$$

Some useful matrix identities.

$$\text{Trace}(AB) = \text{Trace}(BA)$$

$$\text{Trace}(ABC) = \text{Trace}(CAB) = \text{Trace}(BCA)$$

$$\frac{\partial}{\partial A} \text{Trace}(A^T B) = B, \quad \frac{\partial}{\partial A} \text{Trace}(AB) = B^T$$

$$z^T A z = \text{Trace}(z^T A z)$$

$$= \text{Tr}(z z^T A)$$

$$= \text{Trace}(A z z^T)$$

$$\therefore \frac{\partial}{\partial A} z^T A z = \frac{\partial}{\partial A} \text{Trace}(A z z^T)$$

$$= (z z^T)^T$$

$$= z z^T$$

now we need other derivative $\frac{\partial}{\partial A} \ln(|A|)$

There is a indentity

$$\beta \in \mathbb{R}$$

$$\frac{\partial}{\partial \beta} \ln(|A|) = \text{Trace}\left(A^{-1} \frac{\partial}{\partial \beta} A\right)$$

$$\frac{\partial}{\partial A_{ij}} \ln(|A|) = \text{Trace}\left(A^{-1} \frac{\partial}{\partial A_{ij}} A\right)$$

$\frac{\partial A}{\partial A_{ij}}$ is a matrix with 1 in position $(i,j)$

and zeros else where.

$$\left(A^{-1}\frac{\partial}{\partial A_{ij}}A\right)_{kl} = \sum_{s} A^{-1}_{ks}\left(\frac{\partial}{\partial A_{ij}}A\right)_{sl}$$

$$= \begin{cases} 0 & \text{if } \quad l \neq j \\ A^{-1}_{ki} & \text{if } \quad l = j \end{cases}$$

$$\text{Trace}\left(A^{-1}\frac{\partial A}{\partial A_{ij}}\right) = \left(A^{-1}\frac{\partial A}{\partial A_{ij}}\right)_{jj}$$

$$= A^{-1}_{ji}$$

$$\therefore \quad \frac{\partial}{\partial A}\ln(|A|) = (A^{-1})^{T}$$

now putting these together then

take the partial derivate & equate to zero

$$\frac{n}{2}\Lambda^{-1} - \frac{1}{2}\sum_{j=1}^{n} z_j^{T} z_j = 0$$

$$\Sigma = \Lambda^{-1} = \frac{1}{n}\sum_{j=1}^{n}(x_j-\mu)(x_i-\mu)^{T}$$

example 2 : Generalized Discrete RV.

$Z$ take values $a_1, \dots, a_m$

with probabilites $p_1, \dots, p_m$

Given data is form of iid realizations of this RV.

we need to estimate $p_i$

the parameters should satisfy

$$p_i \geq 0 \quad \& \quad \sum_{i=1}^{m} p_i = 1$$

For our estimation, we represent the DRV $Z$

by a "m" dimensional vector RV. : $V$

→ $V$ takes $m$ possible values.

$$[1, 0, \dots 0]^T, \ [0, 1, \dots, 0]^T, \ \dots, \ [0, 0, \dots, 1]^T$$

$$V = [v^1, \dots, v^m]$$

$$v^i \in \{0, 1\} \qquad \sum_{j=1}^{m} v^j = 1$$

One hot Representation.

$$p(v) = p_1^{v^1} * p_2^{v^2} * \cdots * p_m^{v^m}$$

$$= \prod_{i=1}^{m} p_i^{v^i}$$

Given iid Data

$$D = \{ v_1, v_2, \ldots, v_n \}$$

$$v_i = [ v_i^1, v_i^2, \ldots, v_i^m ]^T$$

$$v_i^j \in \{0, 1\} \qquad \sum_{j=1}^{m} v_j^j = 1$$

$$p(v_i) = \prod_{j=1}^{m} p_j^{v_i^j}$$

Considering all data points & taking log

$$= \sum_{i=1}^{n} \ln p(v_i)$$

$$= \sum_{i=1}^{n} \ln \left( \prod_{j=1}^{m} p_j^{v_i^j} \right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \omega_i^j \ln(p_j)$$

The above formulation is an Un Constrained maximization.

→ we need to perform MLE of the parameters over the constraints

$$p_i \geq 0 \quad \& \quad \sum_{j=1}^{m} p_j = 1$$

∴ the constrained optimization problem is

$$\max_{p_j} \sum_{i=1}^{n} \sum_{j=1}^{m} \omega_i^j \ln p_j$$

$$\text{s.t} \quad \sum_{j=1}^{m} p_j = 1$$

The lagrangian for this problem is given by

$$\sum_{i=1}^{n} \sum_{s=1}^{m} \omega_i^s \ln p_s + \lambda \left(1 - \sum_{s=1}^{m} p_s\right)$$

now calculate the partial derivative of the lagrangian & equate it to zero.

$$\nabla p_j = \sum_{j=1}^{n} \frac{\omega_i^j}{p_j} - \lambda = 0 \qquad j = 1, \ldots, m$$

$$p_j = \frac{1}{\lambda} \sum_{i=1}^{n} \omega_i^j$$

now we use the constraint $\sum_{j=1}^{m} p_j = 1$

$$\sum_j p_j = \sum_{j=1}^{m} \frac{1}{\lambda} \sum_{i=1}^{n} \omega_i^j = 1$$

$$\lambda = \sum_{j=1}^{m} \sum_{i=1}^{n} \omega_i^j = \sum_{i=1}^{n} \sum_{j=1}^{m} \omega_i^j = n$$

$$\therefore \; p_j^* = \frac{1}{n} \sum_{i=1}^{n} \omega_i^j , \; j = 1, \ldots, m \quad \blacksquare$$