Exercise 2

Maximum likelihood estimation.

Gaussian distribution expressed in terms of the parameters μ and Σ.

μ: nx1 vector

Σ: nxn symmetric matrix

We have the following form for the density function:

$$p(x|\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}(\Sigma)^{\frac{1}{2}}} exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}.$$

Where x is a vector in $\mathcal{R}^n$.

In our problem we want to estimate the parameter Σ. Let us suppose that D is our dataset.

$$l(\mu,\Sigma|D) = -\frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu).$$

Taking the derivative with respect to μ is straightforward:

$$\frac{\partial l}{\partial \mu} = \sum_{\iota=1}^{N}(x_i-\mu)^T\Sigma^{-1}.$$

and setting to zero we obtain a pleasant result:

$$\hat{\mu} = \frac{1}{N}\sum_{\iota=1}^{N} x_i.$$

We use this form in order to calculate the mean of the pdf.

We now turn to the problem of computing the maximum likelihood estimate of the covariance matrix. Letting $l(\Sigma|D)$ denote the terms in the likelihood that are a function of Σ, we obtain:

$$l(\Sigma|D) = -\frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu).$$

We need to take the derivative with respect to Σ and set to zero. Rewriting the log-likelihood using "track trace"

$$l(\Sigma|D) = -\frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$$

$$= \frac{N}{2}\log|\Sigma^{-1}| - \frac{1}{2}\sum_n tr[(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)]$$

$$= \frac{N}{2}\log|\Sigma^{-1}| - \frac{1}{2}\sum_n tr[(x_i-\mu)^T(x_i-\mu)\Sigma^{-1}]$$

We now take the derivative with respect to the matrix Σ^{-1} and using 1) $\frac{\partial}{\partial A}\log|A| = A^{-T}$, 2) $\frac{\partial}{\partial A}Tr[AB] = \frac{\partial}{\partial A}Tr[BA] = B^T$.

$$\frac{\partial l}{\partial \Sigma^{-1}} = \frac{N}{2}\Sigma - \frac{1}{2}\sum_n (x_i-\mu)^T(x_i-\mu).$$

Finally, setting to zero yields the maximum likelihood estimator:

$$\hat{\Sigma}_{ML} = \frac{1}{N}\sum_n (x_n-\hat{\mu}_{ML})^T(x_n-\hat{\mu}_{ML}).$$

In our case, the covariance matrices are diagonal, with all diagonal elements equal.

$$\Sigma = \begin{bmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{bmatrix}$$

$$l(\Sigma|D) = -\frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_n (x_i-\mu)^T\Sigma^{-1}(x_i-\mu) + const$$

$$= -\frac{N}{2}\log\sigma^{2k} - \frac{1}{2}\sum_n (x_i-\mu)^T\Sigma^{-1}(x_i-\mu) + const$$

$$= -\frac{N}{2}\log\sigma^{2k} - \frac{1}{2}\sum_n (x_i-\mu)^T I\sigma^{-1}(x_i-\mu) + const$$

$$= -Nk\log\sigma - \frac{\sigma^{-2}}{2}\sum_n (x_i-\mu)^T(x_i-\mu) + const$$

$$\frac{dl}{d\sigma} = -Nk - \frac{1}{\sigma^2}\sum_n (x_i-\mu)^T(x_i-\mu) = 0$$

$$\sigma^2_{ML} = \frac{1}{Nk}\sum_n (\vec{x_i}-\vec{\mu})^T(\vec{x_i}-\vec{\mu})$$