coursera

Menu

MLE for Gaussian population

In the videos, you got an intuition of what the Maximum Likelihood Estimation (MLE) should look like for the mean and variance of a Gaussian population.

In this reading item, you will learn the derivation of both results.

Mathematical formulation

Suppose you have n samples $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ from a Gaussian distribution with mean μ and variance σ^2. This means that $X_i \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$.

If you want the MLE for μ and σ the first step is to define the likelihood. If both μ and σ are unknown, then the likelihood will be a function of these two parameters. For a realization of X, given by $x = (x_1, x_2, \ldots, x_n)$:

$$\begin{aligned} L(\mu, \sigma; \boldsymbol{x}) &= \prod_{i=1}^n f_{X_i}(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}} \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^n \sigma^n} e^{-\frac{1}{2}\frac{\sum_{i=1}^n (x_i-\mu)^2}{\sigma^2}} \end{aligned}$$

Now all you have to do is find the values of μ and σ that maximize the likelihood $L(\mu, \sigma; \boldsymbol{x})$.

You might remember from the calculus course that one way to do this analytically is by taking the derivative of the Likelihood function and equating it to 0. The values of μ and σ that make the derivative zero, are the extreme points. In particular, for this case, they will be maximums.

Taking the derivative of the likelihood is a cumbersome procedure, because of all the products involved. However, there is a nice trick you can use to simplify things. Note that the logarithm function is always increasing, so the values that maximize $L(\mu, \sigma; \boldsymbol{x})$ will also maximize its logarithm. This is the **log-likelihood**, and it is defined as

$$\ell(\mu, \sigma) = \log(L(\mu, \sigma; \boldsymbol{x}))$$

The logarithm has the property of turning a product into a sum, this means that