Maximum Likelihood parameter estimation of Multivariate Gaussian

Chiranjib Bhattacharyya

March 15, 2020

Abstract

This note can be used as reading material for the topic of parameter estimation of Multivariate Gaussian discussed in a lecture for the course E0-270.

Let $\mu \in \mathbb{R}^d$ and $C \in \mathbb{R}^{d\times d}$ be the mean and covariance matrix associated with a d-dimensional Gaussian random variable, X. The pdf of X at x will be denoted by $N(x|\mu, C)$ and will be defined by

$$P(X = x|\mu, C) = \frac{1}{\left(\sqrt{2\pi}\right)^d |C|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^\top C^{-1}(x-\mu)} = N(x|\mu, C)$$

where $|C|$ denotes the determinant of C.

Assume that we have access to a dataset $D = \{x_1, x_2, \ldots, x_n\}$ consisting of n i.i.d. realizations of X. Define

$$m = \frac{1}{n}\sum_{i=1}^n x_i, \quad S = \frac{1}{n}\sum_{i=1}^n (x_i - m)(x_i - m)^\top$$

We will further assume that both C and S are invertible, and hence both of them are positive definite.

The learning problem is to estimate μ, C from D by maximizing the likelihood function

Claim: Let μ, C, m, S be defined earlier. The maximum likelihood estimates of μ and C are given by $\mu = m, C = S$.

Proof. By definition

$$\mu^*, C^* = \underset{\mu, C}{\operatorname{argmax}}\, LL(\mu, C)$$

where LL denotes the log-likelihood of dataset D,

$$LL(\mu, C) = \sum_{i=1}^n \log N(x_i|\mu, C) = -n\log\left(\sqrt{2\pi}\right)^d - \frac{1}{2}\log|C| - \frac{1}{2}\sum_{i=1}^n (x_i-\mu)^\top C^{-1}(x_i-\mu)$$

Recall that for any twice continuously differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ with negative definite Hessian

$$z^* = \underset{z\in\mathbb{R}^d}{\operatorname{argmin}}\, f(z) \quad \text{iff} \quad \nabla_z f(z^*) = 0$$

This can be used to deduce that for any square symmetric positive definite matrix C

$$\mu(C) = \underset{\mu}{\operatorname{argmax}}\, LL(\mu, C) = m.$$

The deduction follows by noting that by direct computation for any fixed C the Hessian of $LL(\mu, C)$ when evaluted with respect to μ evaluates to $-nC$. Since C is positive definite the Hessian is negative definite. To find the maximum we need to evaluate the stationary points of LL which is obtained by taking the derivative of LL with respect to μ namely,

$$C^{-1} \sum_{i=1}^{n} (x_i - \mu) = 0$$

which is satisfied by $\mu = m$, and the deduction is complete. Consequently

$$LL(\mu, C) \leq LL(m, C) \tag{1}$$

Further recall that

$$Trace(ABC) = Trace(BCA)$$

where A, B, C are suitably defined matrices. This allows rewriting the term as

$$(x - m)^\top C^{-1} (x - m) = nTrace(C^{-1} S)$$

and due to (1)

$$LL(\mu, C) \leq LL(m, C) = -\frac{n}{2} \left(\log(2\pi)^d + \log |C| + Trace(C^{-1} S) \right)$$

Putting $C = S$ gives

$$LL(m, S) = -\frac{n}{2} \left(\log(2\pi)^d + \log |S| + d \right)$$

where $trace(I) = d$ was used. To complete the proof we need to show that

$$LL(m, S) \geq LL(m, C)$$

In other words we need to show that

$$LL(m, S) - LL(m, C) \geq 0$$

Direct substitution yields

$$LL(m, S) - LL(m, C) = -\frac{n}{2} \left(d + \log |S| - \log |C| - Trace(C^{-1} S) \right)$$

$$LL(m, S) - LL(m, C) = -\frac{n}{2} \left(d + \log |C^{-1} S| - Trace(C^{-1} S) \right) \tag{2}$$

We now use the relationship between Trace and Determinants with the eigenvalues. Since C and S are both positive definite it follows that $C^{-1}S$ is also positive definite. Recall that for any square symmetric positive definite matrix $d \times d$ matrix A with real entries

$$trace(A) = \sum_{i=1}^{d} \lambda_i, \quad \log |A| = \sum_{i=1}^{d} \log \lambda_i$$

where λ_i are the eigenvalues of A. Let λ_i be the eigenvalues of the matrix $C^{-1}S$ and so (2) writes as

$$\begin{aligned} LL(m,S) - LL(m,C) &= -\frac{n}{2}\left(d + \sum_{i=1}^{d} \log \lambda_i - \sum_{i=1}^{d} \lambda_i\right) \\ &= -\frac{n}{2}\left(\sum_{i=1}^{d} g(\lambda_i)\right) \end{aligned}$$

where $g : (0, \infty) \to \mathbb{R}$ and $g(z) = 1 + \log z - z$. It is easy to see that $g(z) \leq 0$ for all $z \geq 0$ and consequently

$$LL(m,S) - LL(m,C) \geq 0$$

Thus it follows from (1)

$$LL(\mu, C) \leq LL(m,S)$$

for any $\mu \in \mathbb{R}^d$ and any positive definite square symmetric matrix C.

$\square$