

Assignment 3 : CS215

Akshat Chugh : 170050019

Satvik Ambati : 170050101

October 21, 2018

3. We know,

$$P(x; \mu, C) = \frac{1}{\sqrt{(2\pi)^2 |C|}} \exp(-0.5(x - \mu)^T C (x - \mu))$$

where x is a 2×1 vector such that $x_1^2 + x_2^2 = r^2$. Here μ and C are the mean and the covariance matrix respectively and $|C|$ is the determinant of C .

Therefore, the likelihood function is given by $P(x; \mu, C)$.

So, the log likelihood for an individual data-point is,

$$LL = -0.5(x - \mu)^T C (x - \mu) - 0.5 \ln(|C|) - \ln(2\pi)$$

Now, for MLE,

$$\frac{\partial LL}{\partial C} = \frac{\partial LL}{\partial \mu} = 0$$

Thus, using derivatives calculated online,

$$\frac{\partial LL}{\partial C} = \sum_{j=1}^n 0.5(C^T)^{-1}((x^j - \mu)(x^j - \mu)^T(C^T)^{-1} - I) = O \quad (1)$$

$$\frac{\partial LL}{\partial \mu} = \sum_{j=1}^n 0.5(x^j - \mu)^T(C^{-1} + (C^{-1})^T) = O \quad (2)$$

where x^j is the j th data-point, n is the number of data-points.

Now, since C is a covariance matrix, $C = C^T$. Thus, by right-multiplying both sides by C in (2), we get

$$\begin{aligned} \sum_{j=1}^n (x^j - \mu)^T &= O \\ \implies \hat{\mu}^{ML} &= \frac{\sum_{j=1}^n x^j}{n} \end{aligned} \quad (3)$$

Similarly, by left-multiplying by C in (1), we get

$$\hat{C}^{ML} = \frac{\sum_{j=1}^n (x^j - \hat{\mu}^{ML})(x^j - \hat{\mu}^{ML})^T}{n} \quad (4)$$

Now, we can take $x = \begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix}$

Since we have to consider large n , we can take average of $\cos \theta$ and $\sin \theta$ over $[0, 2\pi)$, which is 0 for both. Thus, $\hat{\mu}^{ML} = O$.

Now, for \hat{C}^{ML} , we have to take average of all four entries. This ends up being,

$$\hat{C}^{ML} = r^2 \text{avg}_{[0,2\pi)} \begin{pmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{pmatrix}$$

By integrating to get average, we get,

$$\hat{C}^{ML} = r^2 \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} = \frac{r^2}{2} I_2$$

Now, with these values of μ and C , the Gaussian is,

$$P(x; \mu, C) = \frac{1}{\pi r^2} \exp \left(-0.5 \begin{pmatrix} x & y \end{pmatrix} \frac{2}{r^2} I_2 \begin{pmatrix} x \\ y \end{pmatrix} \right) = \frac{1}{\pi r^2} \exp \left(-\frac{(x^2 + y^2)}{r^2} \right)$$

The mode of this Gaussian clearly lies at the point where $(x^2 + y^2)$ is minimum, i.e the origin.

Clearly, this Gaussian does not fit the data well, as our data points all lie only on one circle, whereas the Gaussian gives probability for all points as non-zero. Even then, the mode of the Gaussian is not the original circle our points come from. Thus, it is not a good model.

Using MATLAB code in the folder, we get, for $n = 10000, r = 1$, (for one run of the code)

$$\hat{\mu}^{ML} = \begin{pmatrix} -0.0078 \\ -0.0028 \end{pmatrix}$$

$$\hat{C}^{ML} = \begin{pmatrix} 0.5004 & -0.0039 \\ -0.0039 & 0.4995 \end{pmatrix}$$

This matches extremely well with our theoretical values.