Q3) For every ~~d~~ datapoint, $X = \begin{pmatrix} x \\ y \end{pmatrix}$,

$$X^T X = r^2, \text{ or } X = r\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \text{ where}$$

θ is uniformly distributed on $(0, 2\pi)$.

$$P(x_i \mid \mu, C) = \frac{1}{2\pi\sqrt{|C|}} \exp\left(-\frac{(x_i-\mu)^T \cdot C^{-1} \cdot (x_i-\mu)}{2} \right)$$

Log Likelihood $= -\frac{n}{2}\log|C| - \frac{1}{2}\sum_{i=1}^{N}\left((x_i-\mu)^T C^{-1}(x_i-\mu)\right) + c$

$$\frac{\partial LL}{\partial \mu} = 0 \Rightarrow \left(\sum\left((x_i-\hat{\mu})^T \cdot C^{-T} + (x_i-\hat{\mu})^T C^{-1}\right)\right) = 0$$

$\Rightarrow$ [using $d(A^TBC) = C^TB^T d(A^T) + A^TB\,dC$]

$$\Rightarrow \sum (x_i - \hat{\mu})^T \cdot (C^{-1} + C^{-T}) = 0$$

$$(C^{-1}+C^{-T}) \sum (x_i - \hat{\mu}) = 0$$

$$\Rightarrow \sum x_i - N\hat{\mu} = 0$$

$$\Rightarrow \hat{\mu} = \frac{\sum x_i}{N} = \frac{r\sum\begin{pmatrix} \cos\theta_i \\ \sin\theta_i \end{pmatrix}}{N}$$

Now, to calculate $\hat{C} \rightarrow$

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

$$= \frac{n}{2}\log|C^{-1}| - \frac{1}{2}\sum_{i=1}^{N}\left((x_i-\mu)^T C^{-1}(x_i-\mu)\right)$$

Since inverse of a matrix is unique, we can differentiate the expression w.r.t C^{-1}, & set it to 0 to obtain MLE for C.

$$\Rightarrow \frac{\partial \mathcal{L}}{\partial C^{-1}} = 0 \Rightarrow \frac{N}{2}\hat{C} - \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu).(x_i-\mu)^T = 0$$

The first term is obtained by the identity

$$\frac{\partial}{\partial X}\log|X| = X^{-1},$$

& the second by the identity

$$\frac{d}{da}\left(tr(xx^T(Ya+z))\right) = xx^TY$$

[$\because$ $(x_i-\mu)^T.C^{-1}(x_i-\mu) = tr((x_i-\mu)^T C^{-1}(x_i-\mu))$,
since ~~both are~~ LHS is 1×1 in dimension.
further, since $tr(XY) = tr(YX)$,
$tr((x_i-\mu)^T C^{-1}(x_i-\mu)) = tr(C^{-1}(x_i-\mu).(x_i-\mu)^T)$]

$$\Rightarrow \hat{C} = \frac{1}{N}\left(\sum_{i=1}^{N}(x_i-\hat{\mu}).(x_i-\hat{\mu})^T\right)$$

$$\hat{C} = \frac{1}{N} r^2 \sum \begin{pmatrix} (\cos\theta_i - \sum_N \cos\theta_j)^2 & (\cos\theta_i - \sum_N \cos\theta_j).(\sin\theta_i - \sum_N \sin\theta_j) \\ (\cos\theta_i - \sum_N \cos\theta_j).(\sin\theta_i - \sum_N \sin\theta_j) & (\sin\theta_i - \sum_N \sin\theta_j)^2 \end{pmatrix}$$

Mode of Gaussian $\rightarrow$ x for which $P(x|\mu, C)$ is max.

$\Rightarrow \dfrac{dP}{dx} = 0$.

$$\Rightarrow P(x|\mu, C) \cdot \left[\frac{-(x-\mu)^T C^{-T} - (x-\mu)^T C^{-1}}{2} \right] = 0$$

$$\Rightarrow x_{mode} - \mu = 0 \quad [\because (C^{-1} + C^{-T}) \neq 0]$$

$$\Rightarrow \underline{x_{mode} = \mu}$$

Now, as $N \to \infty$,

$$\hat{\mu} = \frac{r}{N} \begin{pmatrix} \sum_{i=1}^{N} \cos\theta_i \\ \sum_{i=1}^{N} \sin\theta_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

[$\because$ θ_i are uniformly distributed RV over $(0, 2\pi)$, & so by law of large numbers, means of $\cos\theta_i$ & $\sin\theta_i \to 0$]

$$\hat{C} = \frac{r^2}{N} \sum \begin{pmatrix} (\cos\theta_i)^2 & \cos\theta_i \sin\theta_i \\ \cos\theta_i \sin\theta_i & (\sin\theta_i)^2 \end{pmatrix}$$

$$= \cancel{r^2 \sum} \; r^2 \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

[similar arguments]

Hence, mode = $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

This means that Gaussian is not a good fit for data, as none of data points lie inside circle, where the probability is highest.

Experimental Values-

N=1000000.

r=2

Mu = (1e-5   2e-5)

C = 2.0000   1e-6
 1e-6   2.0001

These are very close to the theoretical values predicted by the Gaussian kernel.