CS215 Assignment 5
Estimation Transformations

Tathagat Verma 180050111
Neel Aryan Gupta 180050067

PAGE NO.
DATE / /

Question - 3

Let the dataset be $\{x_i\}_{i=1}^{N}$ where each x_i is a 2D vector $\begin{bmatrix} x \\ y \end{bmatrix}$

Following results will be used from Matrix calculus

$$\frac{d}{d\mu}(x-\mu)^T C^{-1}(x-\mu) = 2C^{-1}(x-\mu) = \frac{d}{dx}(x-\mu)^T C^{-1}(x-\mu)$$

In general, $\frac{d}{d\mu}(x-\mu)^T A(x-\mu) = (A+A^T)(x-\mu)$

$$\frac{d}{dC}(x-\mu)^T C^{-1}(x-\mu) = -C^{-T}(x-\mu)(x-\mu)^T C^{-T}$$

$$\frac{d}{dC}\log(|C|) = \frac{1}{|C|}|C| C^{-T} = C^{-T}$$

(i) Likelihood : $\prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^2|C|}} \exp\left(\frac{-1}{2}(x_i-\mu)^T C^{-1}(x-\mu)\right)$

Log likelihood : $\mathcal{L}$

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

For mean estimate,

$$\frac{\partial \mathcal{L}}{\partial \mu} = \frac{1}{2}\sum_{i=1}^{N} 2C^{-1}(x_i-\mu) = 0 \quad \text{(using Matrix calculus)}$$

$$\Rightarrow \sum_{i=1}^{N}(x_i-\mu) = 0$$

$$\Rightarrow \sum_{i=1}^{N} x_i - N\mu = 0$$

$$\Rightarrow \hat{\mu} = \frac{\sum_{i=1}^{N} x_i}{N}$$

PAGE NO.
DATE / /

For covariance estimate,

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

$$\Rightarrow NC^{-T} = C^{-T}\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T C^{-T}$$

$$\Rightarrow N = C^{-T}\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T$$

$$\Rightarrow C^T = \frac{\sum_{i=1}^{N}(x_i-\mu)(x_i-\mu)^T}{N} \qquad [\because (AB)^T = B^T A^T]$$

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

Now, Gaussian $= \frac{1}{\sqrt{(2\pi)^2|\hat{C}|}} \exp\left(\frac{-1}{2}(x-\hat{\mu})^T\hat{C}^{-1}(x-\hat{\mu})\right)$

(ii) For mode, $\frac{\partial G}{\partial x} = 0$ ie. G is max. ie $\log G$ is max.

$$\Rightarrow \frac{\partial(\log G)}{\partial x} = 0$$

$$\Rightarrow \frac{\partial}{\partial x}\left(\frac{-1}{2}(x-\hat{\mu})^T\hat{C}^{-1}(x-\hat{\mu})\right) = 0$$

$$\Rightarrow \frac{-1}{2}\cdot 2\hat{C}^{-1}(x-\hat{\mu}) = 0$$

$$\Rightarrow x = \hat{\mu}$$

For the given data, as the sample size increases, $\hat{\mu}$ will tend to be at the origin (0, 0) [proof ahead]

$\Rightarrow$ Mode will occur at (0, 0)

PAGE NO.
DATE / /

But (0, 0) is not even included in given dataset (here $x^2+y^2 = 0 \neq r^2$).
Gaussian does not fit this data well.
This is not a good model for the given data as gaussian distribution has points situated around the mean uniformly, at different sentences from the mean.
Here all the data is at a fixed distance from the mean (= mode), which means that all the data would belong to a fixed contour (for a fixed c) and not all contours.

(iii) $(x, y) \equiv (r\cos\theta, r\sin\theta)$
$E(X) = (r\,E(\cos\theta), r\,E(\sin\theta))$
$= (0, 0)$
Since for large sample size,
$\int_0^{2\pi} \sin\theta\, d\theta = 0$ & $\int_0^{2\pi} \cos\theta\, d\theta = 0$

$\therefore E(\sin^2\theta) = 1/2 = E(\cos^2\theta) = 1/2$ (Average values)
(Assuming large sample size
$C(\text{theoretical}) = \begin{bmatrix} r^2\cos^2\theta & r^2\sin\cos\theta \\ r^2\sin\theta\cos\theta & r^2\sin^2\theta \end{bmatrix}$

$\langle C \rangle$ (Average) $= \begin{bmatrix} r^2/2 & 0 \\ 0 & r^2/2 \end{bmatrix}$

$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

For $N = 10^6$ & $r = 1$

$cov \equiv \begin{bmatrix} 0.5001 & -0.0003 \\ -0.0003 & 0.5001 \end{bmatrix} \approx \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \propto \begin{bmatrix} r^2/2 & 0 \\ 0 & r^2/2 \end{bmatrix}$

$mean = [-0.8998 \ \ -0.2882]^T \times 10^{-3} \approx \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

These match theoretically predicted values