SC4000 Machine Learning Tutorial
Density Estimation

Question 1: Suppose a dataset of four 3-dimensional instances is shown in Table 1. Estimate the sample mean and covariance matrix (unbiased).

Table 1: Data set for Question 1.

ID	X_1	X_2	X_3
P1	3	5	-1
P2	-1	8	3
P3	2	-4	-4
P4	0	-1	-6

Question 2: The exponential distribution is often used to describe the amount of time between two events (subject to certain assumptions), such as the time between two emails arriving at your inbox, or the time between two people exiting an MRT station. The probability density function is $P(x) = \lambda \exp(-\lambda x)$ with the support $x \in [0, \infty)$. λ is the single parameter that characterizes the distribution. Suppose we have n data points $x_1, x_2, ..., x_n$ drawn independently from an exponential distribution. Derive the maximum likelihood estimate for λ.

Question 3: Suppose a dataset of five scalar (1-dimensional) data instances is shown in Table 2. Use histogram estimator with an origin of 0 and a width of 3, naive estimator with a width of 3, and 3-NN estimator to estimate the density function $\hat{P}(\mathbf{x})$ and compute the value of $\hat{P}(2.6)$ at 2.6, respectively.

Table 2: Data set for Question 2.

P1	P2	P3	P4	P5
1.2	2	10	-6	3.5