Indian Institute of Technology Madras

ID5055 Foundations of Machine learning

Assignment II

Due date: 16^{th} September 2023

Instruction

1. Assignment shall be submitted on the due date. Late submissions will not be entertained. If you cannot submit the assignment due to some reasons, please contact the instructor by email.
2. All the assignments must be the student's own work. The students are encouraged to collaborate or consult friends. In the case of collaborative work, please write every student's name on the submitted solution.
3. If you find the solution in the book or article or on the website, please indicate the reference in the solutions.

Problems

1. The growth rate of a fungus can be described by the following probability distribution function:

$$f(y; \alpha) = \frac{1}{\alpha^2} y e^{-y/\alpha} \tag{1}$$

   with $\alpha \in (0, \infty)$ and $y \in [0, \infty)$. Find the maximum likelihood estimator for the parameter α.

2. The probability distribution functions of Weibull distribution and Rayleigh distribution are given below:

$$\text{Weibull distribution: } f(x; k, \lambda) = \tfrac{k}{\lambda}(\tfrac{x}{\lambda})^{k-1} e^{-(x/\lambda)^k}, x \geq 0$$

$$\text{Rayleigh distribution: } f(x; \sigma) = \tfrac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, x \geq 0$$

   Use the dataset (link) provided to estimate the parameter λ in Weibull distribution using maximum likelihood estimation (MLE) (assume $k = 2$). Use the property of invariance of MLE to estimate the parameter σ of Rayleigh distribution.

3. Find the maximum likelihood estimate of the parameter θ of the following probability distribution function:

$$f(y; \theta) = \frac{3y^2}{\theta^3} \tag{2}$$

   with $\theta \in (0, \infty)$ and $y \in [0, \theta]$.

4. Dr. AAA collects samples of cancer patients to estimate the mean expression levels of an oncogene. Due to technical limitations, (s)he can collect only 20 samples per day and measure the expression levels of the oncogene. It has been known that the gene expression levels follow normal distribution with standard deviation 8 ($\sim \mathcal{N}(\mu, \sigma = 8)$). Help him/her in estimating the mean gene-expression value using recursive Bayesian estimation. The dataset (link) provided has gene expression levels of the oncogene collected for a 10 days period. Do the following:

i) Assume the prior distribution of μ to be a normal distribution. You can take the sample mean of Day 1 samples and variance as prior parameters

ii) Estimate the posterior distribution of μ using samples from Day 1

iii) Update the priors and repeat step ii) using data from each of the days

iv) Plot the probability distribution of the mean of gene expression level each time after the update

5. Obtain the maximum a posteriori estimate of the parameter $\sigma \in [0, 100]$ in Rayleigh distribution. Assume a normal distribution prior for the parameter σ ($\sim \mathcal{N}(\mu = 15, \sigma = 3)$. The dataset is provided in the (link).

$$\text{Rayleigh distribution: } f(x;\sigma) = \tfrac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, x \geq 0$$