Dr. Boqiang Huang

Mathematical Foundations of Data Analysis II — SS 2019

Problem Set 03

Published on 29.04.2019 To be collected on 07.05.2019

Problem 1: (8 points)
Maximum Likelihood Estimator
Assume that our random sample $\mathbf{X} = [X_1, \ldots, X_n]^T \sim F$, where $F = F_\theta$ is a distribution depending on a parameter θ. By the independence property, the continuous joint probability density function (PDF) (the discrete probability mass function (PMF) will be similar) $p_{\mathbf{X}}(\mathbf{x};\theta) = \Pr(\mathbf{X} = \mathbf{x};\theta)$ can be rewritten as

$$p_{\mathbf{X}}(\mathbf{x};\theta) = \prod_{i=1}^{n} p(x_i;\theta). \tag{1}$$

To estimate the most likely parameter θ, we define the likelihood function as the one of parameter θ

$$L(\theta;\mathbf{X}) = \prod_{i=1}^{n} p(X_i;\theta). \tag{2}$$

The maximum likelihood estimator finds the maximizer of the likelihood function by

$$\hat{\theta}_n = \arg\max_{\theta} L(\theta;\mathbf{X}). \tag{3}$$

In many cases, maximizing the likelihood function might not be easy so one considers maximizing the log-likelihood function:

$$\hat{\theta}_n = \arg\max_{\theta} \ell(\theta;\mathbf{X}) = \arg\max_{\theta} \sum_{i=1}^{n} \ell(\theta;X_i) = \arg\max_{\theta} \sum_{i=1}^{n} \log p(X_i;\theta), \tag{4}$$

where $\ell(\theta;X_i) = \log p(X_i;\theta)$. When the log-likelihood function is differentiable with respect to θ, we consider the score function $\mathbf{s}(\theta;X_i) = \nabla_\theta(\ell(\theta;X_i))$ in order to obtain the estimate $\hat{\theta}$ by solving

$$\mathbf{s}(\hat{\theta}_n;\mathbf{X}) = \sum_{i=1}^{n} \mathbf{s}(\hat{\theta}_n;X_i) = 0. \tag{5}$$

Now, please apply the above-mentioned maximum likelihood estimator to the following cases:
(a) Normal distribution - assume $\mathbf{X} = [X_1, \ldots, X_n]^T \sim N(\mu, 1)$, where $\theta = \mu \in \mathbb{R}$,

$$p(x;\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}; \tag{6}$$

(b) Exponential distribution - assume $\mathbf{X} = [X_1, \ldots, X_n]^T \sim \text{Exp}(\lambda)$, where $\theta = \lambda \in \mathbb{R}$,

$$p(x;\lambda) = \lambda e^{-\lambda x}; \tag{7}$$

(c) Uniform distribution - assume $\mathbf{X} = [X_1, \ldots, X_n]^T \sim \text{Uni}[0,\theta]$, where $\theta \in \mathbb{R}$,

$$p(x;\lambda) = \frac{1}{\theta} \mathbf{1}(0 \le x \le \theta); \quad \text{where } \mathbf{1}(\cdot) \text{ is an indicator function;} \tag{8}$$

(d) Bernoulli distribution - assume $\mathbf{X} = [X_1, \ldots, X_n]^T \sim \text{Ber}(p)$, where $\theta = p \in \mathbb{R}$. Then the PMF is $X = 1$ with a probability of p and $X = 0$ with a probability $1 - p$,

$$P(x;p) = p^X (1-p)^{1-X}; \tag{9}$$

Problem 2: (8 points)
A factory produces very expensive but high quality rings that their qualities are measured in terms of curvature

Curvature	Diameter	QC Result
2.95	6.63	Passed
2.53	7.79	Passed
3.57	5.65	Passed
3.16	5.47	Passed
2.58	4.46	Not Passed
2.16	6.22	Not Passed
3.27	3.52	Not Passed

and diameter. Result of quality control (QC) by experts is given in the table below.
As a consultant to the factory, you get a task to set up the criteria for automatic quality control. Then, the manager of the factory also wants to test your criteria upon new type of rings that even the human experts are argued to each other. The new ring has curvature 2.81 and diameter 5.46.
Try to solve this problem with linear discriminant analysis, and quadratic discriminant analysis respectively.

Problem 3: (4 points)
Please derive the form (23), the quadratic discriminant function, for the case that each class of total k classes of observations has its own covariance matrix.