“Machine Learning and Computational Statistics”

5th Homework

Exercise 1:

Consider the Erlang distribution $p(x) = \theta^2 x\, exp(-\theta\, x) u(x)$, (where $u(x) = 1(0)$, if $x \geq 0$ (< 0)), whose mean equals to $2/\,\theta$.

(a) Given a set of N measurements $x_1, \dots, x_N$, for the random variable x that follows the Erlang distribution, prove that the ML estimate of θ is

$$\theta_{ML} = {}^{2N}\!/_{\sum_{i=1}^{N} x_i}$$

(b) For $N = 5$ and $x_1 = 2$, $x_2 = 2.2$, $x_3 = 2.7$, $x_4 = 2.4$, $x_5 = 2.6$, compute the ML estimate of θ (denoted by $\hat{\theta}$) of the random variable x and write explicitly the formula of the Erlang distribution that “explains” these data. Estimate the mean of the random variable x.

(c) Using the distribution derived in (b), estimate the pdf value for $x'_1 = 2.1$, $x'_2 = 2.3$, $x'_3 = 2.9$.

Exercise 2:

Consider again the Erlang distribution $p(x) = \theta^2 x\, exp(-\theta\, x) u(x)$, (where $u(x) = 1$ (0), if $x \geq 0$ ((< 0)). Given

- a set of N measurements $x_1, \dots, x_N$, for the random variable x that follows the Erlang distribution, and
- the a priori probability for the parameter θ is a normal distribution, $N(\theta_0, {\sigma_0}^2)$ (where $\theta_0, {\sigma_0}^2$ are known)

(a) Compute the MAP estimate of the parameter θ.

(b) How this estimate becomes for the case were (i) N→∞, (ii) ${\sigma_0}^2$>> and (c) ${\sigma_0}^2$<<? Give a short justification.

(c) Let $\theta_0 = 1.1$ and ${\sigma_0}^2$ =1. For $N = 5$ and $x_1 = 2$, $x_2 = 2.2$, $x_3 = 2.7$, $x_4 = 2.4$, $x_5 = 2.6$, compute the MAP estimate of θ (denoted by $\hat{\theta}$) of the random variable x and write explicitly the formula of the Erlang distribution that “explains” these data. Estimate the mean of the random variable x.

(d) Using the distribution derived in (b), estimate the pdf value for $x'_1 = 2.1$, $x'_2 = 2.3$, $x'_3 = 2.9$.

(e) Compare the results with those of exercise 1.

Exercise 3:

Consider the model $x=\mu+\eta$ ($x, \mu, \eta \in R$) and a set of measurements $Y=\{x_1, x_2, \ldots, x_N\}$, which are noisy versions of μ. Assume that we have prior knowledge about μ saying that it lies close to μ_0. Formulating the ridge regression problem as follows

$$\min \quad J(\mu) = \sum_{n=1}^{N} (x_n - \mu)^2,$$

$$\text{subject to } (\mu - \mu_0)^2 \le \rho$$

Prove that

$$\hat{\mu}_{RR} = \frac{\sum_{n=1}^{N} x_n + \lambda\mu_0}{N + \lambda}$$

where λ is a user defined parameter.

Hint: Define the Lagrangian function $L(\mu) = \sum_{n=1}^{N}(x_n - \mu)^2 + \lambda((\mu - \mu_0)^2 - \rho)$ (λ is the Lagrange multiplier corresponding to the constraint).

Exercise 4 (python code + text):

Suppose we have a copper wire and we measure (experimentally) its resistance at various temperatures. An Nx2 array, called Data, of relative measurements is given in the file HW5.mat. Each row of the array corresponds to a measurement $(\theta_i, R_i), i = 1, \ldots, N$, where θ_i corresponds to the temperature and R_i to the associated resistance.

(i) Determine the relation $R = f(\theta)$, assuming that the joint pdf of the random variables θ and R, corresponding to the temperature and the resistance, respectively, is normal and utilizing the MSE criterion.

(ii) Use the Data_test matrix to assess the performance of the regressor. Specifically, considering each row of the matrix, use the first element (temperature) and estimate the value of the resistance using the relation derived in (i). Then, compare this estimate with second element of the row (real value).