"Machine Learning and Computational Statistics"

6th Homework

Exercise 1:

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

(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

$$\theta_{ML} = \frac{2N}{\sum_{i=1}^{N} x_i}$$

(b) For $N = 5$ and $x_1 = 2$, $x_2 = 2.2$, $x_3 = 2.7$, $x_4 = 2.4$, $x_5 = 2.6$, estimate the θ_{ML}. Utilizing this estimate, determine $\hat{p}(x)$, for $x = 2.3$ and $x = 2.9$.

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 θ_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) $N \to \infty$, (ii) ${\sigma_0}^2 \gg$ and (c) ${\sigma_0}^2 \ll$? Give a short justification.

Exercise 3:

Consider the model $x=\mu+\eta$ ($x, \mu, \eta \in R$) and a set of measurements $Y = \{x_1, \dots, x_N\}$, 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

$min_\mu\, J(\mu) = \sum_{n=1}^{N} (x_n - \mu)^2$, subject to $(\mu - \mu_0)^2 \leq \rho$

Prove that

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

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:

Consider a data set $Y = \{x_1, \dots, x_N\}$, whose elements have been drawn independently from the exponential distribution

$$p(x|\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

The parameter λ of the distribution is modelled by a prior gamma distribution, i.e.,

$$p(\lambda) \equiv p(\lambda; a, b) = \begin{cases} \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}, & \lambda \geq 0 \\ 0, & \lambda < 0 \end{cases}$$

(a) **Determine** the likelihood $p(Y|\lambda)$.
(b) **Form** the product of the prior and the likelihood and determine the MAP estimate of λ.
(c) **Give** the form of $p(x)$, in terms of the MAP estimate of λ, determined in (b).
(d) **Determine** the posterior distribution for λ, $p(\lambda| Y)$, in the light of the Y.
(e) **Compare** the form of the resulting posterior $p(\lambda| Y)$ with that of the prior $p(\lambda)$ of λ and comment briefly.
(f) **Prove** that $p(x|Y)$ is a lomax distribution.

For the following, assume that $Y = \{2.8, 2.4, 2.9, 2.6, 2.1, 2.2\}$, $a = 2$ and $b = 2$.

(g) **Write down** the $p(x)$ of (c) for the above Y.
(h) **Write down** the $p(x|Y)$ of (f) for the above Y.
(i) **Compute** $p(x)$ (from (g)) and $p(x|Y)$ (from (h)), for $x = 2.5$.

Hints: (i) For (a) and (b) work as in exercise 2.

(ii) For (d): A pdf of the form $C\lambda^r e^{-s\lambda}$ is a gamma distribution with parameters r and s.

(iii) For (f): (I) It is $\int_0^\infty t^b e^{-at} dt = \frac{\Gamma(b+1)}{a^{b+1}}$ and (II) $\Gamma(z+1) = z\Gamma(z)$. (III) The lomax distribution is defined as $p(x; c, d) = \frac{cd^c}{(x+d)^{c+1}}$, for $x \geq 0$ and 0 otherwise. (IV) In order to completely determine $p(x|Y)$, the values c, d of the distribution need to be determined.