1. Statistical Estimation

Given $X \in \{0, 1\}$, the random variable $Y$ is exponentially distributed with rate $3X + 1$.

(a) Assume $P(X = 1) = p \in (0, 1)$ and $P(X = 0) = 1 - p$. Find the MAP estimate of $X$ given $Y$.

(b) Find the MLE of $X$ given $Y$.

Solution:

(a) We know that when $X = 0$, $f_{Y|X}(y \mid 0) = \exp(-y)1\{y > 0\}$ and when $X = 1$, $f_{Y|X}(y \mid 1) = 4\exp(-y)1\{y > 0\}$. The MAP maximizes $f_{X|Y}(x, y)$ over $x$ for the given observation $y$, which is equivalent to maximizing $f_{X,Y}(x, y)$. Thus,

$f_{X,Y}(0, y) = (1 - p)\exp(-y)1\{y > 0\},$

and

$f_{X,Y}(1, y) = 4p\exp(-4y),$

and

$\text{MAP}[X \mid Y] = 1 \iff 4p\exp(-4Y) > (1 - p)\exp(-Y)$

which gives

$\text{MAP}[X \mid Y] = 1\left\{ Y < \frac{1}{3} \ln 4p \right\}.$

(b) The MLE is the MAP estimate with the prior probability $p$ set to $1/2$.

$\text{MLE}[X \mid Y] = 1\left\{ Y < \frac{1}{3} \ln 4 \right\} = 1\{Y < 0.462\}.$

2. Poisson Process MAP

Customers arrive to a store according to a Poisson process of rate 1. The store manager learns of a rumor that one of the employees is sending $1/2$ of the customers to the rival store. Refer to hypothesis $X = 1$ as the rumor being true, that one of the employees is sending every other customer arrival to the rival store and hypothesis $X = 0$ as the rumor being false, where each hypothesis is equally likely. Assume that at time 0, there is a successful sale. After that, the manager observes $S_1, S_2, \ldots, S_n$ where $n$ is a positive integer.
and $S_i$ is the time of the $i$th subsequent sale for $i = 1, \ldots, n$. Derive the MAP rule to determine whether the rumor was true or not.

**Solution:**

Note that both hypotheses are a priori equally likely, so the MAP rule is equivalent to the ML rule. The interarrival times are independent conditioned on $X = 1$ and $X = 0$. The density of an interarrival interval given $X = 1$ is Erlang of order 2, so for $0 \leq s_1 < \cdots < s_n$:

$$f_{S|X}(s_1, s_2, \ldots, s_n | 1) = \prod_{i=1}^{n} (s_i - s_{i-1}) e^{-(s_i - s_{i-1})} = e^{-s_n} \prod_{i=1}^{n} (s_i - s_{i-1})$$

The density of an interarrival interval given $X = 0$ is exponential, so:

$$f_{S|X}(s_1, s_2, \ldots, s_n | 0) = e^{-s_n}$$

We can thus see, by taking the log of both expressions, we declare $X = 1$ if $\sum_{i=1}^{n} \log(S_i - S_{i-1}) \geq 0$, otherwise we declare $X = 0$.

3. **Laplace Prior & $\ell^1$-Regularization**

Suppose you draw $n$ i.i.d. data points $(x_1, y_1), \ldots, (x_n, y_n)$, where $n$ is a positive integer and the true relationship is $Y = WX + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. (That is, $Y$ has a linear dependence on $X$, with additive Gaussian noise.) Further suppose that $W$ has a prior distribution with density

$$f_W(w) = \frac{1}{2\beta} e^{-|w|/\beta}, \quad \beta > 0.$$

(This is known as the Laplace distribution.) Show that finding the MAP estimate of $W$ given the data points $\{(x_i, y_i) : i = 1, \ldots, n\}$ is equivalent to minimizing the cost function

$$J(w) = \sum_{i=1}^{n} (y_i - wx_i)^2 + \lambda|w|$$

(you should determine what $\lambda$ is). This is interpreted as a one-dimensional $\ell^1$-regularized least-squares criterion, also known as LASSO.

**Solution:**

The likelihood for $W$ is

$$\mathcal{L}(w \mid (x_1, y_1), \ldots, (x_n, y_n)) \propto \mathcal{L}(x_1, y_1), \ldots, (x_n, y_n \mid W = w) f_W(w)$$

(technically, the expression on the right should be divided by the likelihood of the data, but this has no dependence on $w$, so we omit the denominator for simplicity)

$$= \prod_{i=1}^{n} \mathcal{L}(x_i, y_i \mid W = w) f_W(w)$$
(the data points are conditionally independent given $W$)

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} e^{-(y_i - wx_i)^2/(2\sigma^2)} \cdot \frac{1}{2\beta} e^{-|w|/\beta}$$

(here we say that the likelihood of $(x_i, y_i)$ given $W$ is the density of $\varepsilon_i$, which is $\mathcal{N}(0, \sigma^2)$, evaluated at $y_i - wx_i$)

$$\propto \prod_{i=1}^{n} e^{-(y_i - wx_i)^2/(2\sigma^2)} e^{-|w|/\beta}$$

(again, we throw out constant factors that do not depend on the data points or $w$).

We wish to maximize this expression w.r.t. $w$, but we will find it more convenient to take the log-likelihood instead.

$$\ell(w \mid (x_1, y_1), \ldots, (x_n, y_n)) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - wx_i)^2 - \frac{1}{\beta} |w|.$$  

Since we want to maximize the log-likelihood, this is equivalent to minimizing the cost function

$$J(w) = \sum_{i=1}^{n} (y_i - wx_i)^2 + \lambda |w|,$$

where $\lambda = 2\sigma^2 / \beta$. 