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

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

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

Solution 1. (a) We know that when $X = 0$, $f_{Y \mid X}(y \mid 0) = \exp(-y)\mathbf{1}\{y > 0\}$ and when $X = 1$, $f_{Y \mid X}(y \mid 1) = 4\exp(-y)\mathbf{1}\{y > 0\}$. 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,

$$f_{X,Y}(0,y) = (1 - p)\exp(-y)\mathbf{1}\{y > 0\},$$
$$f_{X,Y}(1,y) = 4p\exp(-4y)\mathbf{1}\{y > 0\},$$

and

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

which gives

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

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

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

Problem 2. Exponential: MLE & MAP
The random variable $X$ is exponentially distributed with mean 1. Given $X$, the random variable $Y$ is exponentially distributed with rate $X$.

1. Find $\text{MLE}[X \mid Y]$.
2. Find $\text{MAP}[X \mid Y]$.

Solution 2. 1. The density of $Y$, given $X = x$, is $f(y) = x\exp(-xy)$ for $y > 0$, so $\ln f(y) = \ln x - xy$. To maximize this over $x$, we differentiate to obtain $1/x - y = 0$, so $x = 1/y$, that is, $\text{MLE}[X \mid Y] = 1/Y$. 


2. The posterior density of $X$ is

$$f_{X|Y}(x \mid y) \propto f_{Y|X}(y \mid x)f_X(x) = x \exp(-xy) \exp(-x)$$

$$= x \exp(-x(1+y))$$

so we can maximize $\ln x - x(1+y)$ over $x$. Differentiating, we have $1/x - 1 - y = 0$, or $1/x = 1 + y$. Hence, $MAP[X \mid Y] = 1/(1 + Y)$.

**Problem 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 3.** 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 $W$)

$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(y_i - wx_i)^2}{(2\sigma^2)}\right) \frac{1}{2\beta} \exp\left(-\frac{|w|}{\beta}\right)$$

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