1. Gaussians and the MSE

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

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

**Solution:**

The likelihood for the data is

$$L((x_1, y_1), \ldots, (x_n, y_n) \mid W = w) = \prod_{i=1}^{n} L((x_i, y_i) \mid W = w)$$

(the data points are conditionally independent given $W$)

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

(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.

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

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

$$J(w) = \sum_{i=1}^{n} (y_i - wx_i)^2$$
2. Hypothesis Testing for Bernoulli Random Variables

Assume that

- If $X = 0$, then $Y \sim \text{Bernoulli}(1/4)$.
- If $X = 1$, then $Y \sim \text{Bernoulli}(3/4)$.

Using the Neyman-Pearson formulation of hypothesis testing, find the optimal randomized decision rule $r : \{0, 1\} \rightarrow \{0, 1\}$ with respect to the criterion

$$
\min_{\text{randomized } r : \{0, 1\} \rightarrow \{0, 1\}} \mathbb{P}(r(Y) = 0 \mid X = 1)
$$

$$
s.t. \mathbb{P}(r(Y) = 1 \mid X = 0) \leq \beta,
$$

where $\beta \in [0, 1]$ is a given upper bound on the false positive probability.

Solution:

Here, the likelihood ratio is

$$
\frac{f_{Y \mid X}(y \mid 1)}{f_{Y \mid X}(y \mid 0)} = \begin{cases} 
3, & \text{if } y = 1 \\
\frac{1}{3}, & \text{if } y = 0
\end{cases}
$$

If $\beta \leq P(Y = 1 \mid X = 0)$, then the optimal decision rule is to have $r(0) = 0$ and have $r(1) = 1$ with probability $\gamma = \frac{\beta}{1/4}$. Otherwise, the optimal decision rule is $r(1) = 1$ and have $r(0) = 1$ with probability $\gamma$, chosen to make $\mathbb{P}(r(Y) = 1 \mid X = 0) = \beta$. Then,

$$
\frac{1}{4} + \frac{3}{4} \gamma = \beta
$$

so $\gamma = \frac{4}{3} \beta - \frac{1}{3}$.

3. Bayesian Hypothesis Testing for Gaussian Distribution

Assume that $X$ has prior probabilities $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = 1/2$. Further

- If $X = 0$, then $Y \sim \mathcal{N}(\mu_0, \sigma_0^2)$.
- If $X = 1$, then $Y \sim \mathcal{N}(\mu_1, \sigma_1^2)$.

Assume $\mu_0 < \mu_1$ and $\sigma_0 < \sigma_1$.

Using the Bayesian formulation of hypothesis testing, find the optimal decision rule $r : \mathbb{R} \rightarrow \{0, 1\}$ with respect to the minimum expected cost criterion

$$
\min_{r : \mathbb{R} \rightarrow \{0, 1\}} \mathbb{E}[I\{r(Y) \neq X\}],
$$

Solution:

We can write

$$
E[I(r(Y) \neq X)] = P(r(Y) \neq X)
$$

$$
= P(r(Y) = 1 \mid X = 0) \cdot \frac{1}{2} + P(r(Y) = 0 \mid X = 1) \cdot \frac{1}{2}
$$
We can write $P(r(Y) = 1 \mid X = 0)$ as

$$P(r(Y) = 1 \mid X = 0) = \int \begin{cases} f(y \mid X = 0) & \text{if } r(y) = 1 \\ 0 & \text{otherwise} \end{cases} \, dy$$

and do something similar for $P(r(Y) = 0 \mid X = 1)$. Combining everything together, we get

$$E[I(r(Y) \neq X)] = \frac{1}{2} \int \begin{cases} f(y \mid X = 0) & \text{if } r(y) = 1 \\ f(y \mid X = 1) & \text{if } r(y) = 0 \end{cases} \, dy$$

Since we’re free to choose $r(y)$ as 0 or 1, trying to minimize this leads to

$$\begin{cases} 0, & \text{if } f(y \mid X = 0) > f(y \mid X = 1) \\ 1, & \text{if } f(y \mid X = 0) < f(y \mid X = 1). \end{cases}$$

The condition $f(y \mid X = 0) < f(y \mid X = 1)$ can be written as

$$\left( \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) y^2 - 2 \left( \frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2} \right) y + \left( \frac{\mu_0^2}{\sigma_0^2} - \frac{\mu_1^2}{\sigma_1^2} - 2 \ln \frac{\sigma_1^2}{\sigma_0^2} \right) > 0,$$

and if we let $a < b$ be the two roots of this quadratic, then the optimal decision rule can be written as

$$r(y) = \begin{cases} 0, & \text{if } y \in (a, b) \\ 1, & \text{if } y \in (-\infty, a) \cup (b, \infty). \end{cases}$$