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

2. Flipping Coins and Hypothesizing

You flip a coin until you see heads. Let

$$X = \begin{cases} 1 & \text{if the bias of the coin is } q > p, \\ 0 & \text{if the bias of the coin is } p. \end{cases}$$

Find a decision rule $\hat{X}(Y)$ that maximizes $P[\hat{X} = 1 \mid X = 1]$ subject to $P[\hat{X} = 1 \mid X = 0] \leq \beta$ for $\beta \in [0, 1]$. Remember to calculate the randomization constant $\gamma$.

3. Projections

The following exercises are from the note on the Hilbert space of random variables. See the notes for some hints.

(a) Let $\mathcal{H} := \{X : X$ is a real-valued random variable with $\mathbb{E}[X^2] < \infty\}$. Prove that $(X, Y) := \mathbb{E}[XY]$ makes $\mathcal{H}$ into a real inner product space.$^1$

(b) Let $U$ be a subspace of a real inner product space $V$ and let $P$ be the projection map onto $U$. Prove that $P$ is a linear transformation.

(c) Suppose that $U$ is finite-dimensional, $n := \dim U$, with basis $\{v_i\}_{i=1}^{n}$. Suppose that the basis is orthonormal. Show that $Py = \sum_{i=1}^{n} (y, v_i)v_i$. (Note: If we take $U = \mathbb{R}^n$ with the standard inner product, then $P$ can be represented as a matrix in the form $P = \sum_{i=1}^{n} v_i v_i^T$.)

---

$^1$To be perfectly correct, it is possible for $X \neq 0$ but $\mathbb{E}[X^2] = 0$; this occurs if $X = 0$ with probability 1. To fix this, we need to define two random variables $X$ and $Y$ to be equal if $P(X = Y) = 1$. In other words, we consider equivalence classes of random variables, defined by the relation $\equiv$. With this definition, then if $X \neq 0$ we do indeed have $\mathbb{E}[X^2] > 0$. 

1
4. MMSE and Conditional Expectation

Let $X, Y_1, \ldots, Y_n$ be square integrable random variables. The MMSE of $X$ given $(Y_1, \ldots, Y_n)$ is defined as the function $\phi(Y_1, \ldots, Y_n)$ which minimizes the mean square error

$$E[(X - \phi(Y_1, \ldots, Y_n))^2].$$

(a) For this part, assume $n = 1$. Show that the MMSE is precisely the conditional expectation $E[X|Y]$. Hint: expand the difference as $(X - E[X|Y] + E[X|Y] - \phi(Y))$.

(b) Argue that

$$E[(X - E[X|Y_1, \ldots, Y_n])^2] \leq E\left[\left(X - \frac{1}{n} \sum_{i=1}^{n} E[X|Y_i]\right)^2\right].$$

That is, the MMSE does better than the average of the individual estimates given each $Y_i$.

5. Exam Difficulties

The difficulty of an EECS 126 exam, $\Theta$, is uniformly distributed on $[0, 100]$ (i.e. continuous distribution, not discrete), and Alice gets a score $X$ that is uniformly distributed on $[0, \Theta]$. Alice gets her score back and wants to estimate the difficulty of the exam.

(a) What is the MLE of $\Theta$? What is the MAP of $\Theta$?

(b) What is the LLSE for $\Theta$?