ST790 Homework 1 Solution

1. (Optimal Decision Rule)

(a) Let $g(a) = E(X-a)^2 = E(X^2) - 2aE(X) + a^2$. Setting its partial derivative wrt a to zero gives us

$$\frac{\partial f(a)}{\partial a} = 2a - 2E(X) = 0,$$

or $a = \mathbb{E}(X)$. Also, check that the second derivative is

$$\frac{\partial^2 f(a)}{\partial a^2} = 2 > 0.$$

Thus $g(a)$ is minimized when $a = E(X)$ and $\min_a g(a) = E(X - EX)^2$.

(b) Let $g(a) = E|X-a| = \int_a^\infty (x-a)f(x)dx + \int_{-\infty}^a (a-x)f(x)dx$. Setting its partial derivative wrt a to zero gives us

$$\frac{\partial g(a)}{\partial a} = \int_a^\infty -f(x)dx + \int_{-\infty}^a f(x)dx = 0,$$

which means $F(x) = 1/2$ or $a = m$. Also, check that the second derivative is

$$\frac{\partial^2 g(a)}{\partial a^2} = 2f(a) > 0.$$

Thus $g(a)$ is minimized when $a = m$ and $\min_a g(a) = E|X - m|$.

2. (Bayes Estimator Under Squared Error Loss)

(a) The likelihood function is

$$f(X_1, \cdots, X_n|p) = p^{\sum_{i=1}^n X_i}(1-p)^{\sum_{i=1}^n (1-X_i)},$$

and the prior kernel of p is

$$\pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$$

The posterior kernel is then

$$\pi(p|X_1, \cdots, X_n) \propto f(X_1, \cdots, X_n|p)\pi(p) = p^{\sum_{i=1}^n X_i + \alpha - 1}(1-p)^{n - \sum_{i=1}^n X_i + \beta - 1}.$$

So the posterior distribution is $B(\sum_{i=1}^n X_i + \alpha, n - \sum_{i=1}^n X_i + \beta)$.

(b)

$$\hat{p}_{\text{bayes}} = \mathbb{E}(p|X_1, \cdots, X_n) = \frac{\sum_{i=1}^n X_i + \alpha}{\sum_{i=1}^n X_i + \alpha + n - \sum_{i=1}^n X_i + \beta} = \frac{\sum_{i=1}^n X_i + \alpha}{\alpha + \beta + n}$$

(c) From (b) we know

$$\mathbb{E}(\hat{p}_{\text{bayes}}) = \frac{np + \alpha}{\alpha + \beta + n}, \quad Var(\hat{p}_{\text{bayes}}) = \frac{np(1-p)}{(\alpha+\beta+n)^2}.$$

So the risk of $\hat{p}_{\text{bayes}}$ is

$$\begin{aligned} R(p, \hat{p}_{\text{bayes}}) = \mathbb{E}(p - \hat{p}_{\text{bayes}})^2 &= \mathbb{E}(\hat{p}^2_{\text{bayes}}) - 2p\mathbb{E}(\hat{p}_{\text{bayes}}) + p^2 \\ &= (\frac{np+\alpha}{\alpha+\beta+n})^2 + \frac{np(1-p)}{(\alpha+\beta+n)^2} - 2p \times \frac{np+\alpha}{\alpha+\beta+n} + p^2 \\ &= \frac{np(1-p)}{(\alpha+\beta+n)^2} + (\frac{np+\alpha}{\alpha+\beta+n} - p)^2 \end{aligned}$$

(d) Taking $\alpha = \beta = \sqrt{n/4}$, we have

$$\hat{p}_{\text{bayes}} = \frac{\sum_{i=1}^n X_i + \sqrt{n/4}}{n + \sqrt{n}}, \quad R(p, \hat{p}_{\text{bayes}}) = \frac{n}{4(n+\sqrt{n})^2}.$$

3. (Exercise 2.7)

(a) For linear regression, denote the design matrix as $X = (x_1, \cdots, x_N)^T$ and $y = (y_1, \cdots, y_N)^T$. The linear regression predictor is

$$\hat{f}(x_0) = x_0^T \hat{\beta} = x_0^T (X^T X)^{-1} X^T y = \sum_{i=1}^N x_0^T (X^T X)^{-1} x_i y_i.$$

Therefore $l_i(x_0; \mathcal{X}) = x_0^T (X^T X)^{-1} x_i$. For k-nearest-neighbor, the predictor is

$$\hat{f}(x_0) = \sum_{i=1}^N \frac{1}{k} I\{x_i \in N_k(x_0)\} y_i,$$

where $N_k(x)$ is the set of the k nearest neighbors of x. So $l_i(x_0; \mathcal{X}) = I\{x_i \in N_k(x_0)\}/k$.

(b)

$$\begin{aligned}
\mathbb{E}_{\mathcal{Y}|\mathcal{X}}(f(x_0) - \hat{f}(x_0))^2 &= \mathbb{E}_{\mathcal{Y}|\mathcal{X}}(f(x_0) - \mathbb{E}_{\mathcal{Y}|\mathcal{X}}\hat{f}(x_0) + \mathbb{E}_{\mathcal{Y}|\mathcal{X}}\hat{f}(x_0) - \hat{f}(x_0))^2 \\
&= \left(f(x_0) - \mathbb{E}_{\mathcal{Y}|\mathcal{X}}\hat{f}(x_0)\right)^2 + Var_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0)) \\
&= \left(f(x_0) - \sum_{i=1}^N \ell_i(x_0; \mathcal{X}) f(x_i)\right)^2 + \sum_{i=1}^N (\ell_i(x_0; \mathcal{X}))^2 \sigma^2.
\end{aligned}$$

(c)

$$\begin{aligned}
&\mathbb{E}_{\mathcal{Y},\mathcal{X}}(f(x_0) - \hat{f}(x_0))^2 \\
=&\mathbb{E}_{\mathcal{Y},\mathcal{X}}(f(x_0) - \mathbb{E}_{\mathcal{Y},\mathcal{X}}\hat{f}(x_0) + \mathbb{E}_{\mathcal{Y},\mathcal{X}}\hat{f}(x_0) - \hat{f}(x_0))^2 \\
=&\left(f(x_0) - \mathbb{E}_{\mathcal{Y},\mathcal{X}}\hat{f}(x_0)\right)^2 + Var_{\mathcal{Y},\mathcal{X}}(\hat{f}(x_0)) \\
=&\int \cdots \int \left(\left(f(x_0) - \sum_{i=1}^N \ell_i(x_0; \mathcal{X}) f(x_i)\right)^2 + \sum_{i=1}^N (\ell_i(x_0; \mathcal{X}))^2 \sigma^2\right) \prod_{i=1}^N h(x_i) dx_1 \cdots dx_N.
\end{aligned}$$

(d) The result in (c) is the expectation of the result in (b). Other reasonable observations are also acceptable.