Homework 2
Due Friday Feb 22 3:00 pm
Submit a pdf file on Canvas

1. Consider data $(X_1, Y_1), \ldots, (X_n, Y_n)$ where $X_i \in \mathbb{R}$ and $Y_i \in \mathbb{R}$. Inspired by the fact that $\mathbb{E}[Y|X = x] = \int y p(x, y) dy / p(x)$, define

$$\widehat{m}(x) = \frac{\int y \widehat{p}(x, y) dy}{\widehat{p}(x)}$$

where

$$\widehat{p}(x) = \frac{1}{n} \sum_i \frac{1}{h} K\left(\frac{X_i - x}{h}\right)$$

and

$$\widehat{p}(x, y) = \frac{1}{n} \sum_i \frac{1}{h^2} K\left(\frac{X_i - x}{h}\right) K\left(\frac{Y_i - y}{h}\right).$$

Assume that $\int K(u) du = 1$ and $\int u K(u) du = 0$. Show that $\widehat{m}(x)$ is exactly the kernel regression estimator that we defined in class.

2. Suppose that (X, Y) is bivariate Normal:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ \eta \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{pmatrix} \right).$$

(a) Show that $m(x) = \mathbb{E}[Y|X = x] = \alpha + \beta x$ and find explicit expressions for α and β.

(b) Find the maximum likelihood estimator $\widehat{m}(x) = \widehat{\alpha} + \widehat{\beta} x$.

(c) Show that $|\widehat{m}(x) - m(x)|^2 = O_P(n^{-1})$.

3. Let $m(x) = \mathbb{E}[Y|X = x]$. Let $X \in [0, 1]^d$, Divide $[0, 1]^d$ into cubes $B_1, \ldots, B_N$ whose sides have length h. The data are $(X_1, Y_1), \ldots, (X_n, Y_n)$. In this problem, treat the X_i's as fixed. Assume that the number of observations in each bin is positive. Let

$$\widehat{m}(x) = \frac{1}{n(x)} \sum_i Y_i I(X_i \in B(x))$$

where $B(x)$ is the cube containing x and $n(x) = \sum_i I(X_i \in B(x))$. Assume that

$$|m(y) - m(x)| \leq L||x - y||_2$$

for all x, y. You may further assume that $\sup_x \mathrm{Var}(Y|X = x) < \infty$.

(a) Show that

$$|\mathbb{E}[\widehat{m}(x)] - m(x)| \leq C_1 h$$

for some $C_1 > 0$. Also show that

$$\mathrm{Var}(\widehat{m}(x)) \le \frac{C_2}{n(x)}$$

for some $C_2 > 0$.

(b) Now let X be random and assume that X has a uniform density on $[0,1]^d$. Let $h \equiv h_n = (C \log n/n)^{1/d}$. Show that, for $C > 0$ large enough, $P(\min n_j = 0) \to 0$ as $n \to \infty$ where n_j is the number of observations in cube B_j.

4. Consider the RKHS problem

$$\widehat{f} = \operatorname*{argmin}_{f \in \mathcal{H}} \sum_{i=1}^{n} \big(y_i - f(x_i)\big)^2 + \lambda \|f\|_{\mathcal{H}}^2, \tag{1}$$

for some Mercer kernel function $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. In this problem, you will prove that the above problem is equivalent to the finite dimensional one

$$\widehat{\alpha} = \operatorname*{argmin}_{\alpha \in \mathbb{R}^n} \|y - K\alpha\|_2^2 + \lambda \alpha^T K \alpha, \tag{2}$$

where $K \in \mathbb{R}^{n \times n}$ denotes the kernel matrix $K_{ij} = K(x_i, x_j)$.

Recall that the functions $K(\cdot, x_i)$, $i = 1, \dots n$ are called the *representers of evaluation.* Recall that

- $\langle f, K(\cdot, x_i) \rangle_{\mathcal{H}} = f(x_i)$, for any function $f \in \mathcal{H}$
- $\|f\|_{\mathcal{H}}^2 = \sum_{i,j=1}^n \alpha_i \alpha_j K(x_i, x_j)$ for any function $f = \sum_{i=1}^n \alpha_i K(\cdot, x_i)$.

(a) Let $f = \sum_{i=1}^n \alpha_i K(\cdot, x_i)$, and consider defining a function $\widetilde{f} = f + \rho$, where ρ is any function orthogonal to $K(\cdot, x_i)$, $i = 1, \dots n$. Using the properties of the representers, prove that $\widetilde{f}(x_i) = f(x_i)$ for all $i = 1, \dots n$, and $\|\widetilde{f}\|_{\mathcal{H}}^2 \ge \|f\|_{\mathcal{H}}^2$.

(b) Conclude from part (a) that in the infinite-dimensional problem (1), we need only consider functions of the form $f = \sum_{i=1}^n \alpha_i K(\cdot, x_i)$, and that this in turn reduces to (2).

5. Let $X = (X(1), \dots, X(d)) \in \mathbb{R}^d$ and $Y \in \mathbb{R}$. In the questions below, make any reasonable assumptions that you need but state your assumptions.

(a) Prove that $\mathbb{E}(Y - m(X))^2$ is minimized by choosing $m(x) = \mathbb{E}(Y|X = x)$.

(b) Find the function $r(x)$ that minimizes $\mathbb{E}|Y - r(X)|$. (You can assume that the conditional cdf $F(y|X = x)$ is continuous and strictly increasing, for every x.)

(c) Prove that $\mathbb{E}(Y - \beta^T X)^2$ is minimized by choosing $\beta_* = B^{-1}\alpha$ where $B = \mathbb{E}(XX^T)$ and $\alpha = (\alpha_1, \dots, \alpha_d)$ and $\alpha_j = \mathbb{E}(Y X(j))$.

6. Consider the many Normal means problem where we observe $Y_i \sim N(\theta_i, 1)$ for $i = 1, \ldots, d$. Let $\widehat{\theta}$ minimize the penalized loss

$$\sum_i (Y_i - \theta_i)^2 + \lambda J(\theta).$$

Find an explicit expression for $\widehat{\theta}$ in three cases: (i) $J(\theta) = ||\theta||_0$, (ii) $J(\theta) = ||\theta||_1$, (iii) $J(\theta) = ||\theta||_2$.