Practice Questions for Midterm
10/36-702

(1) Let $X_1, \ldots X_n \sim \text{Unif}(0,1)$. Compute the bias and variance of the histogram density estimator with binwidth h for this distribution. Show that the optimal value of h is $h = 1$.

(2) Let $X_1, \ldots, X_n \sim P$ where p has a density p on $\mathbb{R}$. Assume that $p(x) > 0$ for each $x \in \mathbb{R}$. Given $c_1, \ldots, c_k \in \mathbb{R}$, the population k-means risk is

$$R(k) = \inf_{c_1,\ldots,c_k} \mathbb{E}\Big(\min_{j=1,\ldots k} |X - c_j|^2 \Big).$$

Show that $R(k)$ is strictly decreasing in k.

(3) Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be iid. Suppose that $X_1, \ldots, X_n \sim P$ has a density p on $[0,1]$ where $0 < c \leq p(x) \leq C < \infty$ for all $x \in [0,1]$. Assume that the density p is known. Assume that

$$Y_i = m(X_i) + \epsilon_i$$

where $\epsilon_1, \ldots, \epsilon_n$ are iid with mean 0 and variance σ^2. Assume that m, m', m'', m''' p, p', p'', p''' are bounded and continuous functions. Let $x \in (0,1)$ and define

$$\widehat{m}(x) = \frac{\frac{1}{n}\sum_{i=1}^n Y_i \frac{1}{h} K\left(\frac{x - X_i}{h}\right)}{p(x)}$$

where K is a smooth, symmetric, kernel with bounded support. Show that

$$\mathbb{E}[\widehat{m}(x)] = m(x) + Ch^2 + O(h^3).$$

(4) Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be iid. Suppose that $Y_i \in \{0,1\}$ and $X_i \in [0,1]$. Let $\theta = P(Y_i = 1)$. Assume that $0 < \theta < 1$. Suppose that

$$X_i \,|\, Y_i = 1 \ \sim \ p_1$$

and

$$X_i \,|\, Y_i = 0 \ \sim \ p_0$$

where p_0 and p_1 are densities on $[0,1]$. Assume that, for some constants, c and C,

$$0 < c \leq p_j(x) \leq C < \infty$$

for all $x \in [0,1]$ and $j = 0, 1$.

Let $\widehat{p}_0$ be an estimate of p_0 and let $\widehat{p}_1$ be an estimate of p_1. Define

$$\widehat{h}(x) = \begin{cases} 1 & \text{if } \widehat{m}(x) \geq 1/2 \\ 0 & \text{if } \widehat{m}(x) < 1/2 \end{cases}$$

where

$$\widehat{m}(x) = \frac{\widehat{\theta}\,\widehat{p}_1(x)}{\widehat{\theta}\,\widehat{p}_1(x) + (1-\widehat{\theta})\,\widehat{p}_0(x)},$$

$\widehat{\theta} = n^{-1}\sum_{i=1}^n Y_i$,

Suppose that

$$\sup_x |\widehat{p}_0(x) - p_0(x)| \xrightarrow{P} 0, \quad \text{and} \quad \sup_x |\widehat{p}_1(x) - p_1(x)| \xrightarrow{P} 0.$$

Show that

$$\mathbb{P}(Y \neq \widehat{h}(X)) - \mathbb{P}(Y \neq h_*(X)) \xrightarrow{P} 0$$

as $n \to \infty$, where h_* is the Bayes classifier, and $\mathbb{P}$ is probability with respect to X and Y, but not with respect to $\widehat{h}$.

(5) Let p be a bounded continuous density defined on a bounded subset $S \subset \mathbb{R}$. Assume further that p has bounded, continuous first and second derivatives. Let $Y_1, \ldots, Y_n \sim p$ and let

$$\widehat{p}_h(y) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h} K\left(\frac{y - Y_i}{h}\right).$$

Let $p_h(x) = \mathbb{E}[\widehat{p}_h(x)]$.

(a) Show that, for any $t > 0$, $\mathbb{P}(|\widehat{p}_h(x) - p_h(x)| > t) \to 0$ as long as $nh \to \infty$.

(b) Let $C_h = \{x : \; p_h(x) > \lambda\}$ and let $\widehat{C}_h = \{x : \; \widehat{p}_h(x) > \lambda\}$. Show that $\widehat{C}_h$ is a consistent estimator of C_h in the following sense: (i) if $p_h(x) > \lambda$ then $\mathbb{P}(x \in \widehat{C}_h) \to 1$ and (ii) if $p_h(x) < \lambda$ then $\mathbb{P}(x \notin \widehat{C}_h) \to 1$.

(6) Let $Y_i = \beta^T X_i + \epsilon_i$ where $Y_i \in \mathbb{R}$, $X_i \in \mathbb{R}^d$ and $\epsilon_i \sim N(0, \sigma^2)$. Recall that the ridge estimator is

$$\widehat{\beta} = (\mathbb{X}^T\mathbb{X} + \lambda I)^{-1}\mathbb{X}^T Y,$$

where $\mathbb{X} \in \mathbb{R}^{n\times d}$, $Y = (Y_1, \ldots, Y_n)$ and $\lambda \geq 0$. Find $\mathbb{E}[\widehat{\beta}|X_1, \ldots, X_n]$ and $\mathrm{Var}[\widehat{\beta}|X_1, \ldots, X_n]$. Show that $\mathrm{Var}[\widehat{\beta}|X_1, \ldots, X_n] \to 0$ as $\lambda \to \infty$. Show that the bias tends to 0 as $\lambda \to 0$ if $d < n$.

(7) Let $(X, Y) \sim P$, and consider predicting the value of Y from X. That is, consider choosing a function f to minimize

$$\mathbb{E}\big[\big(Y - f(X)\big)^2\big].$$

Show that the function minimizing this is given by

$$f(x) = \frac{\int y \cdot p_{X,Y}(x,y)\,dy}{p_X(x)},$$

where $p_{X,Y}$ is the joint density of (X, Y), and p_X is the density of X.