Assignment 2

1. (a)

$$
\begin{aligned}
E(\hat{m}(x) - m(x))^2 &= bias^2 + Var \\
&= \left(E\left[\frac{1}{k} \sum_{i \in \mathcal{N}_k(x)} m(x_i) + \epsilon_i - m(x) \right] \right)^2 + Var(\hat{m}(x)) \\
&= \left(E\left[\frac{1}{k} \sum_{i \in \mathcal{N}_k(x)} (m(x_i) - m(x)) \right] + \frac{1}{k} \sum_{i \in \mathcal{N}_k(x)} E(\epsilon_i) \right)^2 + \frac{\sigma^2}{k} \\
&= \left(\frac{1}{k} \sum_{i \in \mathcal{N}_k(x)} (m(x_i) - m(x)) \right)^2 + \frac{\sigma^2}{k}
\end{aligned}
$$

(b) If n points are evenly distributed over d-dimensional unit cube, the distance between any two closest points is $d_{min} = \frac{1}{\sqrt[d]{n}-1}$. In k nearest neibor, the maximal distance from the center point is at most $\sqrt{d}(\sqrt[d]{k}-1)d_{min}$. So $||x_i - x||_2 \leq \sqrt{d}(\sqrt[d]{k}-1)\frac{1}{\sqrt[d]{n}-1} \leq C(\frac{k}{n})^{1/d}$, where $C = \sqrt{d}$. The bias term becomes

$$
\begin{aligned}
\left(\frac{1}{k} \sum_{i \in \mathcal{N}_k(x)} (m(x_i) - m(x)) \right)^2 &\leq |m(x_i) - m(x)|^2 \\
&\leq (L||x_i - x||)^2 \\
&\leq \left(LC(\frac{k}{n})^{1/d} \right)^2
\end{aligned}
$$

Hence,

$$
E(\hat{m}(x) - m(x))^2 \leq (CL)^2(\frac{k}{n})^{2/d} + \frac{\sigma^2}{k}
$$

(c) Take the derivative of the above result w.r.t. k and set to 0, we have

$$
\frac{2(CL)^2}{dn^{2/d}} k^{2/d-1} - \frac{\sigma^2}{k^2} = 0 \Rightarrow k = \left(\frac{n^2 d^d \sigma^{2d}}{2^d (CL)^{2d}} \right)^{\frac{1}{d+2}}
$$

2. (a)

$$
\begin{aligned}
\hat{m}_{(-i)}(X_i) &= \sum_{j \neq i} Y_j \frac{K(X_i, X_j)}{\sum_{k \neq i} K(X_i, X_k)} = \sum_{j \neq i} Y_j \frac{K(X_i, X_j)}{\sum_k K(X_i, X_k) - K(X_i, X_i)} \\
\Longrightarrow \hat{m}_{(-i)}(X_i) \sum_k K(X_i, X_k) &= \hat{m}_{(-i)}(X_i) K(X_i, X_i) + \sum_{j \neq i} Y_j K(X_i, X_k) \\
\Longrightarrow \hat{m}_{(-i)}(X_i) &= \sum_{j=1}^{n} S_{ij} Z_j
\end{aligned}
$$

where

$$
Z_j = \begin{cases} Y_j & j \neq i \\ \hat{m}_{(-i)}(X_i) & j = i \end{cases}
$$

(b) Since $\hat{m}(X_i) = \sum_{j=1}^{n} Y_j S_{ij}$,

$$
\hat{m}(X_i) - \hat{m}_{(-i)}(X_i) = \sum_{j=1}^{n} Y_j S_{ij} - \sum_{j=1}^{n} S_{ij} Z_j = (Y_i - \hat{m}_{(-i)}(X_i)) S_{ii}
$$

Hence,

$$
Y_i - \hat{m}_{(-i)}(X_i)) = \frac{Y_i - \hat{m}(X_i))}{1 - S_{ii}}
$$

3. (a) The solution to $||Y - \beta X||^2$ is $\beta = (X^TX)^{-1}X^TY$. Since $X^TX = I$, $\beta = X^TY$.

(b)

$$\begin{aligned}
\frac{1}{2}||Y - X\beta||^2 + \lambda||\beta||_0 &= \frac{1}{2}(Y^TY - 2Y^TX\beta + \beta^TX^TX\beta) + \lambda||\beta||_0 \\
&= \frac{1}{2}[Y^T(I - XX^T)Y + ||\beta - X^TY||^2] + \lambda||\beta||_0 \\
&= \frac{1}{2}[Y^T(I - XX^T)Y] + \sum_{i=1}^{d}[\frac{1}{2}(\beta_i - v_i^TY)^2 + \lambda I(\beta_i \neq 0)]
\end{aligned}$$

Let S be the summation term in the above equation. If $\beta_i = 0$, $S = \frac{1}{2}(v_i^TY)^2$, otherwise, $S = \frac{1}{2}(\beta_i - v_i^TY)^2 + \lambda$.

4. (a)

$$P(X_i \notin C_j) = 1 - P(X_i \in C_j) = 1 - \int p(x)dV \leq 1 - cV_{C_j} = 1 - ch^d = 1 - \frac{c}{N}.$$

Therefore,

$$\begin{aligned}
P(\exists \text{ one cube that has no data}) &\leq \sum_{j=1}^{N} P(X_i \notin C_j \text{ for all } i) \\
&\leq \sum_{j=1}^{N}(1 - \frac{c}{N})^n \\
&= N(1 - \frac{c}{N})^n \\
&\leq N(e^{-c/N})^n \\
&= Ne^{-c(n/N)}
\end{aligned}$$

Because $n/N = nh^d \to \infty$, the above probability goes to 0. So the probability that there exists one cube with no data tends to be 0 as $n \to \infty$.

(b)

5. (a)

$$\begin{aligned}
E\left[\int_0^1 (\hat{m}(x) - m(x))^2 dx\right] &= E\left[\int_0^1 (\sum_{j=1}^{k}(\hat{\beta}_j - \beta_j)\psi_j(x) - \sum_{j=k+1}\beta_j\psi_j(x))^2 dx\right] \\
&= E\left[\sum_{j=1}^{k}(\hat{\beta}_j - \beta_j)^2 + \sum_{j=k+1}\beta_j^2\right] \\
&= \sum_{j=1}^{k}E(\hat{\beta}_j^2 - \beta_j^2) + \sum_{j=k+1}\beta_j^2
\end{aligned}$$

The first term is $MSE(\hat{\beta})$.

$$\begin{aligned}
bias(\hat{\beta}_j) &= E[\hat{\beta}_j] - \beta_j \\
&= E\left[\frac{1}{n}\sum_i \frac{Y_i\psi_j(X_i)}{p(X_i)}\right] - \beta_j \\
&= \frac{1}{n}E\left[E\left(\sum_i \frac{Y_i\psi_j(X_i)}{p(X_i)}|X_i\right)\right] - \beta_j \\
&= \frac{1}{n}\sum_i E\left[\frac{E(Y_i|X_i)\psi_j(X_i)}{p(X_i)}\right] - \beta_j \\
&= \frac{1}{n}\sum_i E\left[\frac{m(X_i)\psi_j(X_i)}{p(X_i)}\right] - \beta_j \\
&= \frac{1}{n}\sum_i \int_0^1 \frac{m(X_i)\psi_j(X_i)}{p(X_i)}p(X_i)dx - \beta_j \\
&= \beta_j - \beta_j = 0
\end{aligned}$$

$$
\begin{aligned}
Var(\hat{\beta}_j) &= Var\left(\frac{1}{n}\sum_i \frac{Y_i\psi_j(X_i)}{p(X_i)}\right) \\
&= \frac{1}{n^2}\sum_i Var\left(\frac{Y_i\psi_j(X_i)}{p(X_i)}\right) \\
&= \frac{1}{n^2}\sum_i \left(E\left[(\frac{Y_i\psi_j(X_i)}{p(X_i)})^2\right] - E^2\left[\frac{Y_i\psi_j(X_i)}{p(X_i)}\right]\right) \\
&= \frac{1}{n^2}\sum_i E\left[E\left((\frac{(m(X_i)+\epsilon_i)\psi_j(X_i)}{p(X_i)})^2|X_i\right)\right] - \frac{\beta_j^2}{n} \\
&= \frac{1}{n^2}\sum_i E\left[\frac{1}{p(X_i)}((m^2(X_i)+E(\epsilon_i^2|X_i))\psi_j^2(X_i)\right] - \frac{\beta_j^2}{n} \\
&\le \frac{1}{n^2}\sum_i E\left[\frac{1}{\inf p(x)}((m^2(X_i)+E(\epsilon_i^2|X_i))C^2\right] - \frac{\beta_j^2}{n} \\
&\le \frac{C^2}{n^2\inf p(x)}\sum_i \left(E[m^2(X_i)] + E[\epsilon_i^2|X_i]\right) \\
&= \frac{C^2}{n\inf p(x)}(||m||_2^2 + e^2)
\end{aligned}
$$

where $||m||_2^2 = E[m^2(X_i)]$ and $e^2 = E[\epsilon_i^2|X_i]$.

$$
\begin{aligned}
\sum_{j=k+1}\beta_j^2 &= \sum_{j=k+1}\left(\beta_j^2\frac{(k+1)^{2q}}{(k+1)^{2q}}\right) \\
&\le \frac{1}{(k+1)^{2q}}\sum_{j=k+1}\beta_j^2 j^{2q} \\
&\le \frac{C^2}{(k+1)^{2q}}
\end{aligned}
$$

Put them together, we get

$$
\begin{aligned}
E\left[\int_0^1(\hat{m}(x)-m(x))^2dx\right] &= \sum_{j=1}^{k}(bias^2(\hat{\beta}_j) + Var(\hat{\beta}_j)) + \sum_{j=k+1}\beta_j^2 \\
&\le \sum_{j=1}^{k}\frac{C^2}{n\inf p(x)}(||m||_2^2+e^2) + \frac{C^2}{(k+1)^{2q}} \\
&= \frac{kC^2}{n\inf p(x)}(||m||_2^2+e^2) + \frac{C^2}{(k+1)^{2q}}
\end{aligned}
$$

Take the derivative of the upper bound and set to 0, solve it, we get

$$
k^* = \left(\frac{2qn\inf p(x)}{||m||_2^2+e^2}\right)^{\frac{1}{2q+1}}
$$

.

(b) According to the assumptions, we know

$$
||\hat{p}-p||_\infty \to 0
$$

and

$$
r_n k_n \to 0
$$

The goal is to prove $|\hat{m}-m| \to 0$.
$||\hat{m}-m|| \le ||\hat{m}-m^*|| + ||m^*-m||$ where m^* is the estimator where density p is known. According to (a), $||m^*-m|| \to 0$ for large n, k.

$$
\begin{aligned}
||\hat{m} - m^*|| &= \sum_j ||\hat{\beta}_j - \beta_j^*|| \\
&= \sum_j ||\frac{1}{n}\sum_i Y_i \psi_j(X_i)(\frac{1}{\hat{p}(X_i} - \frac{1}{p(X_i)})|| \\
&\leq \sum_{i,j} \frac{1}{n^2}|Y_i|^2|\psi_j(X_i)|^2|\frac{\hat{p}(X_i) - p(X_i)}{\hat{p}(X_i)p(X_i)}|^2 \\
&\leq \sum_{i,j} \frac{1}{n^2}|Y_i|^2 C^2 \frac{(\sup|\hat{p} - p|)^2}{|\hat{p} \inf p|^2} \\
&= k|\bar{Y}|^2 C^2 \frac{||\hat{p} - p||_\infty^2}{|\hat{p} \inf p|^2} \\
&= |\bar{Y}|^2 C^2 \frac{O_P(r_n)k_n}{|\hat{p} \inf p|^2} \\
&= |\bar{Y}|^2 C^2 \frac{o_P(1)}{|\hat{p} \inf p|^2} \to 0
\end{aligned}
$$