Assignment 1

1. (a) i.

$$
\begin{aligned}
S_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu+\mu-\bar{X}_n)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2+\frac{2}{n}\sum_{i=1}^{n}(X_i-\mu)(\mu-\bar{X}_n)^2+\frac{1}{n}\sum_{i=1}^{n}(\mu-\bar{X}_n)^2 \\
(\bar{X}_n \xrightarrow{P} \mu) \xrightarrow{P} &\ \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 \\
&= \sigma^2
\end{aligned}
$$

ii. Because $S_n^2 \xrightarrow{P} \sigma^2$, we have $\frac{S_n^2}{\sigma^2} \xrightarrow{P} 1$.

$$
\frac{\sqrt{n}(\bar{X}_n-\mu)}{S_n} \xrightarrow{P} \frac{\sqrt{n}(\bar{X}_n-\mu)}{S_n} \times \sqrt{\frac{S_n^2}{\sigma^2}} \rightsquigarrow N(0,1)
$$

(b) i. Since $X_n = O_P(1)$, $\forall \epsilon > 0$, $\exists M > 0, N_1 > 0$, $n > N_1$ s.t. $P(|X_n| > M) \le \epsilon/2$. Since $Y_n = o_P(1)$, $\forall \epsilon > 0, \delta > 0$, $\exists N_2 > 0$, $n > N_2$, $P(|Y_n| > \delta/M) \le \epsilon/2$.
Then, $P(|X_nY_n| > \delta) \le P(|X_n| > M) + P(|Y_n| > \delta/M) \le \epsilon/2 + \epsilon/2 = \epsilon$.
So $X_nY_n = o_P(1)$.

ii. Let $M' = M + 1 > M + \delta$, then

$$
P(|X_n+Y_n| > M') \le P(|X_n| > M) + P(|Y_n| > \delta) \le P(|X_n| > M) + P(|Y_n| > 1) \le \epsilon
$$

. So $X_n + Y_n = O_P(1)$.

(c) i. $X_n = n$ with probability $\frac{1}{\sqrt{n}}$, 0 otherwise.

ii. Since $X_n = o_P(1)$, $\forall \epsilon > 0$, $\exists N > 0$, $n > N$ s.t. $P(|X_n| > \epsilon) \to 0$. And f is bounded and continuous, say the bound is M. Then

$$
\begin{aligned}
E(f(X_n)) &= \lim_{n\to\infty}\frac{1}{n}\left(\sum_{i=1}^{N} f(X_i) + \sum_{i=N+1}^{n} f(X_i)\right) \\
&\le \lim_{n\to\infty}\frac{1}{n}\left(NM + (n-N)f(0)\right) \\
&\to f(0)
\end{aligned}
$$

iii.

(d) i. $p(x) = \frac{1}{\theta}$ for $x \in [0,\theta]$, 0 otherwise. $\hat{\theta} = \max_i X_i$.

ii.

(e) i. Let $m = E(Y|X = x)$.

$$
\begin{aligned}
E(Y-m(X))^2 &= E(Y-m+m-m(X))^2 \\
&= E(Y-m)^2 + E(m-m(X))^2 + 2E[(Y-m)(m-m(X))]
\end{aligned}
$$

ii. Let $L = (Y - \beta^T X)^2$.

$$
\frac{dL}{d\beta} = E[2(\beta^T X - Y)X^T]
$$

And

$$
\frac{d^2L}{d\beta^2} = 2E(XX^T) = B
$$

Because $B = XX^T$ is positive semi-definite, L can be minimized.
Minimizing L means $\frac{dL}{d\beta} = 0$, which implies $\beta = B^{-1}\alpha$.

(f) i.

$$\begin{aligned} MSE(a) &= E(\hat{\mu}-\mu)^2 \\ &= E(a\bar{X}_n-\mu)^2 \\ &= E[a^2(\bar{X}_n-\mu)^2+(1-a^2)\mu^2-2a(1-a)\bar{X}_n\mu] \\ (X_i \sim N(\mu,1)) &= a^2/n+(1-a^2)\mu^2-2a(1-a)\mu^2 \\ &= a^2/n+(a-1)^2\mu^2 \end{aligned}$$

ii.

$$\begin{aligned} \frac{dMSE(a)}{da} &= 2a/n+2(a-1)\mu^2=0 \\ &\Rightarrow a=\frac{n\mu^2}{1+n\mu^2} \end{aligned}$$

$\therefore a_* = \frac{n\mu^2}{1+n\mu^2}, MSE(a_*) = \frac{\mu^2}{1+n\mu^2}$

iii. Because $\bar{X} \xrightarrow{P} \mu$,

$$M\hat{S}E(a) = a^2/n+(a-1)^2\bar{X}^2$$

Therefore,

$$\hat{a}_* = \frac{n\bar{X}^2}{1+n\bar{X}^2}, M\hat{S}E(a_*) = \frac{\bar{X}^2}{1+n\bar{X}^2}$$

2. Let $f=\sum_i \alpha_i K_{X_i} \in \mathscr{H}$. For any $g \in \mathscr{H}$, we can find $\hat{f}$ s.t. $\langle f-\hat{f}, g\rangle = 0$. Geometrically speaking, any vector can be orthogonal with any given vector after some projection.

$$\begin{aligned} ||f||_K^2 &= \langle \hat{f}+(f-\hat{f}), \hat{f}+(f-\hat{f})\rangle \\ &= \langle \hat{f},\hat{f}\rangle + \langle \hat{f}, f-\hat{f}\rangle + \langle f-\hat{f}, f-\hat{f}\rangle \\ &= ||\hat{f}||_K^2 + ||f-\hat{f}||_K^2 \\ &\geq ||\hat{f}||_K^2 \end{aligned}$$

So $\hat{f}$ is a minimizer of $\sum_{i=1}^n L(f(X_i),Y_i)+\lambda||f||_K^2$. Since $\hat{f} \in \mathscr{H}$, $\hat{f}$ has the form $\sum_i \alpha_i K_{X_i}$.

$$||\hat{f}||_K^2 = \sum_{i,j} \alpha_i \langle K_{X_i}, K_{X_j}\rangle \alpha_j = \alpha^T \mathbb{K}\alpha,$$

and let $Q(f)=\sum_{i=1}^n L(f,Y_i)$, then α minimizes $Q(\mathbb{K}\alpha)+\lambda\alpha^T\mathbb{K}\alpha$.

3. (a)

$$\forall f,g \in \mathscr{H}, |L_x f - L_x g| = |L_x(f-g)| \leq B||f-g||$$

So L_x is Lipschitz, hence it is continuous. $\mathscr{H}$ is RKHS.

(b)

$$\begin{aligned} |L_x f| &= \langle f, K_x\rangle \\ &\leq ||f||_K ||K_x||_K \\ &= \langle K_x, K_x\rangle ||f||_K \\ &\leq \sup_x K(x,x) ||f||_K \\ &= B||f||_K \end{aligned}$$

Set $B=\sup_x K(x,x)$.

4. If $K(x,y)=xy$, let $f=\sum_i \alpha_i K_{x_i}$, then $f(x)=\sum_i \alpha_i K_{x_i}(x) = \sum_i \alpha_i K(x_i,x) = (\sum_i \alpha_i x_i)x$. Because $\alpha_i \in \mathbb{R}, x_i \in [0,1]$, $\sum_i \alpha_i x_i \in \mathbb{R}$ and call it a. So we have constructed $f(x)=ax$, as desired.

5. Suppose the radius of the ball is ϵ. The diameter of B is $2\sqrt{d}r$. Then in each dimension, we need at most $\sqrt{d}r/\epsilon+1$. So the upper bound is $(\sqrt{d}r/\epsilon+1)^d$.

6.

$$\begin{aligned}
\mathbb{E}(|X|^k) &= \int_0^\infty 2P(X^k > \delta)d\delta \\
&= \int_0^\infty 2P(X > \delta^{\frac{1}{k}})d\delta \\
&\leq \int_0^\infty 2\exp(-t\delta^{\frac{1}{k}})\mathbb{E}(e^{tX})d\delta && \text{(Chernoff's method)} \\
&\leq \int_0^\infty 2\exp(ct^2 - t\delta^{\frac{1}{k}})d\delta && \text{(sub-Gaussian)} \\
&= \int_0^\infty 2\exp(-\delta^{2/k}/(4c))d\delta && (\text{pick } t = \frac{\delta^{1/k}}{2c}) \\
&= (4c)^{k/2}k\int_0^\infty \exp(-x)x^{k/2-1}dx \\
&= (4c)^{k/2}k\frac{k-2}{2}\cdots 1 \\
&\leq (4c)^{k/2}k(k/2-1)^{k/2-1}
\end{aligned}$$

So

$$(\mathbb{E}(|X|^k))^{1/k} \leq (4c)^{\frac{1}{2}}k^{1/k}(k/2-1)^{1/2} \leq (4c)^{\frac{1}{2}}e^{1/e}\sqrt{k} = C\sqrt{k}$$

7.

$$\begin{aligned}
\mathrm{Rad}_n(\mathscr{F}) &= \mathbb{E}_\sigma\left(\sup_{f\in\mathscr{F}}|\frac{1}{n}\sum_{i=1}^n \sigma_i f(Z_i)|\right) \\
&\leq \mathbb{E}_\sigma\left(\sum_{j=1}^n|\frac{1}{n}\sum_{i=1}^n \sigma_i f_j(Z_i)|\right) \\
&= \sum_{j=1}^n \mathbb{E}_\sigma\left(|\frac{1}{n}\sum_{i=1}^n \sigma_i f_j(Z_i)|\right) \\
&= \sum_{i=1}^n \mathrm{Rad}_n(|f_i|)
\end{aligned}$$

8. (i)

$$\begin{aligned}
\mathrm{Rad}_n(\alpha\mathscr{F}) &= \mathbb{E}_\sigma\left(\sup_{\alpha f\in\alpha\mathscr{F}}|\frac{1}{n}\sum_{i=1}^n \sigma_i \alpha f(Z_i)|\right) \\
&= \mathbb{E}_\sigma\left(|\alpha|\sup_{f\in\mathscr{F}}|\frac{1}{n}\sum_{i=1}^n \sigma_i f(Z_i)|\right) \\
&= |\alpha|\mathbb{E}_\sigma\left(\sup_{f\in\mathscr{F}}|\frac{1}{n}\sum_{i=1}^n \sigma_i f(Z_i)|\right) \\
&= |\alpha|\mathrm{Rad}_n(\mathscr{F})
\end{aligned}$$

(ii)

$$\begin{aligned}
\mathrm{Rad}_n(\mathscr{F}+\mathscr{G}) &= \mathbb{E}_\sigma\left(\sup_{f+g\in\mathscr{F}+\mathscr{G}}|\frac{1}{n}\sum_{i=1}^n \sigma_i (f+g)(Z_i)|\right) \\
&= \mathbb{E}_\sigma\left(\sup_{f\in\mathscr{F},g\in\mathscr{G}}|\frac{1}{n}\sum_{i=1}^n \sigma_i (f(Z_i)+g(Z_i))|\right) \\
&\leq \mathbb{E}_\sigma\left(\sup_{f\in\mathscr{F},g\in\mathscr{G}}|\frac{1}{n}\sum_{i=1}^n \sigma_i (f(Z_i))| + |\frac{1}{n}\sum_{i=1}^n \sigma_i (g(Z_i))|\right) \\
&\leq \mathbb{E}_\sigma\left(\sup_{f\in\mathscr{F}}|\frac{1}{n}\sum_{i=1}^n \sigma_i (f(Z_i))| + \sup_{g\in\mathscr{G}}|\frac{1}{n}\sum_{i=1}^n \sigma_i (g(Z_i))|\right) \text{(Subadditivity)} \\
&= \mathrm{Rad}_n(\mathscr{F}) + \mathrm{Rad}_n(\mathscr{G})
\end{aligned}$$

(iii) If $x > y$, $x \vee y = \frac{1}{2}(x+y+x-y) = x = \max(x,y)$. If $x < y$, $x \vee y = \frac{1}{2}(x+y+y-x) = y = \max(x,y)$. Hence,

$$x \vee y = \frac{1}{2}(x+y+|x-y|).$$

$$\begin{aligned}
\mathrm{Rad}_n(\mathscr{F} \vee \mathscr{G}) &= \mathbb{E}_\sigma \left(\sup_{f \vee g \in \mathscr{F} \vee \mathscr{G}} |\frac{1}{n} \sum_{i=1}^n \sigma_i (f \vee g)(Z_i)| \right) \\
&= \mathbb{E}_\sigma \left(\sup_{f, g \in \mathscr{F}, \mathscr{G}} |\frac{1}{n} \sum_{i=1}^n \sigma_i \frac{1}{2}(f + g + |f-g|)(Z_i)| \right) \\
&\leq \frac{1}{2}\mathbb{E}_\sigma \left(\sup_{f \in \mathscr{F}} |\frac{1}{n} \sum_{i=1}^n \sigma_i f(Z_i)| \right) + \frac{1}{2}\mathbb{E}_\sigma \left(\sup_{g \in \mathscr{G}} |\frac{1}{n} \sum_{i=1}^n \sigma_i g(Z_i)| \right) \\
&\quad + \frac{1}{2}\mathbb{E}_\sigma \left(\sup_{f, g \in \mathscr{F}, \mathscr{G}} |\frac{1}{n} \sum_{i=1}^n \sigma_i (|f-g|)(Z_i)| \right) \\
&\leq \frac{1}{2}\mathrm{Rad}_n(\mathscr{F}) + \frac{1}{2}\mathrm{Rad}_n(\mathscr{G}) + \frac{1}{2}\mathbb{E}_\sigma \left(\sup_{f \in \mathscr{F}} |\frac{1}{n} \sum_{i=1}^n \sigma_i f(Z_i)| + \sup_{g \in \mathscr{G}} |\frac{1}{n} \sum_{i=1}^n \sigma_i g(Z_i)| \right) \\
&= \frac{1}{2}\mathrm{Rad}_n(\mathscr{F}) + \frac{1}{2}\mathrm{Rad}_n(\mathscr{G}) + \frac{1}{2}\mathrm{Rad}_n(\mathscr{F}) + \frac{1}{2}\mathrm{Rad}_n(\mathscr{G}) \\
&= \mathrm{Rad}_n(\mathscr{F}) + \mathrm{Rad}_n(\mathscr{G})
\end{aligned}$$