Homework 4

1. • First, recall that the RKHS of the linear kernel on $\mathbb{R}$ $K(a,b) = ab$ is the set of linear functions of the form $f_w(x) = wx$ for $w \in \mathbb{R}$
endowed with the inner product: $\forall a,b \in \mathbb{R}$ $\langle f_b, f_a \rangle_{\mathcal{H}} = ba$
and the norm $\forall a \in \mathbb{R}$, $\|f_a\|_{\mathcal{H}} = a^2$

• Therefore, for $f,g \in \mathcal{H}$, $\exists\, a,b \in \mathbb{R}$, $f = f_a$, $g = f_b$
and
$$\mathrm{cov}_m(f(X), g(Y)) = \mathbb{E}_n(aX\, bY) - \mathbb{E}_n(aX)\, \mathbb{E}_n(bY)$$
$$= \frac{1}{m}\sum_{i=1}^{m} aX_i\, bY_i - \left(\frac{1}{m}\sum_{i=1}^{m} aX_i\right)\left(\frac{1}{m}\sum_{i=1}^{m} bY_i\right)$$
$$= \frac{ab}{m}\left[\sum_{i=1}^{m} X_i Y_i - \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{m} X_i Y_j\right]$$
$$= \underline{\frac{ab}{m}\left[X^t Y - X^t J Y\right]}, \text{ where } J = \begin{bmatrix} 1/n & 1/n & \cdots & 1/n \\ \vdots & \vdots & & \vdots \\ 1/n & 1/n & \cdots & 1/n \end{bmatrix}$$

• $\mathcal{B}_K$ is the unit ball of the RKHS of K, so it reads
$$\mathcal{B}_K = \{ f \in \mathcal{H},\ \|f\|_{\mathcal{H}} \leq 1 \}$$
$$= \{ f_a,\ a \in \mathbb{R},\ a^2 \leq 1 \}$$
$$= \{ f_a,\ a \in \mathbb{R},\ |a| \leq 1 \}$$
So $\forall f,g \in \mathcal{B}_K$, $\exists\, a,b \in \mathbb{R}$, $f = f_a$, $g = f_b$ and $|a| \leq 1, |b| \leq 1$
and
$$\mathrm{cov}_m(f(X), g(Y)) \leq \frac{|ab|}{m} | X^t Y - X^t J Y |$$
$$\leq \frac{1}{m} | X^t Y - X^t J Y |$$
so
$$\underline{C_m^K(X,Y) \leq \frac{1}{m} | X^t Y - X^t J Y |}$$

• and for $a = 1$ and $b = \mathrm{sgn}(X^t Y - X^t J Y)$, one has that:

$$\mathrm{cov}_n(f_a(X), f_b(Y)) = \frac{|X^tY - X^tJY|}{m}, \text{ with } f_a, f_b \in \mathcal{B}_K$$

so $\underline{C_m^K(X,Y) \geq \frac{1}{m}|X^tY - X^tJY|}$

- Finally: $\boxed{C_m^K(X,Y) = \frac{1}{m}|X^tY - X^tJY|}$ $\left(\text{with } J = \begin{pmatrix} 1/n & \cdots & 1/n \\ \vdots & & \vdots \\ 1/n & \cdots & 1/n \end{pmatrix}\right)$

2. Now, let K a general kernel.

- First, we show that the representer theorem holds and that $\left(\max_{f,g \in \mathcal{B}_K} \mathrm{cov}_n(f(X), g(Y))\right)$ solutions admit representations of the form
$$f = \sum_{i=1}^{m} \alpha_i K_{x_i} \quad \text{where } \alpha_i, \beta_i \in \mathbb{R} \quad \forall i \in [n]$$
$$g = \sum_{i=1}^{m} \beta_i K_{y_i}$$

Let $\mathcal{H}_s^X = \{f \in \mathcal{H},\ f = \sum_{i=1}^{m} \alpha_i K_{x_i},\ (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n\}$
$\underline{\mathcal{H}_s^Y = \{g \in \mathcal{H},\ g = \sum_{i=1}^{m} \beta_i K_{y_i},\ (\beta_1, \dots, \beta_n) \in \mathbb{R}^n\}}$

The $\mathcal{H}_s$ are finite-dimensional subspaces, therefore any $f, g \in \mathcal{H}$ can be uniquely decomposed as $f = f_s + f_\perp$, with $f_s \in \mathcal{H}_s^X$
$g = g_s + g_\perp$, with $g_s \in \mathcal{H}_s^Y$

$\mathcal{H}$ being a RKHS, it holds that $\forall i \in [n]$,
$f_\perp(x_i) = \langle f_\perp, K_{x_i} \rangle_{\mathcal{H}} = 0 \Rightarrow f(x_i) = f_s(x_i)$
$g_\perp(y_i) = \langle g_\perp, K_{y_i} \rangle_{\mathcal{H}} = 0 \quad g(y_i) = g_s(y_i)$

Pythagoras' theorem in $\mathcal{H}$ then shows that
$\|f\|_{\mathcal{H}}^2 = \|f_s\|_{\mathcal{H}}^2 + \|f_\perp\|_{\mathcal{H}}^2$ and $\|g\|_{\mathcal{H}}^2 = \|g_s\|_{\mathcal{H}}^2 + \|g_\perp\|_{\mathcal{H}}^2$

Therefore, $\forall f, g \in \mathcal{H}$, $\exists f_s \in \mathcal{H}_s^X$, $g_s \in \mathcal{H}_s^Y$ such that

- $\|f_s\|_{\mathcal{H}} \leq \|f\|_{\mathcal{H}}$ and $\|g_s\| \leq \|g\|_{\mathcal{H}}$
- for $i = 1, \dots, n$, $f(x_i) = f_s(x_i)$ and $g(y_i) = g_s(y_i)$

so $\text{cov}_n(f(X), g(Y)) = \text{cov}_n(f_s(X), g_s(Y))$

This implies that for any couple $(f^*, g^*) \in \mathcal{H}$ optimal solution of $\left(\max\limits_{f,g \in \mathcal{B}_\mathcal{H}} \text{cov}_n(f(X), g(Y))\right)$, one can find $(f_s^*, g_s^*) \in \mathcal{H}_s^X \times \mathcal{H}_s^Y$ such that: $\bullet \|f_s^*\|_{\mathcal{H}} \leq \|f^*\|_{\mathcal{H}} \leq 1$ and $\|g_s^*\|_{\mathcal{H}} \leq \|g_s\|_{\mathcal{H}} \leq 1$, so $f_s^*, g_s^* \in \mathcal{B}_\mathcal{H}$

$\bullet$ $\text{cov}_n(f_s^*(X), g_s^*(Y)) = \text{cov}_n(f^*(X), g^*(Y))$

so (f_s^*, g_s^*) is also solution of $\left(\max\limits_{f,g \in \mathcal{B}_\mathcal{H}} \text{cov}_n(f(X), g(Y))\right)$

Hence, representer theorem holds, and we can restrict to solutions of the form

$$f = \sum_{i=1}^n \alpha_i K_{x_i}, \quad \alpha, \beta \in \mathbb{R}^n$$
$$g = \sum_{i=1}^n \beta_i K_{y_i}$$

- Now, we can rewrite the maximization problem as:

$$\boxed{\begin{array}{ll} \max & \text{cov}_n(f(X), g(Y)) \\ f, g \in \mathcal{B}_\mathcal{H} & \qquad\qquad (*) \\ f, g \in \mathcal{H}_s^X \times \mathcal{H}_s^Y & \end{array}}$$

Let K_X the Gram matrix of X, and K_Y the one of Y

We know that for any $f \in \mathcal{H}_s^X$ and α such that $f = \sum_{i=1}^n \alpha_i K_{x_i}$, one has $f(X) = K_X \alpha$ and $\|f\|_{\mathcal{H}}^2 = \alpha^T K_X \alpha$

(and the same holds for y, K_y and β)

Therefore $(*) \iff \max\limits_{\substack{\alpha^T K_x \alpha \leq 1 \\ \beta^T K_y \beta \leq 1}} \mathrm{cov}_n(K_x \alpha, K_y \beta)$

Now, $\underline{\mathrm{cov}_n(K_x \alpha, K_y \beta)} = \frac{1}{n} \sum_{i=1}^{n} [K_x \alpha]_i [K_y \beta]_i - \frac{1}{n^2} \sum_{i=1}^{n} [K_x \alpha]_i \sum_{j=1}^{n} [K_y \beta]_j$

$$= \frac{1}{n} \left[(K_x \alpha)^T K_y \beta - (K_x \alpha)^T J (K_y \beta) \right], \text{ where } J = \begin{pmatrix} 1/n & \cdots & 1/n \\ \vdots & & \vdots \\ 1/n & \cdots & 1/n \end{pmatrix}$$

$$= \underline{\frac{1}{n} \alpha^T K_x (I_n - J) K_y \beta}$$

Recalling that K_x and K_y are positive semi definite, they admit square root matrices that are also positive semi definite. Thus,

$$\mathrm{cov}_m(K_x \alpha, K_y \beta) = \frac{1}{m} (K_x^{1/2} \alpha)^T K_x^{1/2} (I_n - J) K_y^{1/2} (K_y^{1/2} \beta)$$

and $(*) \iff \max\limits_{\substack{\|K_x^{1/2} \alpha\|_2 \leq 1 \\ \|K_y^{1/2} \beta\|_2 \leq 1}} \frac{1}{m} (K_x^{1/2} \alpha)^T K_x^{1/2} (I_n - J) K_y^{1/2} (K_y^{1/2} \beta)$

↳ if α, β are solution of the maximization problem, then $(\alpha', \beta') = (K_x^{1/2} \alpha, K_y^{1/2} \beta)$ are solutions of the problem:

$$\max\limits_{\|\alpha'\|_2, \|\beta'\|_2 \leq 1} \frac{1}{n} \alpha'^T K_x^{1/2} (I_n - J) K_y^{1/2} \beta'$$

↳ if α', β' are solution of the above problem, we can write $\alpha' = K_x^{1/2}\alpha + \alpha_0$, where $\alpha_0 \in \mathrm{Ker}(K_x^{1/2})$, $\alpha \in \mathbb{R}^n$

$\beta' = K_y^{1/2}\beta + \beta_0$ $\quad\beta_0 \in \mathrm{Ker}(K_y^{1/2})$, $\beta \in \mathbb{R}^n$

because $K_x^{1/2}$ and $K_y^{1/2}$ are positive semi definite matrices (and consequently $\mathbb{R}^n = \mathrm{Ker}(K_x^{1/2}) \oplus \mathrm{Im}(K_x^{1/2})$)

We also have $\begin{cases} K_x^{1/2}\alpha \perp \alpha_0 \\ K_y^{1/2}\beta \perp \beta_0 \end{cases}$ and therefore, by

Pythagoras' theorem, $\begin{cases} \|\alpha'\|_2 = \|K_x^{1/2}\alpha\|_2 + \|\alpha_0\|_2 \geqslant \|K_x^{1/2}\alpha\| \\ \|\beta'\|_2 = \|K_y^{1/2}\beta\|_2 + \|\beta_0\|_2 \geqslant \|K_y^{1/2}\beta\| \end{cases}$

Moreover, since $\alpha_0 \in \mathrm{Ker}\, K_x^{1/2}$ $\beta_0 \in \mathrm{Ker}\, K_y^{1/2}$, we have

$$\frac{1}{m}\alpha'^T K_x^{1/2}(I_n - J)K_y^{1/2}\beta' = \frac{1}{m}(K_x^{1/2}\alpha)^T K_x^{1/2}(I_n - J)K_y^{1/2}(K_y^{1/2}\beta)$$

Finally, we have proved that:

$$(*) \iff \max_{\|\alpha'\|_2, \|\beta'\|_2 \leqslant 1} \frac{1}{m}\alpha'^T \underbrace{K_x^{1/2}(I_n - J)K_y^{1/2}}_{M}\beta'$$

- We have
$$C_n^k(X,Y) = \sup_{\|\alpha'\|_2 \leqslant 1}\ \sup_{\|\beta'\|_2 \leqslant 1} \frac{1}{m}\alpha'^T M\beta'$$
$$= \sup_{\|\alpha'\|_2 \leqslant 1} \frac{1}{m}\alpha'^T M \frac{\alpha'^T M}{\|\alpha'^T M\|_2}$$
$$= \frac{1}{m}\sup_{\|\alpha'\|_2 \leqslant 1} \|\alpha'^T M\|_2 = \frac{1}{m}\|M\|_2$$

Therefore, $\boxed{C_n^k(X,Y) = \frac{1}{m}\|K_x^{1/2}(I_n - J)K_y^{1/2}\|_2}$

where $J = \begin{pmatrix} 1/n & \cdots & 1/n \\ \vdots & & \vdots \\ 1/n & \cdots & 1/n \end{pmatrix}$ and K_x, K_y are the

Gram matrices of X and Y.