به نام خدا

مسئله ۱:

.۱

$K(x,y) = \frac{1}{1-xy} = \sum_{k=0} (xy)^k$ with $\mathcal{X} = \{-1,1\}$ **is a p.d. kernel** as limit of a sum of restricted (to $\{-1,1\}^2$) polynomial kernels.

.۲

$K(x,y) = \log(1+xy)$ with $\mathcal{X} = \mathbb{R}_+$ **is not a p.d. kernel**.
For instance, let us consider the two points $x = 1$ and $y = 2$. The similarity matrix is $[K] = \begin{bmatrix} \log(1+1) & \log(1+2) \\ \log(1+2) & \log(1+4) \end{bmatrix} = \begin{bmatrix} \log 2 & \log 3 \\ \log 3 & \log 5 \end{bmatrix}$ and we have $(-2 \quad 1)\,[K] \begin{pmatrix} -2 \\ 1 \end{pmatrix} = 4\log 2 - 4\log 3 + \log 5 = \log \frac{5\times 2^4}{3^4} = \log \frac{80}{81} < 0$.

.۳

$K(x,y) = \cos(x+y)$ with $\mathcal{X} = \mathbb{R}$ **is not a p.d. kernel**.
For instance, let us consider the two points $x = \frac{\pi}{2}$ and $y = 0$. The similarity matrix is $[K] = \begin{bmatrix} \cos(\pi) & \cos(\frac{\pi}{2}) \\ \cos(\frac{\pi}{2}) & \cos(0) \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ and we have $(1 \quad 0)\,[K] \begin{pmatrix} 1 \\ 0 \end{pmatrix} = -1 < 0$.

.۴

$K(x,y) = \min(x,y)$ with $\mathcal{X} = \mathbb{R}_+$ **is a p.d. kernel**.
First, let us define $\Phi : \mathbb{R} \mapsto L^2(\mathbb{R}_+)$ that maps a real number x, to the square-integrable step function $\Phi(x) = \mathbb{1}_{[0,x]}$ (i.e., $t \mapsto 1$ if $t \leq x$, 0 otherwise). Two interesting properties of these step functions are that for any two real non-negative numbers a and b we have $\int_0^\infty \Phi(a) = \int_0^a 1 = a$ and – since $[0,a] \cap [0,b] = [0, \min(a,b)]$ – $\Phi(\min(a,b)) = \Phi(a)\Phi(b)$; thus we have:

$$\begin{aligned} \min(a,b) &= \int_0^{+\infty} \mathbb{1}_{[0,a]} \mathbb{1}_{[0,b]} \\ &= \langle \Phi(a), \Phi(b) \rangle_{L^2(\mathbb{R}_+)}. \end{aligned}$$

Applying Aronszajn's theorem to Φ then proves that K is a p.d. kernel.

مسئله ۳:

.۱

Since the regular product on $\mathbb{R}$ is a scalar product, the RKHS for the linear kernel is $\mathbb{R}$ with an identity embedding. As such, we have, for $f, g \in \mathbb{R}$:

$$\begin{aligned}
\mathrm{cov}_n(f(X), g(X)) &= \mathbb{E}_n[fXgY] - \mathbb{E}_n[fX]\mathbb{E}_n[gY] \\
&= \frac{1}{n}\sum_{i=1}^{n} fX_i gY_i - (\frac{1}{n}\sum_{i=1}^{n} fX_i)(\frac{1}{n}\sum_{i=1}^{n} gY_i) \\
&= \frac{fg}{n}\left(\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{j=1}^{n} \frac{1}{n} Y_j\right) \\
&= \frac{fg}{n}\left(X^T Y - X^T U Y\right).
\end{aligned}$$

Since the constraints $f, g \in \mathcal{B}_K$ mean that $|f| \leq 1$ and $|g| \leq 1$ in this simple case, we then deduce that:

$$\begin{aligned}
C_n^K(X, Y) &= \max_{f,g \in \mathcal{B}_K} \frac{fg}{n} X^T (I - U) Y \\
&= \frac{|X^T (I - U) Y|}{n}.
\end{aligned}$$

.٢

Let's first show that we can restrict ourselves to f and g with representations of form $f = \sum_{i=1}^{n} F_i K_{X_i}$ and $g = \sum_{i=1}^{n} G_i K_{Y_i}$. Indeed, suppose that we have a solution (f^*, g^*) for the maximization problem defining C_n^K; then f^* is also solution of the maximization problem $\max_{f \in \mathcal{B}_K} \mathrm{cov}_n(f(X), g^*(Y))$. Since $\mathrm{cov}_n(f(X), g^*(Y)) = \frac{1}{n}\sum_{i=1}^{n} \langle f, X_i \rangle g^*(Y_i) - \frac{1}{n^2}\sum_{i,j=1}^{n} \langle f, X_i \rangle g^*(Y_i)$ is linear in f, this optimization problem is a convex optimization problem in f for which strong duality holds (take $f = 0$ to check for Slater's condition).

Since the dual problem satisfies the conditions of the representer theorem, we conclude that f^* admits a representation of the aforementioned form. Using an f^* with this form, we apply the same reasoning to g^* to obtain an optimal pair (f^*, g^*) where both f^* and g^* have the aforementioned forms.

If we design by F the vector $(F_1, \ldots, F_n)$, we have that $f(X_i) = [K_X F]_i$ and $\|f\|^2 = F^T K_X F$ (and similar relations for G, g and Y, mutatis mutandi); thus we can write:

$$\begin{aligned}
\mathrm{cov}_n(X, Y) &= \frac{1}{n}\sum_{i=1}^{n} f(X_i) g(Y_i) - \frac{1}{n}\sum_{i=1}^{n} f(X_i) \frac{1}{n}\sum_{j=1}^{n} g(Y_j) \\
&= \frac{1}{n}\sum_{i=1}^{n} [K_X F]_i [K_Y G]_i - \frac{1}{n}\sum_{i=1}^{n} [K_X F]_i \sum_{j=1}^{n} \frac{1}{n} [K_Y G]_j \\
&= \frac{1}{n}\left((K_X F)^T K_Y G - (K_X F)^T U K_Y G\right) \\
&= \frac{1}{n} F^T K_X (I - U) K_Y G.
\end{aligned}$$

So we can rewrite $n \times C_n^K$ as the solution of the maximization of $F^T K_X (I - U) K_Y G$ subject to $F^T K_X F \leq 1$ and $G^T K_Y G \leq 1$. Recalling that K_X and K_Y are positive semi-definite matrices, and as such admit a positive semi-definite square root, we can rewrite this again as the maximization of $(K_X^{1/2}F)^T K_X^{1/2}(I-U)K_Y^{1/2}(K_Y^{1/2}G)$, subject to $\|K_X^{1/2}F\| \leq 1$ and $\|K_Y^{1/2}G\| \leq 1$.

We now claim that this is equivalent to the maximization of $\tilde{F}^T K_X^{1/2}(I-U)K_Y^{1/2}\tilde{G}$, subject to $\|\tilde{F}\| \leq 1$ and $\|\tilde{G}\| \leq 1$. This is trivial when $K_X^{1/2}$ and $K_Y^{1/2}$ are invertible; however the general case requires more care. We have:

- For any F, G such that $\|K_X^{1/2}F\| \leq 1$ and $\|K_Y^{1/2}G\| \leq 1$, we can define $\tilde{F} = K_X^{1/2}F$ and $\tilde{G} = K_Y^{1/2}G$ satisfying $\|\tilde{F}\| \leq 1$ and $\|\tilde{G}\| \leq 1$ and such that $\tilde{F}^T K_X^{1/2}(I-U)K_Y^{1/2}\tilde{G} = F^T K_X(I-U)K_Y G$.
- Conversely, recall that, as real-valued symmetric matrices, $K_X^{1/2}$ and $K_Y^{1/2}$ are diagonalizable in an orthogonal basis. As such, for any $\tilde{F}, \tilde{G}$ such that $\|\tilde{F}\| \leq 1$ and $\|\tilde{G}\| \leq 1$, we can write $\tilde{F} = K_X^{1/2}F + k_F$ and $\tilde{G} = K_Y^{1/2}G + k_G$ for some vectors F, G, k_F and k_G such that $K_X^{1/2}k_F = K_Y^{1/2}k_G = 0$, and $\langle k_F, K_X^{1/2}F\rangle = \langle k_G, K_Y^{1/2}G\rangle = 0$. By orthogonality, we have $\|K_X^{1/2}F\| = \|\tilde{F}\| - \|k_F\| \leq 1 - \|k_F\| \leq 1$ and similarly $\|K_Y^{1/2}G\| \leq 1$; moreover $\tilde{F}^T K_X^{1/2}(I-U)K_Y^{1/2}\tilde{G} = (K_X^{1/2}(K_X^{1/2}F + k_F))^T(I-U)K_Y^{1/2}(K_Y^{1/2}G + k_G) = F^T K_X(I-U)K_Y G$.

The, these two optimization problems are indeed equivalent, and we have

$$\begin{aligned} n \times C_n^K(X,Y) &= \max_{\|\tilde{F}\| \leq 1, \|\tilde{G}\| \leq 1} \tilde{F}^T K_X^{1/2}(I-U)K_Y^{1/2}\tilde{G}, \\ &= \max_{\|\tilde{G}\| \leq 1} \max_{\|\tilde{F}\| \leq 1} \tilde{F}^T K_X^{1/2}(I-U)K_Y^{1/2}\tilde{G}. \end{aligned}$$

Considering a fixed $\tilde{G}$; if we define $M_G = K_X^{1/2}(I-U)K_Y^{1/2}\tilde{G}$, $\tilde{F}$ is a solution of $\arg\max_{\|\tilde{F}\| \leq 1} \tilde{F}^T M_G$. This is simply the maximization of a scalar product on the unit ball, reached on $\tilde{F} = \frac{M_G}{\|M_G\|}$. Plugging this back in, we get

$$n \times C_n^K(X,Y) = \max_{\|\tilde{G}\| \leq 1} \tilde{F}^T M_G = \max_{\|\tilde{G}\| \leq 1} \frac{M_G^T M_G}{\|M_G\|} = \max_{\|\tilde{G}\| \leq 1} \|K_X^{1/2}(I-U)K_Y^{1/2}\tilde{G}\|.$$

We recognize the spectral norm, and conclude that:

$$C_n^K(X,Y) = \frac{1}{n}\|K_X^{1/2}(I-U)K_Y^{1/2}\|_2.$$

This concludes the problem 3.

مسئله‌ی ۲. توابع کرنل (۱۰ نمره)

فرض کنید $\mathcal{X}$ فضای نمونه، $\mathcal{H}$ فضای هیلبرت و $\Phi : \mathcal{X} \to \mathcal{H}$ یک نگاشت باشد. تابع $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ را نیز تابع کرنل این فضای هیلبرت در نظر بگیرید.

نقاط آموزشی داده‌شده $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$ را در نظر بگیرید و فضای برچسب‌ها را به صورت $y_i \in \{+1, -1\}$ فرض کنید. نقطه‌ی c_y را به صورت زیر تعریف می‌کنیم:

نگاشت x ها به بعد بالای $\mathcal{H}$

$$\boxed{c_y = \frac{1}{m_y} \sum_{i : y_i = y} \Phi(x_i)} \quad (1)$$

که در آن $m_y = |\{i : y_i = y\}|$. فرض کنید m_{+1}, m_{-1} هر دو ناصفر باشند. الگوریتم زیر را در نظر بگیرید:

الگوریتم تصمیم گیری برای نقطه جدید

$$\boxed{h(x) = \begin{cases} +1 & \|\Phi(x) - c_+\| \leqslant \|\Phi(x) - c_-\| \\ -1 & \text{otherwise} \end{cases}}$$

۱. با تعریف $\mathbf{w} = c_+ - c_-$ و $b = \frac{1}{2}(\|c_-\|^2 - \|c_+\|^2)$ نشان دهید:

$$h(x) = \operatorname{sign}(\langle \mathbf{w}, \Phi(x) \rangle + b).$$

۲. روشی برای محاسبه‌ی $h(x)$ بر مبنای تابع کرنل ارائه دهید.

۱. برای حل این بخش نیاز است c_+ و c_- را محاسبه کنیم (با استفاده از تعریف)

$$\|\Phi(x) - c_+\|^2 \underset{\text{استفاده از رابطه (1)}}{=} \Big\|\Phi(x) - \frac{1}{m_+} \sum_{i \in y_i = +1} \Phi(x_i)\Big\|^2 = \|\Phi(x)\|^2 + \|c_+\|^2 - \frac{2}{m_+} \Big\langle \Phi(x) . \sum_{i \in y_i = +1} \Phi(x_i) \Big\rangle$$

مشابه همین عملیات ریاضی برای c_- داریم و سپس:

$$\|\Phi(x) - c_+\|^2 - \|\Phi(x) - c_-\|^2 = \|c_+\|^2 - \|c_-\|^2 + \frac{2}{m_-} \Big\langle \Phi(x) . \sum_{i \in y_i = -1} \Phi(x_i) \Big\rangle - \frac{2}{m_+} \Big\langle \Phi(x) . \sum_{i \in y_i = +1} \Phi(x_i) \Big\rangle$$

$$= \|c_+\|^2 - \|c_-\|^2 - 2 \langle \Phi(x) . c_+ - c_- \rangle$$

بنابراین قاعده تصمیم گیری با اختیار کردن $w = c_+ - c_-$ و $b = \frac{1}{2}(\|c_-\|^2 - \|c_+\|^2)$ به یک طبقه بند خطی به شکل روبرو تبدیل می‌گردد.

$$h(x) = \operatorname{sign}\{(w . \Phi(x) + b\}$$

برای بخش دوم به طرق مختلف می‌توان استدلال کرد برای نمونه می‌توان با محاسبه w که حاصل تفریق دو عبارتی است که از روی تعریف بدست می‌آید که در آن شاهد ترم $\phi(x_i)$ ها خواهیم بود و بعد ضرب آنها در $\phi(x)$ که به صورت ضرب داخلی $\phi(x_i).\phi(x)$ ظاهر می‌شود که می‌شود نگارش $h(x)$ برحسب توابع کرنل $K(x_i, x) = \langle \phi(x_i).\phi(x) \rangle$

مسئله‌ی ۴. تشخیص کرنل‌ها (۱۰ نمره)

- **بخش ۱**
  فضای احتمال $(\Omega, \mathcal{F}, \mathbb{P})$ را در نظر بگیرید. نشان دهید که تابع زیر که روی $\mathcal{F} \times \mathcal{F}$ تعریف شده، یک کرنل PDS است.

$$K(A,B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)$$

حل: برای تشخیص کرنل روش‌های مختلفی وجود دارند اما معمولاً روش راحت‌تر این است با ضرب داخلی‌هایی که بلدید بتوانید $K(x,y) = \langle \phi(x) \cdot \phi(y) \rangle_H$ بنویسید و سپس استدلال کنید که K یک کرنل است.

استدلال می‌کنیم که $K(A,B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)$ یک کرنل است چگونه می‌توان به صورت شکل زیر آن را بیان کرد $K(A,B) = \langle \Phi_A \cdot \Phi_B \rangle_H$. اما چگونه؟ می‌شود Φ_A را به صورت زیر تعریف کرد:

$$\Phi_A = I_A(\circ) - \mathbb{P}(A), \quad A \in \mathcal{F}$$

$\boxed{I_A(\circ) \text{ is an indicator function}}$

و سپس ضرب داخلی را نیز به صورت زیر در نظر می‌گیرید:

$$\langle \Phi_A \cdot \Phi_B \rangle = \int \Phi_A(s)\Phi_B(s)\, d\mathbb{P}(s)$$

سپس برای هر انتخاب A, B می‌توان نوشت که:

$$\langle \Phi_A \cdot \Phi_B \rangle = \int \left[I_A(s) - \mathbb{P}(A) \right] \cdot \left[I_B(s) - \mathbb{P}(B) \right] d\mathbb{P}(s)$$

$$= \int I_A(s) I_B(s)\, d\mathbb{P}(s) - \mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)$$

- **بخش ۲**

S را یک مجموعه‌ی متناهی در نظر بگیرید. نشان دهید که تابع زیر که $P(A)$ در آن مجموعه توانی مجموعه‌ی A است، یک کرنل PSD است.

$$\begin{cases} K : P(S) \times P(S) \to \mathbb{R} \\ K(A,B) = ۲^{|A \cap B|} \end{cases}$$

برای اثبات این که ادعای بالا یک کرنل PSD است می‌توان به شکل بخش ۱ در نظر گرفت که $K(A,B) = \langle \phi_A \cdot \phi_B \rangle_H$. می‌توان یک A دلخواه در یک مجموعه توانی دلخواه در نظر گرفت $A \in \mathbb{P}(S)$ و سپس A عضو مجموعه توانی را می‌توان نمایش داد به $\phi_A = \mathbb{I}_{\mathbb{P}(A)}(a)$ و ضرب داخلی را نیز به صورت زیر تعریف کنید.

$$\langle \phi_A \cdot \phi_B \rangle = \sum_{a \in \mathbb{P}(S)} \phi_A(a)\, \phi_B(a)$$

تعبیر عملکرد این معادله نیز به صورت زیر است که بررسی می‌کند که آیا a دلخواهی آیا در مجموعه توانی A و B هم زمان هست؟ در حقیقت بررسی می‌کند که a دلخواه زیر مجموعه A و B هست یا نه اگر هست ضرب داخلی برابر یک است اگر نیست برابر صفر است، پس $\sum_{a \in \mathbb{P}(S)} \phi_A(a)\phi_B(a)$ داره تعداد a هایی که هم زیرمجموعه‌ی A و B هستند را می‌شمارد که برابر می‌شود با $2^{|A \cap B|}$

- **بخش ۳**

تابع زیر را روی $\mathbb{R}^n \times \mathbb{R}^n$ در نظر بگیرید. قصد داریم PDS بودن این کرنل را نشان دهیم.

$$K_\alpha(\mathbf{x}, \mathbf{x}') = \sum_{k=۱}^{N} \min(|x_k|^\alpha, |x'_k|^\alpha)$$

ابتدا باید نشان دهیم که $\min(x,y)$ یک کرنل است و سپس از ویژگی های توابع کرنل بهره می‌بریم.

در این بخش مشابه بخش های قبلی می‌توان هر $a \in \mathbb{R}$ نمایشی را در نظر گرفت که بررسی می‌کند ورودی برای نمایش کوچکتر یا مساوی x یا y هست یا نه؟

$a \in R$

$\Phi_a(o) = I_{x \leq a}(o)$

$$\langle f, g \rangle = \int_0^\infty f(t) g(t)\, dt$$

سپس با توجه به انتگرال مدنظر که بررسی می‌کند که آیا t کوچکتر یا مساوی a و b هست یا نه

که در این صورت حاصل می‌شود $\min(x,y)$

$$\langle I_{t \leq b} \cdot I_{t \leq a} \rangle = \min(a, b)$$

که بخش می‌گویند Sobolev-like space.

حال که اثبات شد $K(x,y) = \min(x,y)$ یک کرنل است پس با استفاده از قضیه

اینکه جمع کرنل‌ها یک کرنل است می‌توان به ازای یک a ثابت بیان کرد که

$$K(x, x') = \sum_{K=1}^{N} \min\{|x_k|, |x'_k|\}$$

یک کرنل است.

۲. به کمک بخش قبل، ابتدا نشان دهید که K_1 یک کرنل PDS است. سپس همین نتیجه را به صورت مشابه برای یک α دلخواه نیز به دست آورید.

به طرق مشابه می‌توان نوشت که

$$K_a(x,y) = \langle I_{|x_k|^a \leq a}, I_{|x_k|^b \leq b} \rangle$$

که در حقیقت فقط نماد بالایی را تغییر داده‌ایم به شیوه مشابه بالا می‌توان نوشت که

$$K_a(x, x') = \sum_{K=1}^{N} \min\{|x_k|^a, |x'_k|^a\}$$

یک کرنل است.