KERNEL METHODS IN M.L

Homework 4 - by DAVID SOTO

Exercise 1 : B_n - splines.

* . Let us prove that for a certain $n \in \mathbb{N}^*$ the function $K(x,y) = B_n(x-y)$ is a positive definite kernel over $\mathbb{R}^2$.

* Let $n \in \mathbb{N}$, $\forall x, y \in \mathbb{R}$:

$$K(x,y) = B_n(x-y) = I^{*n}(x-y) = \mathbb{1}_{[0,1]}(|x-y|)^{*n}$$

$$= \mathbb{1}_{[0,1]}(|y-x|)^{*n}$$

$$= B_n(y-x)$$

$$K(x,y) = K(y,x)$$

$\Longrightarrow$ K is symmetric

* To prove that B_n is a positive definite function we will use Bochner theorem. But, in order to use Bochner theorem we need to prove that $B_n : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous and that $B_n : \mathbb{R} \longrightarrow \mathbb{R}$ is the Fourier-Stieltjes Transform of a symmetric and positive finite Borel measure μ.

(1) Let $n \in \mathbb{N}$. We want to prove that B_n is the Fourier-Stieltjes Transform of a symmetric and positive finite Borel measure μ, that is we want $\forall w \in \mathbb{R}$

$$B_n(w) = \int_{\mathbb{R}} e^{-iwx} \, d\mu(x) = \hat{\mu}(w)$$

that is $B_n(\omega) = \hat{\mu}(\omega)$

$$\iff \hat{B}_n(\omega) = \hat{\hat{\mu}}(\omega) = (2\pi)\,\mu(\omega).$$

Let us compute $\hat{B}_n(\omega)$: $\forall \omega \in \mathbb{R}$:

$$\hat{B}_n(\omega) = \widehat{(B_{n-1} * B_1)}(\omega) = \hat{B}_{n-1}(\omega) \cdot \hat{B}_1(\omega)$$

par récurrence on voit très bien que

$$\hat{B}_n(\omega) = \left(\hat{B}_1(\omega)\right)^n = \left(\hat{I}(\omega)\right)^n$$

$$= \left(\int_{\mathbb{R}} I(t)\, e^{-i\omega t}\, dt\right)^n = \left(\int_{-1}^{1} e^{-i\omega t}\, dt\right)^n$$

$$= \left(\left[-\frac{1}{i\omega} e^{-i\omega t}\right]_{-1}^{1}\right)^n$$

$$= \left(\frac{e^{i\omega} - e^{-i\omega}}{i\omega}\right)^n$$

$$= \left(2\,\frac{\sin(\omega)}{\omega}\right)^n$$

$$\hat{B}_n(\omega) = \left(2\,\mathrm{sinc}(\omega)\right)^n$$

Therefore, $\hat{B}_n(\omega) = 2\pi\,\mu(\omega)$

$$\iff (2\,\mathrm{sinc}(\omega))^n = 2\pi\,\mu(\omega)$$

$$\iff \boxed{\mu(\omega) = \frac{1}{2\pi}\left(2\,\mathrm{sinc}(\omega)\right)^n}$$

$\hookrightarrow$ μ is a symmetric & positive finite Borel measure if and only if n is even!

(2) Let us prove that $\forall n \geq 2$, $B_n : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous.

For $n = 2$: $\forall x \in \mathbb{R}$:

$$B_2(x) = I * I\,(x) = \int_{\mathbb{R}} I(u)\, I(x-u)\, du$$

$$= \int_{-1}^{1} \mathbb{1}_{[-1,1]}(x-u)\, du$$

by applying a change of variable, we have:

$$B_2(x) = \int_{x-1}^{x+1} \mathbb{1}_{[-1,1]}(z)\, dz$$

$$B_2(x) = \begin{cases} 0 & \text{if } |x| \geq 2 \\ \int_{x-1}^{1} 1\, dz & \text{if } 0 \leq x < 2 \\ \int_{-1}^{x+1} 1\, dz & \text{if } -2 < x \leq 0 \end{cases}$$

$$B_2(x) = \begin{cases} 2 - |x| & \text{if } |x| \leq 2 \\ 0 & \text{if } |x| > 2 \end{cases}$$

B_2 is clearly continuous on $\mathbb{R}$.

Now let us suppose that for a given $n \geq 2$ B_n is continous. Let us prove that B_{n+1} is continuous: $\forall x \in \mathbb{R}$:

$$B_{n+1}(x) = (B_n * B_1)(x) = B_n * I\,(x)$$

$$= \int_{\mathbb{R}} B_n(u)\, \mathbb{1}_{[-1,1]}(x-u)\, du$$

$$B_{n+1}(x) = \int_{x-1}^{x+1} B_n(u)\, du$$

$$B_{n+1}(x) = \int_{x-1}^{x+1} B_n(u)\,du \,.$$ B_{n+1} is clearly continuous!

Hence, we can conclude that $\forall n \geq 2$, B_n is continuous

Hence, we have shown that $\forall n \in \mathbb{N}^*$, when n is even B_n is continuous and B_n is the Fourier-Stieltjes transform of a symmetric and positive finite Borel measure.

$\Rightarrow$ Therefore, we have by Bochner theorem that, when n is even, B_n is a positive definite function

And since $K : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ is a translation invariant kernel, defined by:
$\forall x, y \in \mathbb{R}$, $K(x,y) = B_n(x-y)$,

and since B_n is a positive definite function when n is even, we can conclude that when n is even $\Rightarrow$ K is a positive definite kernel over $\mathbb{R} \times \mathbb{R}$

* Now, let us describe the RKHS of K when K is a p.d kernel, that is when n is even.

Let $n \in \mathbb{N}^*$ be an even number. Then we know that $K(x,y) = B_n(x-y)$ is a translation invariant positive definite kernel on $\mathbb{R} \times \mathbb{R}$.
Moreover, B_n is integrable since it is n times the convolution of the characteristic function.

$\longrightarrow$

Furthermore, $\hat{B}_n$ is integrable since it is the Fourier-Stieltjes Transform of μ (a symmetric and positive finite Borel measure).

Therefore, by using the theorem of slide 311 (RKHS of translation invariant kernels), we can conclude that the subset H of $L_2(\mathbb{R})$ defined as:

$$H = \left\{ f \in L_2(\mathbb{R}),\ f \text{ integrable and continuous and } \frac{1}{(2\pi)} \int_{\mathbb{R}} \frac{|\hat{f}(\omega)|^2}{\hat{B}_n(\omega)}\, d\omega < +\infty \right\}$$

endowed with the inner product:

$$\langle f, g \rangle_H = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\hat{f}(\omega)\, \overline{\hat{g}(\omega)}}{\hat{B}_n(\omega)}\, d\omega$$

is a RKHS with K as reproducing kernel.

(where $\forall x, y \in \mathbb{R}$, $K(x,y) = B_n(x-y)$ and n is even!)

$\square$

Exercise 2 : Diffusion kernel on a grid.

Setting: We denote $G_1 = (V_1, E_1)$ the line graph with n vertices, where:

$V_1 = \{1, \dots, n\}$ is the vertices set of G_1

$E_1 = \{ (i,j) \in V_1 \times V_1$ such that $|i-j| = 1\}$ is the edge set of G_1.

We denote L_1 the laplacian of the line graph G_1.
Let $0 = \lambda_1 \leq \dots \leq \lambda_n \in \mathbb{R}$ and $e_1, \dots, e_n \in \mathbb{R}^n$ be respectively the eigenvalues and the eigenvectors of the laplacian L_1.

We denote $G_2 = (V_2, E_2)$ the $n \times n$ square grid graph, where:

$V_2 = V_1 \times V_1$ is the vertice set of G_2 and

$E_2 = \{ ((i,j),(i',j')) \in V_2 \times V_2$ such that $|i-i'| + |j-j'| = 1\}$ is the edge set of G_2.

We denote L_2 the laplacian of the square grid graph G_2.

N.B: It is patent that G_2 is the cartesian product of G_1 itself, that is $G_2 = G_1 \times G_1$.

N.B: In this exercise, we will use the Kronecker product $\otimes$.
For any matrices $A \in \mathcal{M}_{m \times n}(\mathbb{R})$ and $B \in \mathcal{M}_{p \times q}(\mathbb{R})$, the Kronecker product of A and B, denoted by $A \otimes B$, is defined as:

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ \vdots & a_{22}B & & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix}$$

1. We have $G_2 = G_1 \times G_1$ (cartesian product)

$$\Rightarrow L_2 = L(G_2) = L(G_1 \times G_1) = [L(G_1) \otimes I_n] + [I_n \otimes L(G_1)]$$

that is the laplacian L_2 of G_2 is given by:

$$L_2 = L_1 \otimes I_n + I_n \otimes L_1$$ (I admit this result)

where L_1 is the laplacian of G_1 and I_n is the $n \times n$ unit matrix.

* By definition of the eigenvalues and eigenvectors of L_1, we have that $\forall i = 1, \dots, n \quad L_1 e_i = \lambda_i e_i$

* Let us show that the eigenvalues of the laplacian L_2 are $\lambda_{ij} = \lambda_i + \lambda_j \quad \forall i = 1, \dots, n, \ \forall j = 1, \dots, n$,

and let us compute the corresponding eigenvectors $e_{ij} \in \mathbb{R}^{n^2}$

→ Let $(i,j) \in \{1, \dots, n\} \times \{1, \dots, n\}$
Let λ_i be an eigenvalue of L_1 and let λ_j be an eigenvalue of L_1.
Let $e_i = (e_{i1}, \dots, e_{in})^T$ and $e_j = (e_{j1}, \dots, e_{jn})^T$ be the eigenvectors corresponding to the eigenvalues λ_i and λ_j of L_1.
Therefore we have:

$$L_1 e_i = \lambda_i e_i \quad \text{and} \quad L_1 e_j = \lambda_j e_j$$

Using the Kronecker product and the fact that $L_2 = L_1 \otimes I_n + I_n \otimes L_1$, we have: →

$$L_2 (e_i \otimes e_j) = \left([L_1 \otimes I_n] + [I_n \otimes L_1] \right) (e_i \otimes e_j)$$

$$= (L_1 \otimes I_n)(e_i \otimes e_j) + (I_n \otimes L_1)(e_i \otimes e_j)$$

using the property that $(A \otimes B)(C \otimes D) = AC \otimes BD$, we have:

$$= (L_1 e_i \otimes I_n e_j) + (I_n e_i \otimes L_1 e_j)$$

$$= (\lambda_i e_i \otimes I_n e_j) + (I_n e_i \otimes \lambda_j e_j)$$

$$= (\lambda_i e_i \otimes e_j) + (e_i \otimes \lambda_j e_j)$$

using the property that $(\lambda A \otimes B) = (A \otimes \lambda B) = \lambda (A \otimes B)$, we have:

$$= \lambda_i (e_i \otimes e_j) + \lambda_j (e_i \otimes e_j)$$

$$L_2 (e_i \otimes e_j) = (\lambda_i + \lambda_j)(e_i \otimes e_j)$$

where $e_i \otimes e_j = (e_{i1} e_{j1} \dots e_{i1} e_{jn} \dots e_{in} e_{j1} \dots e_{in} e_{jn})^T \in \mathbb{R}^{n^2}$

If we denote $\forall (i,j) \in \{1, \dots, n\} \times \{1, \dots, n\}$

$\lambda_{ij} = \lambda_i + \lambda_j \in \mathbb{R}$ and $e_{ij} = e_i \otimes e_j \in \mathbb{R}^{n^2}$

we have $L_2 e_{ij} = \lambda_{ij} e_{ij}$

Thus, we can conclude that the eigenvalues of the Laplacian L_2 are $\lambda_{ij} = \lambda_i + \lambda_j \in \mathbb{R}$ with corresponding eigenvectors $e_{ij} = e_i \otimes e_j \in \mathbb{R}^{n^2}$

$\forall\, i,j = 1, \dots, n$

2. Let $K_1 = e^{-tL_1} \in \mathbb{R}^{n \times n}$ and $K_2 = e^{-tL_2} \in \mathbb{R}^{n^2 \times n^2}$ be diffusion kernels, respectively on graph G_1 and graph G_2.

* Using the eigenvalues $\lambda_1, \ldots, \lambda_n$ of L_1 and the eigenvectors $e_1, \ldots, e_n$ of L_1 and by exploiting orthogonality we can write L_1 as:

$$L_1 = \sum_{i=1}^{n} \lambda_i e_i e_i^T$$

We then have:

$$K_1 = e^{-tL_1} = \sum_{i=1}^{n} e^{-t\lambda_i} e_i e_i^T$$

* Likewise, using the eigenvalues $(\lambda_{ij})_{1 \leq i,j \leq n} \in \mathbb{R}$ of L_2 and the eigenvectors $(e_{ij})_{1 \leq i,j \leq n} \in \mathbb{R}^{n^2}$ of L_2 and by exploiting orthogonality we can write L_2 as:

$$L_2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{ij} e_{ij} e_{ij}^T$$

Recall: $\forall i,j = 1, \ldots, n$, $\lambda_{ij} = \lambda_i + \lambda_j \in \mathbb{R}$
and $e_{ij} = e_i \otimes e_j \in \mathbb{R}^{n^2}$.

We then have:

$$K_2 = e^{-tL_2} = \sum_{i=1}^{n} \sum_{j=1}^{n} e^{-t\lambda_{ij}} e_{ij} e_{ij}^T$$

We recall that $\forall i,j = 1, \dots, n$:

$$e_{ij} = e_i \otimes e_j$$
$$e_{ij} = (e_{i1}e_{j1} \dots e_{i1}e_{jn} \; e_{i2}e_{j1} \dots e_{i2}e_{jn} \dots e_{in}e_{j1} \dots e_{in}e_{jn})^T$$

$\Rightarrow \; \forall i,j = 1, \dots, n, \quad \forall a,b = 1, \dots, n$:

$$e_{ij\,ab} = e_{ia} \, e_{jb} \in \mathbb{R} .$$

* Now, let us prove that $\forall i,j,k,l \in \{1, \dots, n\}$,
$$K_2((i,j),(k,l)) = K_1(i,k) \, K_1(j,l)$$

$\rightarrow$ Let $i,j,k,l \in \{1, \dots, n\}$. We have:

$$\begin{aligned}
K_2((i,j),(k,l)) &= \sum_{a=1}^{n} \sum_{b=1}^{n} e^{-t\lambda_{ab}} \, e_{ijab} \, e_{klab} \\
&= \sum_{a=1}^{n} \sum_{b=1}^{n} e^{-t(\lambda_a + \lambda_b)} \left[e_{ia} e_{jb} \; e_{ka} e_{lb} \right] \\
&= \sum_{a=1}^{n} \sum_{b=1}^{n} e^{-t\lambda_a} e^{-t\lambda_b} \left[e_{ia} e_{ka} \; e_{jb} e_{lb} \right] \\
&= \left[\sum_{a=1}^{n} e^{-t\lambda_a} \, e_{ia} e_{ka} \right] \left[\sum_{b=1}^{n} e^{-t\lambda_b} \, e_{jb} e_{lb} \right]
\end{aligned}$$

$$K_2((i,j),(k,l)) = K_1(i,k) \times K_1(j,l)$$

Hence, we have shown that $\forall i,j,k,l \in \{1, \dots, n\}$

$$K_2((i,j),(k,l)) = K_1(i,k) \, K_1(j,l) .$$

3. K_1 is the exponential of an $n \times n$ matrix, therefore the complexity of computing K_1 is $O(n^3)$.

K_2 is the exponential of an $n^2 \times n^2$ matrix
$\Rightarrow$ the complexity of computing K_2 is $O(n^6)$.