Kernels Methods homework1

February 15, 2016

Exercise 1

Question 1

k_1 and k_2 are p.d.s so their gram matrix (associated to any data examples $\{x_1, ..., x_n\}$) are positive semi-definite: $\forall A \in \mathbb{R}^n$

$$A^T K_1 A \geq 0 \ and \ A^T K_2 A \geq 0$$

Or

$$A^T(\alpha K_1 + \beta K_2)A = \alpha A^T K_1 A + \beta A^T K_2 A \ \geq 0$$

The gram matrix associated is positive semi-definite. Let's ϕ_1 and ϕ_2 the features maps associated to k_1 and k_2, $k = \alpha k_1 + \beta k_2$ and its associated feature map ϕ.

$$k(x,y) = <\phi(x), \phi(y)> = \alpha\phi_1(x)\phi_1(y) + \beta\phi_2(x)\phi_2(y)$$

$$k(y,x) = \alpha\phi_1(y)\phi_1(x) + \beta\phi_2(y)\phi_2(x) = k(x,y)$$

k is then symmetric.
We can conclude that $\alpha K_1 + \beta K_2$ is p.d.

Question 2

(Schur Product wiki)
Again, we will use the fact that the gramm matrix associated to k_1 and k_2 are positive semi-definite.
Let's $k(x,y) = k_1(x,y).k_2(x,y)$, we can easily check that k_1 and k_2 being

symmetric, k is also symmetric.
we want to prove that the gram matrix K is positive semi definite.
$\forall A \in \mathbb{R}^n$:

$$A^T K A = \sum_{i,j} A_i A_j K(x_i, x_j) = \sum_{i,j} A_i A_j K_1(x_i, x_j) * K_2(x_i, x_j)$$

Or K_1 (same goes for K_2) is a positive semi-definite matrix so it's eigenvalues decomposition follows $K_1 = U \Sigma U^T = U(\Sigma^{1/2})^T \Sigma^{1/2} U^T = M_1^T M_1$ where $M_1 = \Sigma^{1/2} U^T$

$$K_1 = M_1^T M_1 \rightarrow [K_1]_{i,j} = \sum_k (M_1)_{ik} (M_1)_{jk}$$

$$K_2 = M_2^T M_2 \rightarrow [K_2]_{i,j} = \sum_l (M_2)_{il} (M_2)_{jl}$$

If we plug it in the previous equation:

$$A^T K A = \sum_{i,j} A_i A_j \sum_{k,l} (M_1)_{ik} (M_1)_{jk} \sum_l (M_2)_{il} (M_2)_{jl}$$

$$A^T K A = \sum_{k,l} \sum_{i,j} A_i A_j (M_1)_{ik} (M_1)_{jk} (M_2)_{il} (M_2)_{jl}$$

$$A^T K A = \sum_{k,l} \sum_i A_i (M_1)_{ik} (M_2)_{il} \sum_j A_j (M_1)_{jk} (M_2)_{jl}$$

$$A^T K A = \sum_{k,l} (\sum_i A_i (M_1)_{ik} (M_2)_{il})^2 \geq 0$$

This prove that the matrix K is positive semi definite.

Question 3

Using notations in the exercise, for a given k_n being a p.d kernel implies that it's associated gram matrix is positive semi definite: $\forall A \in \mathbb{R}^n$:

$$A^T K_n A = \sum_{i,j} A_i A_j K_n(x_i, x_j) \geq 0$$

$$lim_{n \rightarrow \infty} A^T K_n A = A^T K A \geq 0$$

So k is a p.d. kernel

Question 4

The Taylor decomposition of exponential gives us

$$e^{K_1}(x,y) = lim_{N\to\infty} \sum_{n=0}^{N} \frac{1}{n!} k_1(x,y)^n$$

- $k_1(x,y)^n$ are p.d. kernels as they are products of positive definite kernels (c.f. Question 2)
- $\frac{1}{n!} k_1(x,y)^n$, we have a positive definite kernel multiplied by a positive constant. The resulting kernel will also be p.d. (it is easy to see that the associated gram matrix is positive semi definite).
- $\sum_{n=0}^{N} \frac{1}{n!} k_1(x,y)^n$ are p.d. kernels as they are sums of positive definite kernels (c.f. Question 1)
- e^{K_1} finally is p.d. because it is limit a p.d kernel sequence.

Exercise 2

Question 1

I don't think it is a p.d kernel because $k(x,x) = 1/0 = \inf$ so $||\phi(x)|| = \inf$

Question 2

I don't think it is a p.d kernel because $k(x,x) = 1/0 = \inf$ so $||\phi(x)|| = \inf$

Exercise 3

Question 1 :

$$K(a,b) = a.b \to f(x) = \sum_i \lambda_i x_i.x \; with \; x \in \mathbb{R}$$

$$f(x) = \lambda x \; with \; \lambda = \sum_i \lambda_i x_i \; and \; ||f|| = |\lambda|$$

The same way apply for g: $g(y) = \beta y \ with \ \beta = \sum_i \beta_i y_i$
The criterion can then be written as :

$$C_n^K(X,Y) = \max_{\lambda,\beta \in [-1,1]} Cov_n(\lambda X, \beta Y)$$

$$C_n^K(X,Y) = \max_{\lambda,\beta \in [-1,1]} \mathbb{E}_n(\lambda X \beta Y) - \mathbb{E}_n(\lambda X).\mathbb{E}_n(\beta Y)$$

By linearity we have

$$C_n^K(X,Y) = \max_{\lambda,\beta \in [-1,1]} \lambda\beta(\mathbb{E}_n(XY) - \mathbb{E}_n(X).\mathbb{E}_n(Y))$$

$$C_n^K(X,Y) = \max_{\lambda,\beta \in [-1,1]} \lambda\beta Cov_n(X,Y)$$

giving the constraints on $\lambda \ and \ \beta$, the criterion above is maximized when $\lambda.\beta = sign(Cov_n(X,Y))$ which means :

$$f(x) = x \ or \ f(x) = -x \rightarrow f \ is \ Id \ or f \ is \ -Id$$

$$g(y) = y \ or \ g(y) = -y \rightarrow g \ is \ Id \ or g \ is \ -Id$$

Finally we will have:

$$C_n^K(X,Y) = |Cov_n(X,Y)|$$