Assignment 1: Kernel Methods in ML

Pen & Paper exercises

Arttu Häkkinen
596077

Question 1

$$K_1(x,y) = (\langle x, y\rangle + c)^m, \quad c \geq 0$$

linear kernel ($\phi(x)=x$), m is a positive integer, $x, y \in \mathbb{R}^d$

Proving that the polynomial kernel $K_1(x,y)$ is a valid kernel:

$$K_1(x,y) = (\langle x,y\rangle + c)^m = (x^T y + c)^m = \left(\sum_{i=1}^{d} x_i y_i + c\right)^m$$

Applying the binomial theorem

$$K_1(x,y) = \sum_{k=1}^{m} \binom{m}{k} c^{m-k} \langle x,y\rangle^k$$

Here we can see that we have a conic sum of kernels, which is also a kernel (lecture 1 slide 40):

"For kernels $(k_j)_{j=1}^{k}$ and $(\alpha_j)_{j=1}^{k} > 0$, $\sum_{j=1}^{k} \alpha_j k_j$ is also a kernel."

This is true in our case because

→ $\binom{m}{k} c^{m-k} > 0$, for all $m, k \in \mathbb{N}_+$ and $c \geq 0$

→ $\langle x,y\rangle$ is a linear kernel and $\langle x,y\rangle^k$ is a product of linear kernels and any product of kernels is also a kernel as is proved on lecture 1 slide 45

→ So because $\binom{m}{k} c^{m-k}$ is a positive scalar and $\langle x,y\rangle^k$ is a kernel, we can verify that $\sum_{k=1}^{m} \binom{m}{k} c^{m-k} \langle x,y\rangle^k$ is a conic sum of kernels, so therefore $K_1(x,y)$ is a valid kernel. ∎

Question 2

Showing that by deriving appropriate expressions for α_i and b, the decision function

$$h(x) = \begin{cases} +1 & \text{if } \|\phi(x) - c_-\|^2 > \|\phi(x) - c_+\|^2 \\ -1 & \text{otherwise} \end{cases}$$

can be written in the following form

$$h(x) = \operatorname{sgn}\left(\sum_{i=1}^{n} \alpha_i k(x, x_i) + b\right), \text{ such that } k(x, x_i) = \langle \phi(x), \phi(x_i) \rangle$$

$$\operatorname{sgn} x = \begin{cases} -1 & , \text{ if } x < 0 \\ 0 & , \text{ if } x = 0 \\ 1 & , \text{ if } x > 0 \end{cases}$$

$$c_+ = \frac{1}{n_+} \sum_{i \in I^+} \phi(x_i) \quad \text{and} \quad c_- = \frac{1}{n_-} \sum_{i \in I^-} \phi(x_i)$$

Let's modify the $h(x)$ in to the form of sgn-function.

$$h(x) = \begin{cases} +1 & \text{if } \|\phi(x) - c_-\|^2 - \|\phi(x) - c_+\|^2 > 0 \\ -1 & \text{otherwise} \end{cases}$$

$$\Rightarrow h(x) = \operatorname{sgn}(\|\phi(x) - c_-\|^2 - \|\phi(x) - c_+\|^2)$$

$$= \operatorname{sgn}(\langle \phi(x) - c_-, \phi(x) - c_- \rangle - \langle \phi(x) - c_+, \phi(x) - c_+ \rangle)$$

$$= \operatorname{sgn}(\cancel{\langle \phi(x), \phi(x) \rangle} - 2\langle \phi(x), c_- \rangle + \langle c_-, c_- \rangle - \cancel{\langle \phi(x), \phi(x) \rangle} + 2\langle \phi(x), c_+ \rangle - \langle c_+, c_+ \rangle)$$

$$= \operatorname{sgn}(2(\langle \phi(x), c_+ \rangle - \langle \phi(x), c_- \rangle) + \langle c_-, c_- \rangle - \langle c_+, c_+ \rangle)$$

$$= \operatorname{sgn}(2\langle \phi(x), c_+ - c_- \rangle + \langle c_-, c_- \rangle - \langle c_+, c_+ \rangle)$$

$$= \operatorname{sgn}(2c_+\phi(x) - 2c_-\phi(x) + c_-^2 - c_+^2) \quad \| \cdot \frac{1}{2}$$

$$= \operatorname{sgn}(c_+\phi(x) - c_-\phi(x) + \frac{1}{2}c_-^2 - \frac{1}{2}c_+^2)$$

$$= \operatorname{sgn}\left(\frac{1}{2}c_-^2 - \frac{1}{2}c_+^2 + c_+\phi(x) - c_-\phi(x)\right)$$

$$= \operatorname{sgn}\left(\frac{1}{2m_-^2} \sum_{i,j \in I^-} \phi(x_i)\phi(x_j) - \frac{1}{2m_+^2} \sum_{i,j \in I^+} \phi(x_i)\phi(x_j) \right.$$

$$\left. + \frac{1}{m_+} \sum_{i \in I^+} \phi(x_i)\phi(x) - \frac{1}{m_-} \sum_{i \in I^-} \phi(x_i)\phi(x) \right)$$

$$= \operatorname{sgn}\left(\frac{1}{m_+} \sum_{i \in I^+} k(x, x_i) + \left(\frac{-1}{m_-}\right) \sum_{i \in I^-} k(x, x_i) \right.$$

$$\left. + \frac{1}{2m_-^2} \sum_{i,j \in I^-} k(x_i, x_j) - \frac{1}{2m_+^2} \sum_{i,j \in I^+} k(x_i, x_j) \right)$$

$$= \operatorname{sgn}\left(\sum_{i=1}^{m} \alpha_i \, k(x, x_i) + b \right),$$

where

$$\alpha_i = \begin{cases} \frac{1}{m_+} & \text{if } y_i = +1 \\ \frac{-1}{m_-} & \text{if } y_i = -1 \end{cases}$$

and

$$b = \frac{1}{2m_-^2} \sum_{i,j \in I^-} k(x_i, x_j) - \frac{1}{2m_+^2} \sum_{i,j \in I^+} k(x_i, x_j)$$

$\blacksquare$

Question 3

Checking if $K_2(x, y) = \cos(x+y)$ is a valid kernel when $x, y \in \mathbb{R}$.

Proof.

Finding $\phi(x)$ and $\mathcal{H}$ such that $k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$ is not trivial in this case and we would need to use imaginary numbers.

Trying to prove the validity by assessing the positive-definiteness of K_2. As stated in theorem of Moore-Aronszajn, all positive-definite functions are kernels and vice-versa.

Def. of PD function

"A symmetric function $k: \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ is positive-definite if $\forall n \geq 1$, $\forall (a_1, \dots, a_n) \in \mathbb{R}^n$, $\forall (x_1, \dots, x_n) \in \mathcal{X}^n$

$$\sum_{i=1}^{n} \sum_{i=j}^{n} a_i a_j \, k(x_i, x_j) \geq 0. \quad "$$

→ Let $n = 1$, so we need to check whether the following holds.

$$a^2 K_2(x, x) \geq 0, \text{ for all } x, y \in \mathbb{R}.$$

→ Let $x = \frac{\pi}{2} \Rightarrow x + x = \pi$

$$a^2 \cdot \cos(\pi) = -a^2 \leq 0$$

→ So $K_2(x, y) = \cos(x+y)$ is not a valid kernel. ∎

Question 4

Claim. $k_3(x, y) = \dfrac{1}{1 - xy}$ is a valid kernel for $x, y \in (-1, 1)$.

Proof.

$$k_3(x, y) = \frac{1}{1 - xy}$$

$$= \sum_{i=0}^{\infty} x^i y^i$$

$$= \langle \phi(x), \phi(y) \rangle,$$

where $\phi(x) = (x^0, x^1, x^2, x^3, \dots)$ ■