AIMS African Masters in Machine Intelligence 2024

Kernel Methods - Quizz #2

July 26, 2024

Exercise 1: Course questions

Answers can be found in class or with other students

Q3(c) $\nabla f(x) = (A + A^T)x + b$

Q7(a) `return (x[None, :]  - x[:, None])**2`

Q7(b) `return np.sum((x[None, :, :]  - x[:, None, :])**2, axis=-1)`

Exercise 2: Kernels

1. In course

2. Idem

3. *Bonus* Show that $\langle X, Y \rangle = \mathbb{E}[XY]$ is an inner product on the space of real-valued random variables.

   - symmetric
   - bilinear, by linearity of expectation
   - $\langle X, X \rangle = \mathbb{E}[X^2] \geq 0$. $\mathbb{E}[X^2] = 0$ implies $X = 0$ almost everywhere.

4. State whether the following functions are p.d. kernels or not. Prove your answer.

   (a) $\mathcal{X} = \mathbb{R}^2$, $k(x, x') = x_1 x'_1 + sign(x_2) sign(x'_2)$
   Choose $\mathcal{H} = \mathbb{R}^2, \Phi(x) = [x_1, sign(x_2)]^T$, then $k(x, x') = \langle \Phi(x), \Phi(x') \rangle_{\mathcal{H}}$

   (b) $\mathcal{X}$ is an inner product space, $k(x, x') = \|x\|^2$

   k is not symmetric ! Choose $x = 0$ and x' nonzero for example

   (c) $\mathcal{X}$ is the space of real-valued random variables, $k(X, X') = Cov(X, X')$

   Seen in class.
   Using the scalar product from Q3, and choosing $\Phi(X) = X - \mathbb{E}[X]$, we have $Cov(X, X') = \langle \Phi(X), \Phi(X') \rangle$

   (d) Let Ω be a set, $\mathcal{X} = \text{Subsets}(\Omega)$, $k(A, B) = \mathbb{P}[A \cap B]$.
   $k(A, B) = \mathbb{P}[A \cap B] = \mathbb{E}[1_{A \cap B}] = \mathbb{E}[1_A 1_B] = \langle 1_A, 1_B \rangle$
   Choose $\mathcal{H}$ as the space of real-valued random variables, and $\Phi(A) = 1_A$

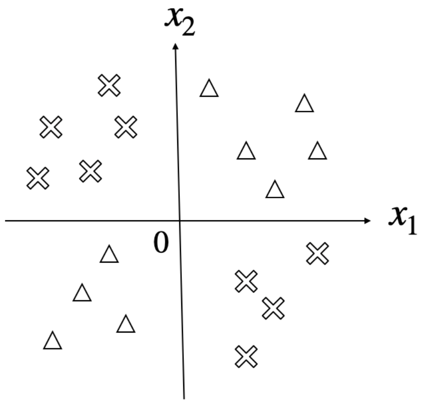
   (e) Same as (c), with $k(A, B) = \mathbb{P}[A \cup B]$
   It is not p.d.
   From **@Attou**, take $A = \Omega$, $B = \emptyset$, then the Gram matrix is equal to $K = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, which is not p.s.d. since $det(K) = -1 < 0$

Exercice 3: Kernel Ridge Regression

All questions are course material.

Exercise 4: Application, XOR

In this exercise we want to classify data points $x = (x_1, x_2) \in \mathbb{R}^2$, with two labels (binary classification), as shown in the picture below. Let $X_\triangle$ and $X_\times$ be the sets of points with respective labels $\triangle$ and $\times$.



1. Are the data $X_\triangle$ and $X_\times$ linearly separable?

   NO

2. (a) Find a space $\mathcal{X}$ and a mapping $\Phi : \mathbb{R}^2 \to \mathcal{X}$ such that the projected data $\Phi(X_\triangle)$ and $\Phi(X_\times)$ are linearly separable.

   There are multiple possibilities for Φ, we choose $\mathcal{X} = \mathbb{R}$ and $\Phi(x) = x_1 x_2$ here, many correctly chose $\Phi(x) = sign(x_1 x_2)$.

   Ask **@Najlaa** for explanation and drawing.

   (b) What is the corresponding kernel function k?

   $k(x, x') = \langle \Phi(x), \Phi(x') \rangle = x_1 x_2 x'_1 x'_2$

   (c) *Bonus* What is the RKHS of k?

   The RKHS is the space $\mathcal{H} = \{f_\beta : (t_1, t_2) \to \beta t_1 t_2,\ \beta \in \mathbb{R}\}$,

   with norm $\|f_\beta\|_{\mathcal{H}} = \beta^2$

   Proof: Let $\mathcal{H} = \{K_x,\ x \in \mathbb{R}^2\}$.

   Prove that $\mathcal{H}$ is a vector space, as it is stable under linear combinations: for $a, b \in \mathbb{R}, x, x' \in \mathbb{R}^2$ there exists $z \in \mathbb{R}^2$ such that $aK_x + bK_{x'} = K_z$.

   As a consequence the closure of the span of $\mathcal{H}$ is $\mathcal{H}$ itself.

   The norm is then $\langle K_x, K_x \rangle = k(x, x) = (x_1 x_2)^2$