Theory of Machine Learning

Exercise sheet 12 — Session 12

Exercise I (Non-expansiveness of the Gaussian kernel) ✎. Consider the Gaussian kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ such that for all pair of points $\mathbf{x}, \mathbf{x}'$ in $\mathbb{R}^d$,

$$K(\mathbf{x}, \mathbf{x}') = e^{-\frac{1}{2\nu^2}\|\mathbf{x}-\mathbf{x}'\|_2^2},$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^d$ and the bandwidth $\nu > 0$. Call $\mathcal{H}$ the RKHS of K and consider its RKHS mapping $\phi : \mathbb{R}^d \to \mathcal{H}$ such that $K(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle_{\mathcal{H}}$ for all $\mathbf{x}, \mathbf{x}'$ in $\mathbb{R}^d$. Show that

$$\|\phi(\mathbf{x}) - \phi(\mathbf{x}')\|_{\mathcal{H}} \leq \frac{1}{\nu} \|\mathbf{x} - \mathbf{x}'\|_2 \,.$$

The mapping is called non-expansive whenever $\nu \geq 1$.

1. Show that
$$\|\phi(\mathbf{x}) - \phi(\mathbf{x}')\|_{\mathcal{H}}^2 = K(\mathbf{x}, \mathbf{x}) + K(\mathbf{x}', \mathbf{x}') - 2K(\mathbf{x}, \mathbf{x}') \,.$$
2. Show that
$$\|\phi(\mathbf{x}) - \phi(\mathbf{x}')\|_{\mathcal{H}}^2 = 2\left(1 - e^{-\frac{1}{2\nu^2}\|\mathbf{x}-\mathbf{x}'\|_2^2}\right) \,.$$
3. Using that $1 + x \leq \exp(x)$ (for all $x \in \mathbb{R}$), show that
$$\|\phi(\mathbf{x}) - \phi(\mathbf{x}')\|_{\mathcal{H}} \leq \frac{1}{\nu} \|\mathbf{x} - \mathbf{x}'\|_2 \,.$$

Exercise II (A useful feature map ϕ) ✎. Given the input space $\mathcal{X} := \mathbb{R}^2$, we define the polynomial kernel as $K(\mathbf{x}, \mathbf{x}') := \langle \mathbf{x} \,,\, \mathbf{x}' \rangle_{\mathbb{R}^2}^2$.

1. Show that K is a p.d. kernel without any computations.
2. Develop the expression of $K(\mathbf{x}, \mathbf{x}') := \langle \mathbf{x} \,,\, \mathbf{x}' \rangle_{\mathbb{R}^2}^2$.
3. Find a feature map $\phi : \mathbb{R}^2 \to \mathbb{R}^3$ such that $K(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}) \,,\, \phi(\mathbf{x}') \rangle_{\mathbb{R}^3}$.
4. Plotting the effect of the feature map ϕ:
   (a) Generate a dataset with points labeled based on whether they lie inside or outside a unit circle in $\mathbb{R}^2$.
   (b) Visualize this dataset using `matplotlib.pyplot.scatter()`.
   (c) Create a new dataset in $\mathbb{R}^3$ by applying the feature map ϕ on each point of the previous dataset.
   (d) Visualize the transformed data in 3D, showcasing their linear separability.

Exercise III (Construction of the RKHS) ✎. Given the input space $\mathcal{X} := \mathbb{R}^d$, we define the polynomial kernel of degree 2 as $K(\mathbf{x}, \mathbf{x}') := \langle \mathbf{x} \,,\, \mathbf{x}' \rangle_{\mathbb{R}^d}^2$. The objective is to construct the associated RKHS.

1. **First step: look for an inner-product.** show that
$$K(\mathbf{x}, \mathbf{y}) = \left\langle \mathbf{x}\mathbf{x}^\top \,,\, \mathbf{y}\mathbf{y}^\top \right\rangle_F ,$$
with $\langle \cdot \,,\, \cdot \rangle_F$ the Froebenius inner-product.
2. **Second step: propose a candidate RKHS.** Show that the RKHS candidate $\mathcal{H}$ contains all the functions of the form:
$$\forall \mathbf{x} \in \mathbb{R}^d, \forall n \in \mathbb{N}^\star, \forall \{a_i\}_{i=1}^n \subset \mathbb{R}, \forall \{\mathbf{x}^{(i)}\}_{i=1}^n \subset \mathbb{R}^d, \qquad f(\mathbf{x}) := \left\langle \sum_{i=1}^n a_i \mathbf{x}^{(i)} {\mathbf{x}^{(i)}}^\top \,,\, \mathbf{x}\mathbf{x}^\top \right\rangle_F \,.$$
3. **Third step: check that the candidate is the RKHS.** Show that the candidate $\mathcal{H}$ which is the set of quadratic functions is the RKHS.

Exercise IV (Massart's lemma) ✎. Let us assume that $\mathcal{G}$ is *finite*, that is, $\mathcal{G} = \{g_1, \dots, g_m\}$. Let us assume further that $\frac{1}{n}\sum_{i=1}^n g_j(X_i)^2 \le R^2$ for all $j \in [d]$. Show that the Rademacher complexity of the function class $\mathcal{G}$ satisfies

$$R_n(\mathcal{G}) \le \sqrt{\frac{2\log m}{n}} R\,.$$

For simplicity's sake, we consider the X_is fixed.

1. Given $\lambda > 0$, show that

$$\exp\left(\lambda \mathbb{E}_\varepsilon\left[\sup_{g\in\mathcal{G}} \sum_{i=1}^n \varepsilon_i g(X_i)\right]\right) \le \sum_{g\in\mathcal{G}} \mathbb{E}_\varepsilon\left[\exp\left(\lambda \sum_{i=1}^n \varepsilon_i g(X_i)\right)\right]\,.$$

   (*Hint: Jensen's inequality and property of* sup.)

2. Show that

$$\sum_{g\in\mathcal{G}} \mathbb{E}_\varepsilon\left[\exp\left(\lambda \sum_{i=1}^n \varepsilon_i g(X_i)\right)\right] = \sum_{g\in\mathcal{G}} \prod_{i=1}^n \frac{1}{2}\left(\exp\left(\lambda g(X_i)\right) + \exp\left(-\lambda g(X_i)\right)\right)\,.$$

   (*Hint*: independence of the ε_is and direct computation of the remaining expectation.)

3. Using $\frac{e^x + e^{-x}}{2} \le e^{x^2/2}$ (for all $x \in \mathbb{R}$), show that

$$\sum_{g\in\mathcal{G}} \prod_{i=1}^n \frac{1}{2}\left(\exp\left(\lambda g(X_i)\right) + \exp\left(-\lambda g(X_i)\right)\right) \le \sum_{g\in\mathcal{G}} \exp\left(\frac{\lambda^2}{2} \sum_{i=1}^n g(X_i)^2\right)\,.$$

4. Show that

$$\sum_{g\in\mathcal{G}} \exp\left(\frac{\lambda^2}{2} \sum_{i=1}^n g(X_i)^2\right) \le m\exp\left(\frac{n\lambda^2 R^2}{2}\right)\,.$$

   (*Hint*: refer to the assumptions in the exercise statement on g and $\mathcal{G}$.)

5. By putting everything together, show that

$$R_n(\mathcal{G}) \le \frac{1}{n\lambda}\log m + \frac{\lambda R^2}{2}\,.$$

6. Show that $\lambda^\star = \frac{1}{R}\sqrt{\frac{2\log m}{n}}$ minimizes the previous bound.

7. Show the Massart's lemma bound.