Theory of Machine Learning

Exercise sheet 11 — Session 11

Exercise I (On some properties of kernels). Given an input space $\mathcal{X}$, prove the following fundamental properties of p.d. kernels:

1. If K_1 and K_2 are p.d. kernels on $\mathcal{X}$, then $K_1 + K_2$, $K_1 \times K_2$ and cK_1 (with $c \geq 0$) are p.d. kernels.
2. If $(K_i)_{i \geq 1}$ is a sequence of p.d. kernels that converges pointwisely to a function K, *i.e.*:
$$\forall (x, x') \in \mathcal{X}^2, \qquad K(x, x') = \lim_{n \to +\infty} K_n(x, x')\,,$$
then K is a p.d. kernel.

Exercise II (On some kernels). Show that the following functions are p.d. kernels:

1. With $\mathcal{X} := \mathbb{R}^d$, $K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2\sigma^2}\left\|\mathbf{x} - \mathbf{x}'\right\|_2^2\right)$ and $\sigma > 0$.
2. With $\mathcal{X} := (-1, 1)$, $K(x, x') = \frac{1}{1 - xx'}$.
3. With $\mathcal{X} := \mathbb{R}, K(x, x') = 2^{xx'}$.
4. With $\mathcal{X} := \mathbb{R}$, $K(x, x') = \cos(x - x')$.
5. With $\mathcal{X} := \mathbb{R}_+$, $K(x, x') = \min(x, x')$.
6. With $\mathcal{X} := \mathbb{R}_+^*$, $K(x, x') = \frac{\min(x, x')}{\max(x, x')}$.

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

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

1. Given $\lambda > 0$, show that
$$\exp\left(\lambda \mathbb{E}_\varepsilon\left[\sup_{g \in \mathcal{G}} \sum_{i=1}^n \varepsilon_i g(X_i)\right]\right) \leq \sum_{g \in \mathcal{G}} \mathbb{E}_\varepsilon\left[\exp\left(\lambda \sum_{i=1}^n \varepsilon_i g(X_i)\right)\right]\,.$$
(*Hint: Jensen's inequality and property of* sup.)
2. Show that
$$\sum_{g \in \mathcal{G}} \mathbb{E}_\varepsilon\left[\exp\left(\lambda \sum_{i=1}^n \varepsilon_i g(X_i)\right)\right] = \sum_{g \in \mathcal{G}} \prod_{i=1}^n \frac{1}{2}\left(\exp(\lambda g(X_i)) + \exp(-\lambda g(X_i))\right)\,.$$
(*Hint*: independence of the ε_is and direct computation of the remaining expectation.)
3. Using $\frac{e^x + e^{-x}}{2} \leq e^{x^2/2}$ (for all $x \in \mathbb{R}$), show that
$$\sum_{g \in \mathcal{G}} \prod_{i=1}^n \frac{1}{2}\left(\exp(\lambda g(X_i)) + \exp(-\lambda g(X_i))\right) \leq \sum_{g \in \mathcal{G}} \exp\left(\frac{\lambda^2}{2}\sum_{i=1}^n g(X_i)^2\right)\,.$$

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.