## CMPUT 654 Fa 23: Theoretical Foundations of Machine Learning

Fall 2023

## Lecture 21: Kernel Methods (November 23)

Lecturer: Csaba Szepesvári

Scribes: Kushagra Chandak

**Note**: *ET<sub>E</sub>X* template courtesy of UC Berkeley EECS dept. (*link* to directory) **Disclaimer**: These notes have <u>not</u> been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.

In the RKHS  $\mathcal{H}$ , we want to minimize the following objective function:

$$\widetilde{\widetilde{Q}} = \frac{1}{n} \sum_{i=1}^{n} \ell(f(x_i), y_i) + \frac{\lambda}{2} \|f\|_{\mathcal{H}}^2.$$

**Theorem 21.1.** For every k symmetric, positive definite kernels,  $\exists (\mathcal{W}, \langle \cdot, \cdot \rangle_{\mathcal{W}})$  Hilbert space and  $\psi : \mathcal{X} \to \mathcal{W}$  such that  $k(u, v) = \langle \psi(u), \psi(v) \rangle_{\mathcal{W}}$ .

Further, the function  $f^{\psi} : \mathcal{W} \to \mathcal{H}_k$  defined by  $w \mapsto (x \mapsto \langle w, \psi(x) \rangle)$  is onto and preserves the norm. That is,  $f^{\psi}(\mathcal{W}) = \mathcal{H}_k$  and  $||w||_{\mathcal{W}}^2 = ||f^{\psi}(w)||_{\mathcal{H}_k}^2$  for all  $w \in \mathcal{W}$ .

We can choose  $\psi(x) = k(x, \cdot)$  and  $\mathcal{W} = \mathcal{H}_k$ .

**Definition 21.2** (Universal kernal).  $k : \mathcal{X}^2 \to \mathbb{R}$ , which is symmetric and positive definite, is a universal kernel if for every  $f \in C(\mathcal{X})$  and for every  $\varepsilon > 0$ , there exists  $g \in \mathcal{H}_k$  such that  $||f - g||_{\infty} \le \varepsilon$ .

To show universality of a kernel, we have the following theorem.

**Theorem 21.3** (Stone-Weierstrass). Let  $\mathcal{X} \subseteq \mathbb{R}^d$  is compact. Then  $\mathbb{R}_{\mathcal{X}}[x] \subseteq C(X)$  is dense w.r.t  $\|\cdot\|_{\infty}$ .

**Corollary 21.4.** Let k be symmetric positive definite kernel. Assume  $\exists \psi_i : \mathcal{X} \to \mathbb{R}$  and  $c_i > 0$  such that  $k(x,y) = \sum_{i=1}^{\infty} c_i \psi_i(x) \psi_i(y)$ . Also assume that

$$\{x \mapsto x_1^{p_1} \dots x_d^{p_d} : p_1, \dots, p_d \ge 0\} \subseteq \psi_i : i \ge 1.$$

Then k is universal.