# Statistical Foundations of Learning - Assignment 6

CIT4230004 (Summer Semester 2024)

## Exercise 6.1: Feature maps of universal kernels are injective

Let $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a universal kernel. Denote by $H$ its Reproducing Kernel Hilbert Space (RKHS), and by $\varphi : \mathbb{R}^d \to H$ its feature map. Recall the reproducing property: For every $f \in H$ and any $x \in \mathbb{R}^d$, we have

$$f(x) = \langle \varphi(x), f \rangle.$$

**Prove that $\varphi$ is injective.**

To prove that $\varphi$ is injective, we need to show that if $\varphi(x_1) = \varphi(x_2)$, then $x_1 = x_2$.

Assume $\varphi(x_1) = \varphi(x_2)$. For any $f \in H$,

$$f(x_1) = \langle \varphi(x_1), f \rangle = \langle \varphi(x_2), f \rangle = f(x_2).$$

Since $k$ is a universal kernel, it implies that the feature map $\varphi$ is characteristic, meaning that it can distinguish between any two distinct points in $\mathbb{R}^d$. Therefore, if $f(x_1) = f(x_2)$ for all $f \in H$, it must be that $x_1 = x_2$.

Thus, $\varphi$ is injective.

## Exercise 6.2: Solving Machine Learning (?)

Consider a neural network $h : [0, 1] \to \mathbb{R}$ with just one hidden neuron:

$$h(x) = \sigma(ax + b)$$

for some activation $\sigma$ and $a, b \in \mathbb{R}$.

**1. Show that there exists an activation function $\sigma$ such that for any polynomial $f : [0, 1] \to \mathbb{R}$ and all $\epsilon > 0$, there exists some $a, b$ satisfying**

$$\sup_{x \in [0,1]} |f(x) - h(x)| \leq \epsilon.$$

Hint: The set of all polynomials with coefficients in $\mathbb{Q}$ is countable (and dense in the set of all polynomials with coefficients in $\mathbb{R}$).

Consider the activation function $\sigma(x) = e^x$. For any polynomial $f(x)$, we can approximate it by finding suitable $a$ and $b$ such that $\sigma(ax+b)$ approximates $f(x)$ within $\epsilon$.

Given the density of polynomials with rational coefficients in the space of all polynomials, for any polynomial $f(x)$ and $\epsilon > 0$, we can find a polynomial $p(x)$ with rational coefficients such that

$$\sup_{x\in[0,1]} |f(x) - p(x)| \leq \epsilon.$$

Since $\sigma(ax+b)$ can approximate $p(x)$ by suitable choice of $a$ and $b$, we have

$$\sup_{x\in[0,1]} |f(x) - h(x)| \leq \epsilon.$$

**2. Junior data scientist Alex is thrilled about this observation and wants to use it for all his machine learning models. What do you tell him?**

While the observation that a neural network with a single hidden neuron can approximate any polynomial is interesting, it has practical limitations. Real-world machine learning models often deal with more complex functions and higher-dimensional data. A single hidden neuron may not have the capacity to capture such complexity.

Additionally, neural networks with more layers (deep neural networks) are known to provide better approximations for complex functions and patterns in data. Therefore, while this observation is theoretically significant, practical machine learning models require more sophisticated architectures to achieve good performance.