The Mathematics of Machine Learning
Homework Set 6

Due 24 April 2024 before 13:00
via Canvas

You are allowed to work on this homework in pairs. One person per pair submits the answers via Canvas. Make sure to put both names on the submission.

1 Introduction

In lecture 1 we saw the benefits of transforming a single feature $x \in \mathbb{R}$ into $\tilde{x} = (1, x, x^2, x^3, \ldots, x^d) \in \mathbb{R}^{d+1}$. Then a linear function of $\tilde{x}$ is a d-th degree polynomial of x. In the next lecture, we will use the idea of transforming features to greatly extend the power of support vector machines. In general let $\tilde{x} = h(x)$ be an arbitrary transformation of the original features. We even want to allow h to construct an *infinite* number of features. In this homework, we will develop the mathematical tools to make this possible. What we need is to construct a suitable *Hilbert space* $\mathcal{H}$ and let h be a map from $x \in \mathcal{X}$ to an element of $\mathcal{H}$. We will be interested in transformations h such that the inner product after applying h,

$$\langle h(x), h(x') \rangle,$$

comes out as a nice function of x and x'. In fact, given a function $k(x, x')$, which we will call a *kernel function*, we want to know whether there exist an $\mathcal{H}$ and h such that

$$\langle h(x), h(x') \rangle = k(x, x') \qquad \text{for all } x, x' \in \mathcal{X}.$$

The goal of this homework is to show that this is always possible as soon as k satisfies a few weak conditions.

2 Understanding SVMs, Part II

A kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called positive semi-definite if, for any finite set of inputs $x_1, \ldots, x_m \in \mathcal{X}$, the matrix $K \in \mathbb{R}^{m \times m}$ with entries $K_{i,j} = k(x_i, x_j)$ is positive semi-definite. That is, K should be symmetric and, for any $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$,

$$\sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j k(x_i, x_j) \geq 0.$$

In particular, symmetry of K implies that k should be symmetric.

Given any positive semi-definite kernel function k, the goal of this homework is to construct a Hilbert space $\mathcal{H}$ and a mapping $h : \mathcal{X} \to \mathcal{H}$ from the original space of features to this Hilbert space, where k corresponds to the inner product in $\mathcal{H}$:

$$k(x, x') = \langle h(x), h(x') \rangle.$$

2.1 Definition of the Reproducing Kernel Hilbert Space

The elements of $\mathcal{H}$ will consist of functions $f : \mathcal{X} \to \mathbb{R}$. In particular, our mapping h will produce the following functions:

$$h(x) := k(\cdot, x) \qquad \text{for any } x \in \mathcal{X}.$$

These functions generate a linear space $\mathcal{L} \subset \mathcal{H}$ that contains all finite linear combinations, i.e. all functions of the form:

$$f(\cdot) = \sum_{i=1}^{m} \alpha_i k(\cdot, x_i)$$

for any $x_1, \ldots, x_m \in \mathcal{X}$, any $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ and any positive integer m. For any two functions $f, g \in \mathcal{L}$ we know there must exist (possibly non-unique) representations

$$f(\cdot) = \sum_{i=1}^{m} \alpha_i k(\cdot, x_i), \qquad\qquad g(\cdot) = \sum_{j=1}^{k} \beta_j k(\cdot, x'_j)$$

for some $\alpha_i \in \mathbb{R}$,$\beta_i \in \mathbb{R}$, $x_1, \ldots, x_m, x'_1, \ldots, x'_k \in \mathcal{X}$. The inner product between $f, g \in \mathcal{L}$ is then defined as

$$\langle f, g \rangle := \sum_{i=1}^{m} \sum_{j=1}^{k} \alpha_i \beta_j k(x_i, x'_j).$$

We will show below that this inner product gives the same value for all possible representations of f and g, so it is well-defined. We will also show that it satisfies all the properties that are required of an inner product. In fact, the inner product also satisfies one more property: for all functions $f \in \mathcal{L}$,

$$\langle k(\cdot, x), f \rangle = f(x) \quad \text{and, in particular,} \quad \langle k(\cdot, x), k(\cdot, x') \rangle = k(x, x').$$

This is called the reproducing property of k and k is called a *reproducing kernel*. Having established all these properties, there is one property of Hilbert spaces that $\mathcal{L}$ may not satisfy: it may not be complete. The Hilbert space $\mathcal{H}$ is therefore defined as the completion of $\mathcal{L}$ in the metric $d(f, g) := \|f - g\| := \sqrt{\langle f - g, f - g \rangle}$. As a result of the reproducing property, this $\mathcal{H}$ is called a *reproducing kernel Hilbert space* (RKHS).

3 Questions: Verifying All Required Properties

We proceed to verify that the definitions above satisfy all requirements.

1. [1 pt] Show that the inner product does not depend on the choice of representation for f and g by showing that

$$\sum_{i=1}^{m}\sum_{j=1}^{k}\alpha_i\beta_j k(x_i, x'_j) = \sum_{j=1}^{k}\beta_j f(x'_j) = \sum_{i=1}^{m}\alpha_i g(x_i).$$

   (The first identity shows that the inner product does not depend on the representation of f; the second identity that it does not depend on the representation of g.)

Then verify the reproducing property:

2. [1 pt] Prove the reproducing property.

Finally, show that the inner product satisfies all the requirements to make $\mathcal{L}$ an inner product space:

3. [1 pt] Symmetry: $\langle f, g\rangle = \langle g, f\rangle$ for any $f, g \in \mathcal{L}$.
4. [1 pt] Linearity: $\langle \alpha_1 f_1 + \alpha_2 f_2, g\rangle = \alpha_1\langle f_1, g\rangle + \alpha_2\langle f_2, g\rangle$ for any $f_1, f_2, g \in \mathcal{L}$, $\alpha_1, \alpha_2 \in \mathbb{R}$.
5. [1 pt] Positive semi-definiteness: $\langle f, f\rangle \geq 0$.
6. [1 pt] If $\langle f, f\rangle = 0$, then $f = 0$ (i.e. $f(x) = 0$ for all $x \in \mathcal{X}$).
   Hint: Combine the reproducing property with the Cauchy-Schwarz inequality $\langle f, g\rangle^2 \leq \langle f, f\rangle\langle g, g\rangle$, which is proved below in the appendix.

This homework is based on the review paper by Hofmann et al. [2008], which contains a great overview of SVMs and other kernel methods.

References

T. Hofmann, B. Schölkopf, and A. J. Smola. Kernel methods in machine learning. *The Annals of Statistics*, 36(3):1171–1220, 2008. doi: 10.1214/009053607000000677. URL `https://doi.org/10.1214/009053607000000677`.

A A Proof of Cauchy-Schwarz Using Only the Proved Properties

The hint in Question 6 suggests to use the Cauchy-Schwarz inequality. For this to be allowed, we have to verify that Cauchy-Schwarz follows from the properties already established in the preceding questions, which is what will be shown here.

It will be convenient to use the shorter notation for norms: $\|f\| := \sqrt{\langle f, f\rangle}$, which are well-defined and non-negative by positive semi-definiteness. Then Cauchy-Schwarz requires us to show that

$$\langle f, g\rangle^2 \leq \|f\|^2 \|g\|^2.$$

To prove this, consider first the case that at least one of the norms $\|f\|$ and $\|g\|$ is strictly positive. By symmetry, we may assume that this is the case for the norm of f, i.e. $\|f\| > 0$. It then follows from linearity that

$$\|f\|^2 \|g\|^2 - \langle f, g\rangle^2 = \frac{\| \|f\|^2 g - \langle f, g\rangle f\|^2}{\|f\|^2} \geq 0,$$

where the inequality holds by non-negativity of the norms.

Alternatively, consider the case that $\|f\| = \|g\| = 0$. Then

$$\begin{aligned} 0 \leq \|f - g\|^2 &= \|f\|^2 - 2\langle f, g\rangle + \|g\|^2 = -2\langle f, g\rangle \quad \Rightarrow \quad \langle f, g\rangle \leq 0 \\ 0 \leq \|f + g\|^2 &= \|f\|^2 + 2\langle f, g\rangle + \|g\|^2 = 2\langle f, g\rangle \quad \Rightarrow \quad -\langle f, g\rangle \leq 0. \end{aligned}$$

Combining these two cases, we get $\langle f, g\rangle = 0$, so

$$\langle f, g\rangle^2 = 0 = \|f\|^2 \|g\|^2,$$

as required.