Kernelized Locality Sensitive Hashing

Daniel Speyer and Geelon So
{dls2192,geelon}@columbia.edu

December 13, 2017

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

1 Introduction

2 Kernels

One of the most natural ways to encode structure into an abstract set $\mathcal{X}$ is to provide a mapping from that set to the reals, $\mathcal{X} \longrightarrow \mathbb{R}$. Or more generally, a mapping to some structure-rich mathematical object $\mathcal{O}$. Indeed, at a high level, we can consider different mathematical objects as the 'canonical' idealization of certain types of structure; then, certain functions from $\mathcal{X}$ to these objects allow us to 'pull back' or instantiate that structure within our specific $\mathcal{X}$.

Of course, the structure we care about with respect to LSH is *similarity*, one idealization of which is the *(real) inner product*. Let us take a moment to reexamine the inner product on finite real vector spaces as a notion of similarity, before generalizing that notion using *kernels* to infinite dimensional vector spaces and arbitrary sets.

2.1 Finite-Dimensional Inner Product Spaces

Definition 1. Let V be a (possibly infinite) vector space over $\mathbb{R}$. An *inner product* on V is a bilinear map $K : V \times V \longrightarrow \mathbb{R}$ that is *symmetric* and *positive-definite*. We call the pair (V, K) an *inner product space*.

Symmetry and positive-definiteness are two of the most obvious requirements for any self-respecting function that calls itself a measure of similarity:

1. the similarity between two points x and y better not depend on the order we

present them to the similarity measure, so we require $K(x, y) = K(y, x)$,

2. a nonzero object should be similar to itself, so if $x \neq 0$, then $K(x, x) > 0$.

These two intuitions of a 'similarity measure' then precisely define the conditions of an inner product on V.[1] Allow us to be (perhaps overly) rigorous in the following discussion, for it is better to bore in the finite case than to confuse in the general case.

For concreteness, let $V = \mathbb{R}^n$ with a fixed basis. With any $x, y \in V$, denote by $\langle x, y \rangle$ the *formal dot product* of x and y as an algebraic construction.[2] Then, as K is a bilinear form, there exists a matrix $\mathbf{K}$ where

$$K(x, y) = \langle x, \mathbf{K}y \rangle. \tag{1}$$

The symmetry condition on the inner product K forces the matrix $\mathbf{K}$ to be symmetric. The spectral theorem on symmetric matrices tell us that V has an eigenbasis with respect to $\mathbf{K}$. Consider an eigenvalue $\mathbf{K}v = \lambda v$. As we require K to be positive-definite, we know that

$$K(v, v) = \langle v, \mathbf{K}v \rangle = \lambda \langle v, v \rangle > 0.$$

Since the dot product of a vector with itself in any basis is nonnegative, this implies that the eigenvalue λ is positive; so in its eigenbasis, $\mathbf{K}$ is diagonal with positive terms on its diagonal—it follows that $\mathbf{K}^{1/2}$ exists, and this allows us to write:

$$K(x, y) = \langle \mathbf{K}^{1/2}x, \mathbf{K}^{1/2}y \rangle.$$

We deduce that there exists a linear map $\Phi : V \longrightarrow \mathbb{R}^n$ (spoiler: we call Φ a *feature map*) where the inner product on V directly corresponds to the usual dot product on $\mathbb{R}^n$. That is,

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle. \tag{2}$$

In other (ridiculously abstract) words, Equation 2 precisely says:

Proposition 2. *Let V be an n-dimensional real vector space. If $K : V \times V \longrightarrow \mathbb{R}$ is an inner product, then there exists a map $\Phi : V \longrightarrow \mathbb{R}^n$ such that the following map commutes:*

$$
\begin{array}{ccc}
V \times V & \xrightarrow{\;\;\Phi \times \Phi\;\;} & \mathbb{R}^n \times \mathbb{R}^n \\
& \searrow_{K} & \downarrow^{\langle \cdot, \cdot \rangle} \\
& & \mathbb{R}
\end{array}
$$

where $\langle \cdot, \cdot \rangle$ is the standard dot product on $\mathbb{R}^n$.

This is the main payoff to all this formality: we may view the pair $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ as the canonical n-dimensional real inner product space. So, no matter which inner product space (V, K) we want to study, we're in fact guaranteed the existence of an isomorphism Φ to the familiar inner product space $\mathbb{R}^n$, and we may freely interchange the two objects in our minds.

Furthermore, we are justified in dropping the distinction between the formal dot product and inner product; let us denote both by $\langle \cdot, \cdot \rangle$.

---

[1]For now, we can just view bilinearity as a necessary condition to ensure that K respects the real numbers as an algebraic object. So, in some sense, bilinearity is not intrisic to K; rather, it is a condition *induced* by the field $\mathbb{R}$. Later on, when we generalize, we won't have a notion of (bi)linearity on arbitrary sets $\mathcal{X}$; however, there is a natural 'completion' of $\mathcal{X}$ as a vector space $\mathcal{H}$. Here, $\mathbb{R}$ induces the linear structure.

[2]One point of confusion may be that $\langle \cdot, \cdot \rangle$ often denotes an inner product—usually unproblematic as the dot product is in fact the inner product *induced by the choice of basis* on V. However, an abstract vector space has no canonical basis; thus, it has no canonical inner product. In particular, here, the inner product corresponding to the dot product is in general unrelated to the inner product K. We will reconcile these two views at Equation 2, which shows that every inner product corresponds to the dot product in an appropriate basis. Conversely, as every dot product is an inner product, if V is an abstract finite-dimensional real vector space, *specifying an inner product on V is equivalent to specifying a basis on V.*

2.2 Generalization: Kernels and Hilbert Spaces

First, we will need to extend the 'similarity measure' from vector spaces V to abstract sets $\mathcal{X}$. In the former, notice that given a basis $\{v_1, \ldots, v_n\}$, the inner product K is fully determined by the collection of values, $K(v_i, v_j)$, by bilinearity of K (this is equivalent to saying that we can represent the bilinear form K by a matrix of $n \times n$ values $\mathbf{K}$). In some sense, this says that there are only 'n ways' or 'n directions' in which objects of V may be (dis)similar.

This suggests that we can generalize Definition 1 because we didn't need all of V to start with! Suppose someone secretly had an n-dimensional inner product space (V, K), but just gave us n linearly independent vectors, say $[n] := \{v_1, \ldots, v_n\}$, along with the corresponding restriction of the similarity measure $K : [n] \times [n] \longrightarrow \mathbb{R}$.

Our view of $[n]$ would just be a collection of n abstract objects, and K just an $n \times n$ matrix over $\mathbb{R}$ (this is often called the *Gram matrix*). Still, we would have produced the same feature map $\Phi_{[n]} : [n] \hookrightarrow \mathbb{R}^n$, albeit restricted to $[n] \subset V$. But in the previous analysis, since V and $\mathbb{R}^n$ are isomorphic via Φ, this implies that $\Phi_{[n]}$ actually *recovers* V by its identification with $\mathbb{R}^n$; it is as though we've discovered that $[n]$ actually lived inside the space V.

Very naturally, the generalization of K from $[n]$ to $\mathcal{X}$ takes the view that K is a similarity 'matrix' on $\mathcal{X}$ (though we call it a *kernel*):

Definition 3. Let $\mathcal{X}$ be a nonempty set. A *kernel*[3] is a map $K : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$. We say that K is *positive-definite* if for any finite subsets A of $\mathcal{X}$, the corresponding Gram matrix of A is symmetric and positive-definite.

In the finite case, we produced an embedding $\Phi_{[n]} : [n] \longrightarrow \mathbb{R}^n$ of $[n]$ satisfying Equation 2 (K corresponds to inner product). As we had only n objects, it is clear that we needed an inner product space with at most n dimensions, hence $\mathbb{R}^n$.

However, in the general case, where $\mathcal{X}$ may be infinite, we may need infinite dimensions. Consider the real vector space $\mathbb{R}^{\mathcal{X}}$ (vector addition and scalar multiplication are defined pointwise), and the map $\Phi : \mathcal{X} \longrightarrow \mathbb{R}^{\mathcal{X}}$ defined by:

$$\Phi(x) := K(\cdot, x).$$

For short, let $k_x := \Phi(x)$. Then, Φ maps $\mathcal{X}$ into a subset of $\mathbb{R}^{\mathcal{X}}$; let $V = \text{span}(\Phi(\mathcal{X}))$ be a linear subspace of $\mathbb{R}^{\mathcal{X}}$, so that elements f of V are of the form:

$$f = \alpha_1 k_{x_1} + \cdots + \alpha_n k_{x_n},$$

with $\alpha_i \in \mathbb{R}$ and n ranging over $\mathbb{N}$. We claim that because K is a positive-definite kernel,

---

[3]Unfortunately, the terminology *kernel* can be terribly confusing. Within functional analysis, kernels are often interchangably called *kernel maps*, *kernel functions*, and *positive-definite kernels*. Yet, each of these also take on other meanings depending on author and context. 'Kernels' and 'kernel maps' are almost always synonymous. Sometimes, 'kernel functions' denote kernels of the form $K(x, y) = k(x - y)$, in situations where subtraction is defined. And in many instances, when context is clear, the author cares only for positive-definite kernels, so omitting the specifier *positive-definite*.

Furthermore, *kernels* used in this sense are unrelated to the algebraic kernel of a linear map. The terminology comes from the theory of integral operators. In particular, infinite-dimensional function spaces often have inner products of the form:

$$\langle f, g \rangle = \int_{X \times X} K(x, y) f(x) g(y) \, \mathrm{d}x \mathrm{d}y,$$

where the K here performs the analogous role as $\langle x, \mathbf{K}y \rangle$ in Equation 1. Actually, it performs precisely the role we desire in the general case. These maps K were historically called kernels.

---

there exists a well-defined bilinear map on V, corresponding to our desired inner product. Of course, we want the inner product to satisfy:

$$\langle k_x, k_y \rangle = K(x, y).$$

But in fact, once we specify this, linearity determines the inner product on general elements $f = \sum_{i=1}^{n_1} \alpha_i k_{x_i}$ and $g = \sum_{j=1}^{n_2} \beta_j k_{y_j}$:

$$\langle f, g \rangle = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \alpha_i \beta_j K(x_i, y_j).$$

This should feel quite familiar, recalling how the Gram matrix of n linearly independent vectors fully determined the inner product in the finite case. However, we must be a little more careful here, as we are not guaranteed that the collection of k_x's are linearly independent. In particular, suppose that f, as follows, is identically zero:

$$0 \equiv f = \sum_{i=1}^{n} \alpha_i k_{x_i}.$$

That is, $f(x) = 0$ for all $x \in \mathcal{X}$. Then, $\langle f, g \rangle$ better equal 0 for all $g \in V$. Notice that it is sufficient to show that $\langle f, k_x \rangle = \langle k_x, f \rangle = 0$ for each $x \in \mathcal{X}$. And indeed, by assumption

$$\langle k_x, f \rangle = \sum_{i=1}^{n} k_{x_i}(x) = f(x) = 0.$$

This proves that V admits an inner product that is compatible with K in the sense of Equation 2. We will in fact go beyond producing an inner product space V; we can complete the space with respect to the norm induced by the inner product, producing a *Hilbert space.*

That is, let $\mathcal{H}$ be the completion of V by taking equivalence classes of Cauchy sequences on V. We claim that $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ is a collection of functions.[4] But let us relegate

the proof to references, say [1, Thm 3.16], for it is not particularly enlightening. What we do attain, however, is the analog to Proposition 2:

Proposition 4 (Moore)**.** *Let $\mathcal{X}$ be a set, and $K : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ a positive-definite kernel. Then, there exists a map $\Phi : \mathcal{X} \longrightarrow \mathcal{H}$ into a Hilbert space such that the following map commutes:*

where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the inner product on $\mathcal{H}$.

The amazing aspect here was that $\mathcal{X}$ was just a set; the only thing we needed was a positive-definite kernel K to obtain $\mathcal{H}$. In fact:

Fact 5. *The Hilbert space $\mathcal{H}$ induced by K is unique. We call $\mathcal{H}$ the* reproducing kernel Hilbert space (RKHS) *on $\mathcal{X}$ with respect to K. (See [1, Prop 3.3]).*

---

[4]Note that in general, it is not the case that the completion of a function space consists of functions. Take any Dirac delta distribution from within $L^2(\mathbb{R})$, for example.

References

[1] Paulsen, Vern I., and Mrinal Raghupathi. *An introduction to the theory of reproducing kernel Hilbert spaces.* Vol. 152. Cambridge University Press, 2016.