Inner Product Spaces (Background)

Stephen Vardeman
Analytics Iowa LLC
ISU Statistics and IMSE

Vectors, scalar products, and sums

Most of applied mathematics in general and statistical machine learning in particular is built on the notions of "linear combinations" of various objects and "inner products" of these (that in turn lead to coherent notions of their "sizes" of and "distances" between them). Here we briefly review what is necessary for a theory of such objects and operations to make sense.

First, a vector (or linear) space $\mathbf{V}$ consists of objects $\mathbf{v}, \mathbf{w}, \ldots$ such that if $\mathbf{v} \in \mathbf{V}$ and $a \in \Re$, then the object $a\mathbf{v}$ makes sense and belongs to $\mathbf{V}$, and for $\mathbf{v}$ and $\mathbf{w}$ in $\mathbf{V}$ the object $\mathbf{v} + \mathbf{w}$ also makes sense and belongs to $\mathbf{V}$. The archetypal vector spaces are the Euclidean spaces $\Re^p$ where elements are "ordinary" p-dimensional vectors. But other kinds of vector spaces are useful in statistical machine learning as well, including function spaces. Take for example the set of functions on $[0, 1]$ that have finite integrals of their squares. (This space is sometimes known as $L_2([0, 1])$.)

Vector spaces of functions

More or less obviously, if $g : [0,1] \longrightarrow \Re$ with $\int_0^1 (g(x))^2 \, dx < \infty$ and $a \in \Re$, then $ag(x)$ makes sense, maps $[0,1]$ to $\Re$ and has

$$\int_0^1 (ag(x))^2 \, dx = a^2 \int_0^1 (g(x))^2 \, dx < \infty$$

Further, if $g : [0,1] \longrightarrow \Re$ with $\int_0^1 (g(x))^2 \, dx < \infty$ and $h : [0,1] \longrightarrow \Re$ with $\int_0^1 (h(x))^2 \, dx < \infty$, then the function $g(x) + h(x)$ makes sense, maps $[0,1]$ to $\Re$ and has finite integral of its square.

Inner products

An inner product on a vector space $\mathbf{V}$ is a symmetric (bi-)linear positive definite function mapping $\mathbf{V} \times \mathbf{V} \to \Re$. That is, $\langle \mathbf{v}, \mathbf{w} \rangle$ is an inner product on $\mathbf{V}$ if it satisfies

1. $\langle \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle \ \forall \ \mathbf{v}, \mathbf{w} \in \mathbf{V}$ (symmetry),
2. $\langle a\mathbf{v}, \mathbf{w} \rangle = a \langle \mathbf{v}, \mathbf{w} \rangle \ \forall \ \mathbf{v}, \mathbf{w} \in \mathbf{V}$ and $a \in \Re$,
   $\langle \mathbf{v} + \mathbf{u}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle \ \forall \ \mathbf{u}, \mathbf{v}$, and $\mathbf{w} \in \mathbf{V}$ (bi-linearity), and
3. $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0 \ \forall \ \mathbf{v} \in \mathbf{V}$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$ (positive definiteness).

Examples

Of course Euclidean p-space is a vector space with inner product defined as the "dot-product" of p-dimensional vectors $\mathbf{v}$ and $\mathbf{w}$ is

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}'\mathbf{w} = \sum_{j=1}^{p} v_j w_j$$

It is possible to argue that in the case of the $L_2([0,1])$ function space, the integral of the product of two elements provides a valid inner product, that is

$$\langle g, h \rangle \equiv \int_0^1 g(x)\, h(x)\, dx$$

satisfies properties 1. through 3.

Norms and distances

An inner product on a vector space **V** leads immediately to notions of size and distance. The norm (i.e. the "size" or "length") of an element of **V** can be defined as

$$\|\mathbf{v}\| \equiv \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

Then the distance between two elements of **V** can be taken to be the size of the difference between them. That is, the distance between **v** and **w** belonging to **V** (say $d(\mathbf{v}, \mathbf{w})$) derived from the inner product is

$$d(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|$$

This satisfies all the properties necessary to qualify as a "metric" or "distance function," including the important triangle inequality.

Examples

In Euclidean p-space, the norm is the geometrical length of a p-vector (the root of the sum of the p squared entries of the vector) and the associated distance is ordinary Euclidean distance.

In the case of the $L_2([0,1])$ function space, the norm/size of an element g is

$$\|g\| = \sqrt{\int_0^1 (g(x))^2 \, dx}$$

and the distance between elements g and h is

$$d(g,h) = \sqrt{\int_0^1 (g(x) - h(x))^2 \, dx}$$

Other concepts like those in Euclidean spaces

Many other useful notions commonly understood in Euclidean spaces generalize directly to inner product spaces. $\mathbf{v}$ and $\mathbf{w} \in \mathbf{V}$ are **perpendicular** or **orthogonal** when $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. **Subspaces** of $\mathbf{V}$ can be generated as all linear combinations of a set of elements of $\mathbf{V}$ and are commonly referred to as the "**span**" of the set of elements. A **basis** for a subspace of $\mathbf{V}$ is a set of linearly independent vectors (no linear combination of them is the $\mathbf{0}$ vector) that span the subspace. "**Orthonormal**" **bases** (whose elements are perpendicular and each of norm 1) for $\mathbf{V}$ (or for subspaces of $\mathbf{V}$) are particularly attractive, as they provide very simple representations for "**projections**" of $\mathbf{v} \in \mathbf{V}$ onto the span of any set of them, as a linear combination of basis vectors where coefficients are the inner products with the corresponding basis vectors.

Projections and "low-dimensional" approximations

In the context of machine learning, projections of a vector **v** are very usefully thought of as "**low-dimensional**" **approximations** to **v** (in terms of a "few" basis vectors). (The dimension of a subspace of **V** is, just as in ordinary Euclidean spaces, the number of vectors in a basis for it.) Geometry of Euclidean cases (where subspaces are geometrical hyperplanes containing the origin and geometrical hyperplanes are subspaces potentially shifted from the origin by addition of a vector not in the subspace) is helpful in interpreting statistical machine learning constructs in more abstract inner product spaces.