Hello and welcome back from spring break!

Before the midterm

We had been discussing orthogonality, how to compute orthogonal projections, and how to find orthonormal bases (Gram-Schmidt).

This time

We discuss how to solve the linear interpolation or "least squares" problem, and then discuss how to think about orthogonality in a more abstract setting.

Least Squares

As you know, the equation $A\mathbf{x} = \mathbf{b}$ need not have a solution.

Now that we have a notion of distance, we can instead ask for the "closest thing to a solution".

More precisely, we ask for $\mathbf{x}$ so that $||A\mathbf{x} - \mathbf{b}||$ is minimized.

This is called the "least squares" problem, because taking length involves squares.

Least Squares

$A\mathbf{x} = \mathbf{b}$ can be solved if and only if $\mathbf{b}$ is in the span of A's columns.

The closest point to $\mathbf{b}$ in the column space of A is the orthogonal projection

$$\hat{\mathbf{b}} := \mathrm{Projection}_{Col\ A}(\mathbf{b})$$

So solving the least squares problem — minimizing $||A\mathbf{x} - \mathbf{b}||$ — is the same as solving the usual linear algebraic problem $A\mathbf{x} = \hat{\mathbf{b}}$.

This sounds like we must first compute $\hat{\mathbf{b}}$.

Least Squares

Actually it's not necessary to compute $\hat{\mathbf{b}}$.

Since $\hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{b}$ to the column space of A, certainly $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to A's columns: $A^T(\mathbf{b} - \hat{\mathbf{b}}) = 0$

Thus if $A\mathbf{x} = \hat{\mathbf{b}}$, then $A^T(A\mathbf{x} - \mathbf{b}) = 0$.

Least Squares

Conversely, suppose $\mathbf{x}$ satisfies $A^T(A\mathbf{x} - \mathbf{b}) = 0$.

Then $\mathbf{b} - A\mathbf{x}$ is orthogonal to the column space of A,

and $A\mathbf{x}$ is in the column space of A,

so $A\mathbf{x}$ is the orthogonal projection of $\mathbf{b}$ to the column space of A.

I.e., $A\mathbf{x} = \hat{\mathbf{b}}$.

Least Squares

Thus solving the "least squares problem"

I.e., finding $\mathbf{x}$ which minimizes $||A\mathbf{x} - \mathbf{b}||$

Is the same as solving the linear equation

$$A^T A\mathbf{x} = A^T\mathbf{b}$$

Example

Find the vector $\mathbf{x}$ minimizing

$$\left\| \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x} - \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\|$$

I.e., solve

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, so

$$\mathbf{x} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

Try it yourself

Find the vector $\mathbf{x}$ minimizing

$$\left\| \begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 0 & 3 \end{bmatrix} \mathbf{x} - \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\|$$

Orthogonal projection revisited

We know that one way to find the orthogonal projection to V is to:

Determine an orthogonal basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$ of V

Use these vectors to find the orthogonal projection:

$$\mathrm{Projection}_V(\mathbf{b}) = \sum \left(\frac{\mathbf{b} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} \right) \mathbf{v}_i$$

Orthogonal projection revisited

Here is another way. Take any basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$ of V; let A be the matrix with these columns.

Finding the orthogonal projection of $\mathbf{b}$ to V is the same as finding a least squares solution $\mathbf{x}$ for $A\mathbf{x} = \mathbf{b}$ and then computing $A\mathbf{x}$.

In other words,

$$\hat{\mathbf{b}} = A(A^T A)^{-1} A^T \mathbf{b}$$

In the case that the basis was orthonormal $A^T A$ is the identity.

Fitting lines

Suppose given data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ for which you want the closest fit line.

What does closest fit line mean?

Let's say, the sum of the squares of the discrepancies, $\sum (y_i - f(x_i))^2$, is minimized.

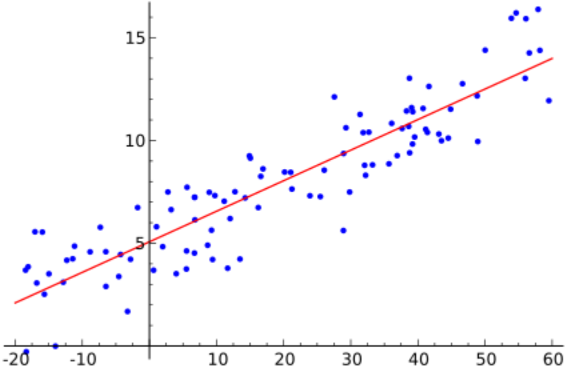
Fitting lines

Since $f(x_i) = ax_i + b$, this means minimizing

$$\left\| \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \right\|$$

You know how to do this.

Fitting lines



Properties of the dot product

Distributivity (aka "bilinearity")

$$(a\mathbf{v}+b\mathbf{v}')\cdot(c\mathbf{w}+d\mathbf{w}') = ac(\mathbf{v}\cdot\mathbf{w})+bc(\mathbf{v}'\cdot\mathbf{w})+ad(\mathbf{v}\cdot\mathbf{w}')+bd(\mathbf{v}'\cdot\mathbf{w}')$$

Commutativity

$$\mathbf{v}\cdot\mathbf{w} = \mathbf{w}\cdot\mathbf{v}$$

Positivity

$$\mathbf{v}\cdot\mathbf{v} \geq 0, \qquad \text{with equality only when } \mathbf{v} = 0$$

Inner products

Definition

An inner product on a vector space V is a map

$$\langle \cdot , \cdot \rangle : V \times V \to \mathbb{R}$$

which is distributive, commutative, and positive.

Inner product properties

Distributivity (aka "bilinearity")

$$\langle a\mathbf{v} + b\mathbf{v}', c\mathbf{w} + d\mathbf{w}' \rangle = ac\langle \mathbf{v}, \mathbf{w} \rangle + bc\langle \mathbf{v}', \mathbf{w} \rangle + ad\langle \mathbf{v}, \mathbf{w}' \rangle + bd\langle \mathbf{v}', \mathbf{w}' \rangle$$

Commutativity

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$$

Positivity

$$\langle \mathbf{v}, \mathbf{v} \rangle \geq 0, \qquad \text{with equality only when } \mathbf{v} = 0$$

Example: Evaluation inner products

Recall that a polynomial of degree n is determined by its evaluation at any $n+1$ distinct real numbers.

Choosing such, say $r_0, \ldots, r_n$, determines an isomorphism

$$\begin{aligned} Ev : \mathbb{P}_n &\to \mathbb{R}^{n+1} \\ f &\mapsto (f(r_0), f(r_1), \ldots, f(r_n)) \end{aligned}$$

This determines a function on pairs of polynomials:

$$\langle f, g \rangle = Ev(f) \cdot Ev(g)$$

It is an inner product on $\mathbb{P}_n$ — the properties hold because they held for the dot product.

Example: The integral inner product

Consider the vector space of continuous functions on the interval $[a, b]$, and the pairing

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx$$

This is an inner product: it is clearly distributive and commutative and it is positive because, if f is continuous and nonzero, then $\langle f, f \rangle = \int_a^b f(x)^2 dx$ is the integral of a nonnegative continous function which is somewhere nonzero.

Example: The integral inner product

The integral inner product is a continuum limit of evaluation inner products:

$$\int_a^b f(x)g(x)dx = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} f(a + \frac{t}{n}(b-a))g(a + \frac{t}{n}(b-a))$$

Length, distance, orthogonality

An inner product on a vector space determines notions of:

Length: the length of a vector $\mathbf{v}$ is $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.

Distance: the distance from $\mathbf{a}$ to $\mathbf{b}$ is $||\mathbf{a} - \mathbf{b}||$.

Orthogonality: the vectors $\mathbf{v}, \mathbf{w}$ are orthogonal if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

Example: ℓ^2

Consider the vector space *Seq* of all infinite sequences of real numbers, $(a_1, a_2, a_3, \ldots)$. It is tempting to define

$$\langle (a_1, a_2, a_3, \ldots), (b_1, b_2, b_3, \ldots) \rangle := \sum_{i=1}^{\infty} a_i b_i$$

However, the sum need not converge.

In fact, already the putative length

$$||a|| = \langle (a_1, a_2, a_3, \ldots), (a_1, a_2, a_3, \ldots) \rangle^{1/2} = \left(\sum a_i^2 \right)^{1/2}$$

may fail to be defined.

Example: ℓ^2

Consider the set of sequences $(a_1, a_2, \ldots)$ for which $\sum a_i^2$ is finite.
This set is called ℓ^2.

E.g., $(1, \frac{1}{2}, \frac{1}{3}, \ldots) \in \ell^2$ because $\sum \frac{1}{n^2} = \frac{\pi^2}{6}$.

However, $(1, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{3}}, \ldots) \notin \ell^2$ because $\sum \frac{1}{n}$ diverges.

Example: ℓ^2

In fact, ℓ^2 is a subspace. Let's check: given $(a_1, a_2, \ldots)$ and $(b_1, b_2, \ldots)$ in ℓ^2, we have for each finite n:

$$\begin{aligned}
\left(\sum_{i=1}^{n}(a_i + b_i)^2\right)^{1/2} &= ||(a_1 + b_1, \ldots, a_n + b_n)|| \\
&\leq ||(a_1, \ldots, a_n)|| + ||(b_1, \ldots, b_n)|| \\
&= \left(\sum_{i=1}^{n} a_i^2\right)^{1/2} + \left(\sum_{i=1}^{n} b_i^2\right)^{1/2}
\end{aligned}$$

This final quantity is, by assumption, bounded as $n \to \infty$, hence so is the first, so $(a_1 + b_1, a_2 + b_2, \ldots) \in \ell^2$.

Example: ℓ^2

The formula

$$\langle (a_1, a_2, a_3, \ldots), (b_1, b_2, b_3, \ldots) \rangle := \sum_{i=1}^{\infty} a_i b_i$$

is finite on ℓ^2:

$$\sum a_i b_i = \frac{1}{2} \left(\sum (a_i + b_i)^2 - \sum a_i^2 - \sum b_i^2 \right)$$

The distributivity, commutativity, and positivity properties hold for the same reasons as for the dot product.

We have defined an inner product on ℓ^2.

Example: ℓ^2

Consider the set *FinSeq* of all sequences whose entries are eventually all zero.

FinSeq $\subset \ell^2$ since $\sum a_i^2$ is actually a finite sum in this case.

FinSeq has a basis $\mathbf{e}_i$, where $\mathbf{e}_i$ is the sequence with all zeroes, except in the i'th position, where it has a 1.

Example: ℓ^2

For any $\mathbf{a} = (a_1, a_2, \ldots) \in \ell^2$, we have $\langle \mathbf{a}, \mathbf{e}_i \rangle = a_i$. Thus the orthogonal complement of *FinSeq* is $\{\mathbf{0}\}$.

Unlike in the finite dimensional case:

$(1, \frac{1}{2}, \frac{1}{3}, \ldots)$ cannot be written as a sum of something in *FinSeq* and something in the orthogonal complement to *FinSeq*.

$(FinSeq^{\perp})^{\perp} = \ell^2 \supsetneq FinSeq$.

The orthogonal projection to *FinSeq* is not well defined.

Orthogonal projections, orthogonal bases, Gram-Schmidt

In a finite dimensional vector space V with an inner product,

given a subspace W, we write $W^\perp$ for the set of vectors whose inner product with anything in W is zero.

Any vector v can be uniquely decomposed as $v = w + w^\perp$ with $w \in W$ and $w^\perp \in W^\perp$.

This holds by the same proof as before. Challenge question: where did that proof use finite dimensionality?

Orthogonal projections, orthogonal bases, Gram-Schmidt

This gives a notion of orthogonal projection.

Given an orthogonal basis $\mathbf{w}_1, \ldots, \mathbf{w}_k$ for W — i.e., $\langle \mathbf{w}_i, \mathbf{w}_j \rangle = 0$ when $i \neq j$ — one has

$$Projection_W(\mathbf{v}) = \sum \frac{\langle \mathbf{v}, \mathbf{w}_i \rangle}{\langle \mathbf{w}_i, \mathbf{w}_i \rangle} \mathbf{w}_i$$

Orthonormal bases can be found by the Gram-Schmidt process.

Writing inner products in a basis

Let V be a finite dimensional vector space with an inner product, and let $\mathcal{B} := \mathbf{b}_1, \ldots, \mathbf{b}_n$ be a basis of V.

Recall that for a vector $\mathbf{v} \in V$, we can uniquely expand

$$\mathbf{v} = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 + \cdots + v_n\mathbf{b}_n$$

We write $[\mathbf{v}]_{\mathcal{B}}$ for the vector $(v_1, \ldots, v_n)$.

Since $\mathbf{v} \to [\mathbf{v}]_{\mathcal{B}}$ gives an isomorphism $V \to \mathbb{R}^n$, we should be able to express the inner product just in terms of $[\mathbf{v}]_{\mathcal{B}}$.

Writing inner products in a basis

Consider the matrix $[\langle \mathbf{b}_i, \mathbf{b}_j \rangle]$, whose entry in the i'th row and j'th column is $\langle \mathbf{b}_i, \mathbf{b}_j \rangle$.

$$
\begin{aligned}
\langle \mathbf{v}, \mathbf{w} \rangle &= \langle \sum_i v_i \mathbf{b}_i, \sum_j w_j \mathbf{b}_j \rangle \\
&= \sum_{i,j} \langle v_i \mathbf{b}_i, w_j, \mathbf{b}_j \rangle \\
&= \sum_{i,j} v_i \langle \mathbf{b}_i, \mathbf{b}_j \rangle w_j \\
&= [\mathbf{v}]_{\mathcal{B}}^T [\langle \mathbf{b}_i, \mathbf{b}_j \rangle] [\mathbf{w}]_{\mathcal{B}}
\end{aligned}
$$

Writing inner products in a basis

If the basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$ was orthonormal, then the matrix $[\langle \mathbf{b}_i, \mathbf{b}_j \rangle]$ is the identity matrix.

That is, in an orthonormal basis $\mathcal{B}$,

$$\langle \mathbf{v}, \mathbf{w} \rangle = [\mathbf{v}]_{\mathcal{B}}^T [\mathbf{w}]_{\mathcal{B}}$$

Orthonormal bases always exist (Gram-Schmidt) so any inner product looks like the dot product in an appropriate basis.

Classifying inner products

What are all the inner products on $\mathbb{R}^n$?

In the preceding discussion, we could have taken $V = \mathbb{R}^n$ and the basis $\mathcal{B}$ to be the standard basis.

Then the inner product $\langle \, , \, \rangle$ is determined by its Gram matrix $[\langle \mathbf{e}_i, \mathbf{e}_j \rangle]$, and is given explicitly as

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T [\langle \mathbf{e}_i, \mathbf{e}_j \rangle] \mathbf{w}$$

Classifying inner products

The matrix $[\langle \mathbf{e}_i, \mathbf{e}_j \rangle]$ is always symmetric — recall this means equal to its transpose — because the inner product is commutative.

Conversely, for any symmetric matrix M, the map

$$\begin{array}{rcl} \mathbb{R}^n \times \mathbb{R}^n & \to & \mathbb{R} \\ \mathbf{v}, \mathbf{w} & \mapsto & \mathbf{v}^T M \mathbf{w} \end{array}$$

is automatically distributive and commutative.

To classify inner products in terms of matrices, it remains to understand when this is positive.

What does positive mean?

Example

The matrix $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ is positive. Indeed,

$$[\,a \;\; b\,]\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = 2a^2 + 2b^2 - 2ab = (a-b)^2 + a^2 + b^2$$

which vanishes only when $a = b = 0$.

What does positive mean?

Example

The matrix $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ is not positive. Indeed,

$$[1 \quad -1] \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2 - 4 + 2 = 0$$

What does positive mean?

More generally, $\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} x & y \\ y & z \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2x + b^2z - 2aby =$

$$b^2 \left(x \left(\frac{a}{b}\right)^2 - 2\left(\frac{a}{b}\right) y + z \right)$$

This quantity will always be positive so long as x is positive, and the parenthetical quantity has no real roots in the variable b/a, which is true so long as the discriminant $4y^2 - 4xz < 0$.

Note this quantity is -4 times the determinant of $\begin{bmatrix} x & y \\ y & z \end{bmatrix}$.

What does positive mean?

We saw that $\begin{bmatrix} x & y \\ y & z \end{bmatrix}$ was positive when $x > 0$ and $xz - y^2 > 0$.
Note that in this case, since $y^2 > 0$, also $z > 0$.

The eigenvalues of the above matrix are the roots of
$t^2 - (x+z)t + (xz - y^2)$, i.e.,

$$2t = (x+z) \pm \sqrt{(x+z)^2 - 4(xz - y^2)} = (x+z) \pm \sqrt{(x-z)^2 + y^2}$$

From the second expression, we see both are real numbers; from the first, since x and z and $xz - y^2 > 0$, we see they are positive.