

The Representer Theorem

David S. Rosenberg

New York University

February 13, 2018

1 Inner Product Spaces and Projections (Hilbert Spaces)

2 Representer Theorem

Inner Product Spaces and Projections (Hilbert Spaces)

Inner Product Space (or “Pre-Hilbert” Spaces)

An **inner product space** (over reals) is a vector space \mathcal{V} and an **inner product**, which is a mapping

$$\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \rightarrow \mathbf{R}$$

that has the following properties $\forall x, y, z \in \mathcal{V}$ and $a, b \in \mathbf{R}$:

- Symmetry: $\langle x, y \rangle = \langle y, x \rangle$
- Linearity: $\langle ax + by, z \rangle = a \langle x, z \rangle + b \langle y, z \rangle$
- Positive-definiteness: $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0 \iff x = 0$.

Norm from Inner Product

For an inner product space, we define a norm as

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Example

\mathbf{R}^d with standard Euclidean inner product is an inner product space:

$$\langle x, y \rangle := x^T y \quad \forall x, y \in \mathbf{R}^d.$$

Norm is

$$\|x\| = \sqrt{x^T x}.$$

What norms can we get from an inner product?

Theorem (Parallelogram Law)

A norm $\|\cdot\|$ can be written in terms of an inner product on \mathcal{V} iff $\forall x, x' \in \mathcal{V}$

$$2\|x\|^2 + 2\|x'\|^2 = \|x + x'\|^2 + \|x - x'\|^2,$$

and if it can, the inner product is given by the **polarization identity**

$$\langle x, x' \rangle = \frac{\|x\|^2 + \|x'\|^2 - \|x - x'\|^2}{2}.$$

Example

ℓ_1 norm on \mathbf{R}^d is NOT generated by an inner product. [Exercise]

Is ℓ_2 norm on \mathbf{R}^d generated by an inner product?

Orthogonality (Definitions)

Definition

Two vectors are **orthogonal** if $\langle x, x' \rangle = 0$. We denote this by $x \perp x'$.

Definition

x is orthogonal to a set S , i.e. $x \perp S$, if $x \perp s$ for all $s \in S$.

Pythagorean Theorem

Theorem (Pythagorean Theorem)

If $x \perp x'$, then $\|x + x'\|^2 = \|x\|^2 + \|x'\|^2$.

Proof.

We have

$$\begin{aligned}\|x + x'\|^2 &= \langle x + x', x + x' \rangle \\ &= \langle x, x \rangle + \langle x, x' \rangle + \langle x', x \rangle + \langle x', x' \rangle \\ &= \|x\|^2 + \|x'\|^2.\end{aligned}$$



Projection onto a Plane (Rough Definition)

- Choose some $x \in \mathcal{V}$.
- Let M be a subspace of inner product space \mathcal{V} .
- Then m_0 is the **projection of x onto M** ,
 - if $m_0 \in M$ and is the closest point to x in M .
- In math: For all $m \in M$,

$$\|x - m_0\| \leq \|x - m\|.$$

Hilbert Space

- Projections exist for all finite-dimensional inner product spaces.
- We want to allow infinite-dimensional spaces.
- Need an extra condition called **completeness**.
- A space is **complete** if all Cauchy sequences in the space converge.

Definition

A **Hilbert space** is a complete inner product space.

Example

Any finite dimensional inner product space is a Hilbert space.

The Projection Theorem

Theorem (Classical Projection Theorem)

- \mathcal{H} a Hilbert space
- M a closed subspace of \mathcal{H} (picture a hyperplane through the origin)
- For any $x \in \mathcal{H}$, there **exists a unique** $m_0 \in M$ for which

$$\|x - m_0\| \leq \|x - m\| \quad \forall m \in M.$$

- This m_0 is called the **[orthogonal] projection of x onto M** .
- Furthermore, $m_0 \in M$ is the projection of x onto M iff

$$x - m_0 \perp M.$$

Projection Reduces Norm

Theorem

Let M be a closed subspace of \mathcal{H} . For any $x \in \mathcal{H}$, let $m_0 = \text{Proj}_M x$ be the projection of x onto M . Then

$$\|m_0\| \leq \|x\|,$$

with equality only when $m_0 = x$.

Proof.

$$\begin{aligned}\|x\|^2 &= \|m_0 + (x - m_0)\|^2 \text{ (note: } x - m_0 \perp m_0 \text{ by Projection theorem)} \\ &= \|m_0\|^2 + \|x - m_0\|^2 \text{ by Pythagorean theorem} \\ \|m_0\|^2 &= \|x\|^2 - \|x - m_0\|^2\end{aligned}$$

Then $\|x - m_0\|^2 \geq 0$ implies $\|m_0\|^2 \leq \|x\|^2$. If $\|x - m_0\|^2 = 0$, then $x = m_0$, by definition of norm. □

Representer Theorem

Generalize from SVM Objective

- SVM objective:

$$\min_{w \in \mathbf{R}^d} \frac{1}{2} \|w\|^2 + \frac{c}{n} \sum_{i=1}^n \max(0, 1 - y_i [\langle w, x_i \rangle]).$$

- Generalized objective:

$$\min_{w \in \mathcal{H}} R(\|w\|) + L(\langle w, x_1 \rangle, \dots, \langle w, x_n \rangle),$$

where

- $R: [0, \infty) \rightarrow \mathbf{R}$ is nondecreasing (**Regularization term**)
- and $L: \mathbf{R}^n \rightarrow \mathbf{R}$ is arbitrary. (**Loss term**)

General Objective Function for Linear Hypothesis Space (Details)

- Generalized objective:

$$\min_{w \in \mathcal{H}} R(\|w\|) + L(\langle w, x_1 \rangle, \dots, \langle w, x_n \rangle),$$

where

- $w, x_1, \dots, x_n \in \mathcal{H}$ for some Hilbert space \mathcal{H} . (We typically have $\mathcal{H} = \mathbf{R}^d$.)
- $\|\cdot\|$ is the norm corresponding to the inner product of \mathcal{H} . (i.e. $\|w\| = \sqrt{\langle w, w \rangle}$)
- $R: [0, \infty) \rightarrow \mathbf{R}$ is nondecreasing (**Regularization term**), and
- $L: \mathbf{R}^n \rightarrow \mathbf{R}$ is arbitrary (**Loss term**).

General Objective Function for Linear Hypothesis Space (Details)

- **Generalized objective:**

$$\min_{w \in \mathcal{H}} R(\|w\|) + L(\langle w, x_1 \rangle, \dots, \langle w, x_n \rangle),$$

- What's “linear”?
- The prediction/score function $x \mapsto \langle w, x_i \rangle$ is linear – in what?
 - in parameter vector w , and
 - in the feature vector x_i .
- Why? [Real-valued] inner products are linear in each argument.
- **The important part is the linearity in the parameter w .**

General Objective Function for Linear Hypothesis Space (Details)

- **Generalized objective:**

$$\min_{w \in \mathcal{H}} R(\|w\|) + L(\langle w, x_1 \rangle, \dots, \langle w, x_n \rangle),$$

- Ridge regression and SVM are of this form.
- What if we penalize with $\lambda\|w\|_2$ instead of $\lambda\|w\|_2^2$? Yes!.
- What if we use lasso regression? No! ℓ_1 norm does not correspond to an inner product.

The Representer Theorem

Theorem (Representer Theorem)

Let

$$J(w) = R(\|w\|) + L(\langle w, x_1 \rangle, \dots, \langle w, x_n \rangle),$$

where

- $w, x_1, \dots, x_n \in \mathcal{H}$ for some Hilbert space \mathcal{H} . (We typically have $\mathcal{H} = \mathbf{R}^d$.)
- $\|\cdot\|$ is the norm corresponding to the inner product of \mathcal{H} . (i.e. $\|w\| = \sqrt{\langle w, w \rangle}$)
- $R: [0, \infty) \rightarrow \mathbf{R}$ is nondecreasing (**Regularization term**), and
- $L: \mathbf{R}^n \rightarrow \mathbf{R}$ is arbitrary (**Loss term**).

If $J(w)$ has a minimizer, then it **has a minimizer of the form** $w^* = \sum_{i=1}^n \alpha_i x_i$.
[If R is strictly increasing, then all minimizers have this form. (Proof in homework.)]

The Representer Theorem (Proof)

- 1 Let w^* be a minimizer.
- 2 Let $M = \text{span}(x_1, \dots, x_n)$. [the “span of the data”]
- 3 Let $w = \text{Proj}_M w^*$. So $\exists \alpha$ s.t. $w = \sum_{i=1}^n \alpha_i x_i$.
- 4 Then $w^\perp := w^* - w$ is orthogonal to M .
- 5 Projections decrease norms: $\|w\| \leq \|w^*\|$.
- 6 Since R is nondecreasing, $R(\|w\|) \leq R(\|w^*\|)$.
- 7 By (4), $\langle w^*, x_i \rangle = \langle w + w^\perp, x_i \rangle = \langle w, x_i \rangle$.
- 8 $L(\langle w^*, x_1 \rangle, \dots, \langle w^*, x_n \rangle) = L(\langle w, x_1 \rangle, \dots, \langle w, x_n \rangle)$
- 9 $J(w) \leq J(w^*)$.
- 10 Therefore $w = \sum_{i=1}^n \alpha_i x_i$ is also a minimizer.

Q.E.D.