The Kernel of a Linear Loss Functional

Adarsh Senthil

November 21, 2023

Abstract

The kernel of a linear loss functional is given by the set of functions orthogonal to its inverse Fourier transform. If the range of the loss functional is the non-negative real numbers, its global minimum is zero, which implies that the functions in the kernel are the models that minimize the loss functional.

In machine learning, a loss functional is used to measure the training error of a model [4:129]. Loss functionals are usually of the form $L(\mathbf{y}, f(\mathbf{x}))$, where $\mathbf{x}$ is an input vector, $\mathbf{y}$ is a label vector and f is a model. Some examples are mean squared error $L(\mathbf{y}, f(\mathbf{x})) = \frac{1}{2n}||\mathbf{y} - f(\mathbf{x})||^2$ and cross-entropy $L(\mathbf{y}, f(\mathbf{x})) = -\langle \mathbf{y}, \log(f(\mathbf{x}))\rangle$. The main objective of training is to find the model that minimizes the loss functional. If the range of the loss functional is the set of non-negative real numbers, its global minimum is zero, which implies that the model that solves the loss functional, minimizes it. Additionally, if the loss functional is linear, the set of models that minimize the training error is given by the kernel of the loss functional.

Let $\mathcal{H}$ be a Hilbert space of models with Hilbert basis $\mathcal{O} = \{u_1, u_2, \dots\}$. Let $L(\mathbf{y}, f(\mathbf{x})) \equiv L(f)$ be a bounded linear functional on $\mathcal{H}$. We have,

$$\begin{aligned} L(f) &= L(\sum_{k=1}^{\infty} \langle f, u_k \rangle u_k) \\ &= \sum_{k=1}^{\infty} \langle f, u_k \rangle L(u_k) \end{aligned}$$

For any bounded linear functional L, the Riesz representation theorem [1:333] states that there is a unique $r_L \in \mathcal{H}$ such that $L(f) = \langle f, r_L \rangle$ for all $f \in \mathcal{H}$. Thus,

$$\begin{aligned} ker(L) &= \{f \in H \mid L(f) = 0\} \\ &= \{f \in H \mid \langle f, r_L \rangle = 0\} \\ &= \{r_L\}^{\perp} \end{aligned}$$

Let $\mathcal{H} = L^2(R^n)$. Every $f \in \mathcal{H}$ can be represented by $f(\mathbf{x}) = \frac{1}{\sqrt{2\pi}} \int_{R^n} \langle f, e^{i\mathbf{k}\cdot\mathbf{x}} \rangle e^{i\mathbf{k}\cdot\mathbf{x}} \, d\mathbf{k}$ [2:265, 3:175] where, $\langle f, g \rangle = \int_{R^n} f(\mathbf{x}) g(\mathbf{x}) \, d\mathbf{x}$. Since $\langle r_L, u_k \rangle = L(u_k)$,

$$\begin{aligned} \int_{R^n} r_L(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}} \, d\mathbf{x} &= L(\mathbf{y}, e^{i\mathbf{k}\cdot\mathbf{x}}) \\ \frac{1}{\sqrt{2\pi}} \int_{R^n} r_L(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}} \, d\mathbf{x} &= \frac{1}{\sqrt{2\pi}} L(\mathbf{y}, e^{i\mathbf{k}\cdot\mathbf{x}}) \\ \mathcal{F}[r_L(\mathbf{x})] &= \frac{1}{\sqrt{2\pi}} L(\mathbf{y}, e^{i\mathbf{k}\cdot\mathbf{x}}) \\ \mathcal{F}^{-1}[\mathcal{F}[r_L(\mathbf{x})]] &= \mathcal{F}^{-1}[\tfrac{1}{\sqrt{2\pi}} L(\mathbf{y}, e^{i\mathbf{k}\cdot\mathbf{x}})] \\ \implies r_L(\mathbf{x}) &= \frac{1}{\sqrt{2\pi}} \mathcal{F}^{-1}[L(\mathbf{y}, e^{i\mathbf{k}\cdot\mathbf{x}})] \end{aligned}$$

Therefore, the kernel of the linear loss functional $L(\mathbf{y}, f(\mathbf{x}))$ is,

$$\begin{aligned} ker(L) &= \{f(\mathbf{x}) \in L^2(R^n) \mid \langle f(\mathbf{x}), \mathcal{F}^{-1}[L(\mathbf{y}, e^{i\mathbf{k}\cdot\mathbf{x}})] \rangle = 0\} \\ &= \{f(\mathbf{x}) \in L^2(R^n) \mid \int_{R^n} f(\mathbf{x}) \mathcal{F}^{-1}[L(\mathbf{y}, e^{i\mathbf{k}\cdot\mathbf{x}})] \, d\mathbf{x} = 0\} \\ &= \{\mathcal{F}^{-1}[L(\mathbf{y}, e^{i\mathbf{k}\cdot\mathbf{x}})]\}^{\perp} \end{aligned}$$

References

1. Steven Roman. *Advanced Linear Algebra.* Springer, New York, 2nd edition, 2005.

2. Peter J. Olver. *Introduction to Partial Differential Equations.* Springer International Publishing Switzerland, 2016.

3. Victor Ivrii. *Partial Differential Equations.* Department of Mathematics, University of Toronto, Canada, 2021.

4. Aston Zhang, Zachary C. Lipton, Mu Li, Alexander J. Smola. *Dive into Deep Learning*.