Learning with Reproducing Kernel Banach Spaces

Haizhang Zhang and Jun Zhang

Abstract. The major obstacle in building Banach space methods for machine learning is the lack of an inner product. We give justifications of substituting inner products with semi-inner-products in Banach spaces as a remedy. By using semi-inner-products, we are able to establish the notion of reproducing kernel Banach spaces (RKBS), and develop regularized learning schemes in the spaces.

Mathematics Subject Classification (2010). Primary 46E15; Secondary 46E22.

Keywords. Machine learning, Banach spaces, sparse approximation, reproducing kernels.

1. Introduction

In the task of learning a function dependency from finite sample data, patterns are usually preprocessed in order to obtain their features. In machine learning, extracting features by mapping the patterns into a Hilbert space is dominant. There are many advantages of this approach thanks to the existence of an inner product in a Hilbert space. In particular, the similarity between patterns can be measured by the inner product of their features in the Hilbert space. This leads to reproducing kernels and gives birth to the popular and successful kernel methods for machine learning [3, 12, 13].

There are some occasions where it might be more appropriate to use a Banach space, a generalization of Hilbert spaces. Firstly, Hilbert spaces is a very limited class of Banach spaces. Any two Hilbert spaces over a common number field with the same dimension are isometrically isomorphic. By reaching out to other Banach spaces, one obtains more variety in geometric

This work was partially supported by the Natural Science Foundation of China under grants 11222103 and 11571377, and by the US Army Research Office under grant W911NF-12-1-0163.

structures and norms that are potentially useful for learning and approximation. Secondly, many training data come with intrinsic structures that make them impossible to be embedded into a Hilbert space. Learning algorithms based on features in a Hilbert space may not work well for them. Thirdly, in some applications, a norm from a Banach space is invoked without being induced from an inner product for some particular purpose. A typical example is the linear programming regularization in coefficient-based regularization for machine learning [12]:

$$\min_{\boldsymbol{c}\in\mathbb{R}^n} Q(K^{\mathbf{x}}(x_1)\boldsymbol{c}, K^{\mathbf{x}}(x_2)\boldsymbol{c}, \cdots, K^{\mathbf{x}}(x_n)\boldsymbol{c}) + \lambda\|\boldsymbol{c}\|_1, \tag{1.1}$$

where $\mathbf{x} := (x_j : j \in \mathbb{N}_n)$ with $\mathbb{N}_n := \{1, 2, \ldots, n\}$ is a given sequence of sampling points, K is a positive-definite reproducing kernel, and $K^{\mathbf{x}}(x_j)$ denotes the row vector $(K(x_i, x_j) : i \in \mathbb{N}_n)$, Q is a loss function, λ is a positive regularization parameter, and $\|\boldsymbol{c}\|_1 := \sum_{j=1}^n |c_j|$ is employed to obtain sparsity in the resulting minimizer.

There has been research in understanding learning of functions in Banach spaces. Minimizing a loss function subject to a regularization condition on a norm in Banach space was studied by [1, 10, 11, 17]. References [4] and [7] considered on-line learning in finite-dimensional Banach spaces, and learning of an L^p function, respectively. Classifications in Banach spaces, and more generally in metric spaces were discussed in [1, 2, 6, 14].

The major obstacle in learning with Banach spaces is caused by the absence of an inner product. As a consequence, kernel methods were not developed in those studies. In particular, it is unknown whether the linear programming regularization (1.1) results from a minimization problem in an infinite-dimensional Banach space. As a consequence, in the learning rate estimates, the hypothesis error will not go away automatically as it does in the reproducing kernel Hilbert space (RKHS) case. In Banach spaces, semi-inner-products [9, 5] in mathematics seem to be a natural substitute for the inner product. The purpose of this note is to introduce this useful tool for developing Banach space methods for machine learning. To this end, we shall give the definition of semi-inner products and justify their capabilities of substituting the important roles of inner products in Section 2. The notion of reproducing kernel Banach spaces (RKBS) was recently established in [15, 16]. We shall review the main results in Section 3.

2. Semi-inner-products

Semi-inner-products were introduced for the purpose of extending Hilbert space type arguments to Banach spaces [9, 5]. A semi-inner-product on a Banach space $\mathcal{B}$ is a function, usually denoted by $[\cdot,\cdot]_{\mathcal{B}}$, from $\mathcal{B}\times\mathcal{B}$ to $\mathbb{R}$ such that for all $f, g, h \in \mathcal{B}$ and $\alpha, \beta \in \mathbb{R}$

1. (linearity with respect to the first variable) $[\alpha f + \beta g, h]_{\mathcal{B}} = \alpha[f, h]_{\mathcal{B}} + \beta[g, h]_{\mathcal{B}}$;

2. (norm compatibility) $[f, f] = \|f\|_{\mathcal{B}}^2$, where $\|f\|_{\mathcal{B}}$ denotes the norm of f in $\mathcal{B}$;
3. (homogeneity with respect to the second variable) $[f, \alpha g]_{\mathcal{B}} = \alpha[f, g]_{\mathcal{B}}$;
4. (Cauchy-Schwartz inequality) $|[f, g]_{\mathcal{B}}| \leq [f, f]_{\mathcal{B}}^{1/2}[g, g]_{\mathcal{B}}^{1/2}$.

A Banach space always has a semi-inner-product [9, 5]. We see that the only property of inner products that a semi-inner-product is not required to possess is symmetry, that is, $[f, g] \neq [g, f]$ in generally. Semi-inner-products were invoked in the machine learning context by [2] for the study of large margin hyperplane classification in Banach spaces. Below we give justifications for the significant roles that semi-inner-products could play in learning and approximation in Banach spaces.

The classical Riesz representation theorem states that every continuous linear functional on a Hilbert space is representable as an inner product. This theoretical result is of fundamental importance to kernel methods for machine learning, in which finite sample data are usually modeled as the point evaluations of the desired function at some inputs. In the general regularization framework for machine learning, the desired function is approximated by functions from an RKHS through a regularized minimization problem. In an RKHS, point evaluations are continuous linear functionals. It follows by the Riesz representation theorem that the point evaluation can be represented by the inner product with a kernel function. This is the starting point for the development of kernel methods for machine learning. For Banach spaces with appropriate conditions, we have an analogue to the classical Riesz representation theorem. Some definitions and notations are needed to present this fundamental fact.

A Banach space $\mathcal{B}$ is *uniformly convex* if for all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\|f + g\|_{\mathcal{B}} \leq 2 - \delta \text{ for all } f, g \in \mathcal{B} \text{ with } \|f\|_{\mathcal{B}} = \|g\|_{\mathcal{B}} = 1 \text{ and } \|f - g\|_{\mathcal{B}} \geq \varepsilon.$$

We also say that $\mathcal{B}$ is *uniformly Fréchet differentiable* if for all $f, g \in \mathcal{B} \setminus \{0\}$

$$\lim_{t\in\mathbb{R},\, t\to 0} \frac{\|f + tg\|_{\mathcal{B}} - \|f\|_{\mathcal{B}}}{t} \tag{2.1}$$

exists and the limit is approached uniformly for all f, g in the unit ball of $\mathcal{B}$. For simplicity, we call a Banach space *uniform* if it is both uniformly convex and uniformly Fréchet differentiable.

Let $\mathcal{B}$ be a Banach space with the semi-inner-product $[\cdot, \cdot]_{\mathcal{B}}$. For each $f \in \mathcal{B}$, the mapping sending $g \in \mathcal{B}$ to $[g, f]_{\mathcal{B}}$ is a continuous linear functional on $\mathcal{B}$. We denote this linear functional by f^* and call it the *dual element* of f. The mapping $f \to f^*$ is called the *duality mapping* from $\mathcal{B}$ to $\mathcal{B}^*$ and is denoted by $\mathcal{J}_{\mathcal{B}}$.

Lemma 2.1. [5] *Let $\mathcal{B}$ be a uniform Banach space. Then it has a unique semi-inner product $[\cdot, \cdot]_{\mathcal{B}}$ and the duality mapping $\mathcal{J}_{\mathcal{B}}$ is bijective and norm-preserving from $\mathcal{B}$ to $\mathcal{B}^*$. In other words, for each μ in the dual space $\mathcal{B}^*$*

there exists a unique $f \in \mathcal{B}$ *such that*

$$\mu(g) = [g, f]_{\mathcal{B}} \text{ for all } g \in \mathcal{B}.$$

and

$$\|f^*\|_{\mathcal{B}^*} = \|f\|_{\mathcal{B}} \text{ for all } f \in \mathcal{B}. \tag{2.2}$$

Moreover,

$$[f^*, g^*]_{\mathcal{B}} := [g, f], \quad f, g \in \mathcal{B} \tag{2.3}$$

defines a semi-inner-product on $\mathcal{B}^*$*.*

3. RKBS

In this section, we give a brief introduction to RKBS established in [15, 16]. This class of Banach spaces of functions are applicable to learning a single task.

In machine learning, we are concerned with Banach space of functions with bounded point evaluation functionals. Let X be an input space. We call $\mathcal{B}$ a Banach space of functions on X if it is a Banach space consisting of certain functions on X such that for every $f \in \mathcal{B}$, $\|f\|_{\mathcal{B}} = 0$ if and only if f vanishes everywhere on X. We stress this definition to make sure that point evaluations are well-defined. For instance, the space of continuous functions on a compact metric space with the usual maximum norm is a Banach space of functions, while $L^p([0,1])$ is not. We call $\mathcal{B}$ a *pre-RKBS* on X if it is a Banach space of functions on X and for each $x \in X$, the point evaluation functional

$$\delta_x(f) := f(x), \quad f \in \mathcal{B} \tag{3.1}$$

is continuous on $\mathcal{B}$. When $\mathcal{B}$ is also a Hilbert space, it is well-known that it possesses a reproducing kernel. In this case, $\mathcal{B}$ is actually an RKHS. The term "pre" is used because there might not exist a reproducing kernel for Banach spaces $\mathcal{B}$. However, when $\mathcal{B}$ is uniform, it does have a reproducing kernel induced by the semi-inner product. In the following, we always denote by $[\cdot, \cdot]_V$ the unique semi-inner-product on a uniform Banach space V.

Theorem 3.1. [15, 16] *Let* $\mathcal{B}$ *be a uniform pre-RKBS on* X*. Then there exists a unique function* $K : X \times X \to \mathbb{R}$ *such that* $K(x, \cdot) \in \mathcal{B}$ *for all* $x \in X$ *and*

$$f(x) = [f, K(x, \cdot)]_{\mathcal{B}} \text{ for all } f \in \mathcal{B} \text{ and } x \in X.$$

In view of the above theorem, we call a uniform pre-RKBS an RKBS and regard the unique function K as the reproducing kernel of $\mathcal{B}$. When $\mathcal{B}$ is also a Hilbert space, it coincides with the reproducing kernel in the usual sense. Similar to the RKHS case, we also have a feature map characterization of reproducing kernels for RKBS.

Theorem 3.2. [15] *A function* $K : X \times X \to \mathbb{R}$ *is the reproducing kernel of some RKBS on* X *if and only if there exists a mapping* Φ *from* X *to a uniform Banach space* $\mathcal{W}$ *such that*

$$K(x, y) = [\Phi(x), \Phi(y)]_{\mathcal{W}} \text{ for all } x, y \in X. \tag{3.2}$$

The function Φ and the space $\mathcal{W}$ are called a pair of *feature map* and *feature space* for K. Feature map representations lead to useful constructions of explicit examples of RKBS. For a mapping Φ from X to a uniform Banach space $\mathcal{W}$, we denote by Φ^* the associated mapping from X to $\mathcal{W}^*$ defined by $\Phi^*(x) := (\Phi(x))^*$, $x \in X$.

Theorem 3.3. [15] *Let $\mathcal{W}$ be a uniform Banach space and Φ a mapping from X to $\mathcal{W}$ such that*

$$\overline{\operatorname{span}}\,\Phi(X) = \mathcal{W}, \quad \overline{\operatorname{span}}\,\Phi^*(X) = \mathcal{W}^*.$$

Then $\mathcal{B} := \{[u, \Phi(\cdot)]_{\mathcal{W}} : u \in \mathcal{W}\}$ equipped with the semi-inner-product

$$\Big[[u, \Phi(\cdot)]_{\mathcal{W}}, [v, \Phi(\cdot)]_{\mathcal{W}}\Big]_{\mathcal{B}} := [u, v]_{\mathcal{W}}$$

and norm

$$\Big\|[u, \Phi(\cdot)]_{\mathcal{W}}\Big\|_{\mathcal{B}} := \|u\|_{\mathcal{W}}$$

is an RKBS on X. Moreover, the reproducing kernel K of $\mathcal{B}$ is given by (3.2).

With the above theoretical preparations, regularized learning schemes were investigated in [15, 16]. We shall present the major result on the representer theorem of the minimizer. Consider a general regularized learning scheme in an RKBS $\mathcal{B}$ on X:

$$\inf_{f \in \mathcal{B}} Q(f(\mathbf{x})) + \lambda \phi(\|f\|_{\mathcal{B}}), \tag{3.3}$$

where $\mathbf{x} := (x_j : j \in \mathbb{N}_n)$ is a sequence of sampling points in X, Q and ϕ are nonnegative loss function and regularization function, and λ is a positive regularization parameter. The loss function Q and the regularization function ϕ should satisfy some minimal requirements for (3.3) to be useful. For this consideration, the learning scheme (3.3) is said to be *acceptable* if both Q and ϕ are continuous and

$$\lim_{t \to \infty} \phi(t) = +\infty. \tag{3.4}$$

The above condition is imposed on ϕ to ensure that it can really put a constraint on the complexity of functions in $\mathcal{B}$ used for learning.

Theorem 3.4. [16] *Let $\mathcal{B}$ be an RKBS on X. Then every acceptable regularized learning scheme (3.3) has at least one minimizer f_0 of the form*

$$f_0^* = \sum_{j=1}^{n} c_j (K(x_j, \cdot))^* \tag{3.5}$$

for some constants $c_j \in \mathbb{R}$. If additionally, ϕ is strictly increasing then every minimizer of (3.3) must have the form (3.5). Furthermore, if Q is convex and ϕ is strictly increasing and strictly convex then an acceptable (3.3) has a unique minimizer, which satisfies (3.5).

When $\mathcal{B}$ is an RKHS then the dual element of a function in $\mathcal{B}$ is itself. Therefore, in this case, Theorem 3.4 recovers the classical representer theorem [8].

References

[1] K. Bennett and E. Bredensteiner, *Duality And Geometry in Svm Classifier.* Proceeding of the Seventeenth International Conference on Machine Learning, P. Langley, eds., Morgan Kaufmann, San Francisco, 2000, 57-64.

[2] R. Der and D. Lee, *Large-margin Classification in Banach Spaces.* JMLR Workshop and Conference Proceedings **2: AISTATS** (2007), 91–98.

[3] T. Evgeniou, C. A. Micchelli, and M. Pontil, *Learning Multiple Tasks with Kernel Methods. J. Mach. Learn. Res.* **6** (2005), 615–637.

[4] C. Gentile, *A New Approximate Maximal Margin Classification Algorithm.* J. Mach. Learn. Res. **2** (2001), 213–242.

[5] J. R. Giles, *Classes of Semi-inner-product Spaces.* Trans. Amer. Math. Soc. **129** (1967), 436–446.

[6] M. Hein, O. Bousquet and B. Schölkopf, *Maximal Margin Classification for Metric Spaces.* J. Comput. System Sci. **71** (2005), 333–359.

[7] D. Kimber and P. M. Long, *On-Line Learning of Smooth Functions of a Single Variable.* Theoret. Comput. Sci. **148** (1995), 141–156.

[8] G. Kimeldorf and G. Wahba, *Some Results on Tchebycheffian Spline Functions.* J. Math. Anal. Appl. **33** (1971), 82–95.

[9] G. Lumer, *Semi-inner-product Spaces.* Trans. Amer. Math. Soc. **100** (1961), 29–43.

[10] C. A. Micchelli and M. Pontil, *A Function Representation for Learning in Banach Spaces.* Learning Theory, 255–269, Lecture Notes in Computer Science, 3120, Springer, Berlin, 2004.

[11] C. A. Micchelli and M. Pontil, *Feature Space Perspectives for Learning the Kernel.* Machine Learning **66** (2007), 297–319.

[12] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* MIT Press, Cambridge, Mass, 2002.

[13] V. N. Vapnik, *Statistical Learning Theory.* Wiley, New York, 1998.

[14] U. von Luxburg and O. Bousquet, *Distance-based Classification with Lipschitz Functions.* J. Mach. Learn. Res. **5** (2004), 669–695.

[15] H. Zhang, Y. Xu, and J. Zhang, *Reproducing Kernel Banach Spaces for Machine Learning.* J. Mach. Learn. Res. **10** (2009), 2741–2775.

[16] H. Zhang and J. Zhang, *Regularized Learning in Banach Spaces as an Optimization Problem: Representer Theorems.* J. Global Optim. **54** (2012), 235–250.

[17] T. Zhang, *On the Dual Formulation of Regularized Linear Systems with Convex Risks.* Machine Learning **46** (2002), 91–129.

Haizhang Zhang
Sun Yat-sen University
Guangzhou 510275, China
e-mail: `zhhaizh2@sysu.edu.cn`

Jun Zhang
University of Michigan
Ann Arbor, MI 48109, USA
e-mail: `junz@umich.edu`