Summary and Overview of “Shannon’s Sampling Theorem II”

By Stephen Thacker

Goals and Scope of Presentation

The goal of the presentation is to summarize the key results of Smale's paper "Shannon's Sampling Theorem II: A Connection to Learning Theory"

Due to the nature of this presentation, the nature of this course, and presentation time constraints, I will be focusing on the concepts most relevant to the topics discussed in this course.

The main focus of this paper is the interplay between Smale's Generalized version of Shannon's Sampling Theorem and Regression Analysis

Overview

Regression Analysis is one method in data science that is used to determine a classification function.

For this reason, it is meaningful to investigate new approaches to studying regression functions.

In this paper, Smale starts by describing a Hilbert Space oriented function reconstruction method using point values. He then takes this result, extends it to Frames in a Reproducing Kernel Hilbert Space Framework, illustrates the interplay between his generalized version of Shannon's Sampling Theorem in RKHS and regression functions, and proves some error bounds related to the approximation of regression functions.

Regression Function

$$\mathcal{E}(f) := \int_Z (f(x) - y)^2 d\rho.$$

$$f_\rho(x) := \int_Y y d\rho(y|x)$$

$$Z := X \times Y \text{ with } Y := \mathbb{R}.$$

With Regression problems, there is an **error function** and this is minimized by a **regression function**

Smale summarizes the goal of solving Regression Problems in the following way, “The problem with a regression function is to find good approximations from a set of random samples”

As we will see on the next slides, Smale now shows that what we have covered so far can be used to approximate functions solving Regression analysis problems.

Function Reconstruction with Point Evaluation

Before discussing regression analysis, we need some preparation.

Function Reconstruction from Point Values

Let $\mathcal{H}$ be a Hilbert space of continuous functions on a complete metric space X and the inclusion $J : \mathcal{H} \to C(X)$ is bounded with $\|J\| < \infty$.

Smale begins by defining a **sampling operator** and using the Riesz Representation Theorem to define Function E_x within H

$$f(x) = < f, E_x >_{\mathcal{H}}, \qquad \forall f \in \mathcal{H}. \tag{2.1}$$

Let $\overline{x}$ be a discrete subset of X. Define the sampling operator $S_{\overline{x}} : \mathcal{H} \to \ell^2(\overline{x})$ by

$$S_{\overline{x}}(f) = \big(f(x)\big)_{x \in \overline{x}}.$$

x represents the sample points in the domain, used for function reconstruction

Function Reconstruction from Point Values

Using the E_x functions in the previous slide, Smale constructs the adjoint operator for S

$$S_{\overline{x}}^T c = \sum_{x \in \overline{x}} c_x E_x, \qquad \forall c \in \ell^2(\overline{x}). \qquad y_x = f^*(x) + \eta_x$$

Term for noise

Smale also gives us the following problem to solve

$$\widetilde{f} := \arg\min_{f \in \mathcal{H}} \left\{ \sum_{x \in \overline{x}} (f(x) - y_x)^2 + \gamma \|f\|_{\mathcal{H}}^2 \right\}$$

This problem is essentially saying "lets find the function in our Hilbert Space that minimizes the least squares problem related to our sampled values".

Function Reconstruction from Point Values

Smale then goes on to give a condition related to the sampling operator that guarantees a solution for our problem.

Theorem 1. *If $S_{\bar{x}}^T S_{\bar{x}} + \gamma I$ is invertible, then $\widetilde{f}$ exists, is unique and*

$$\widetilde{f} = Ly, \qquad L := \left(S_{\bar{x}}^T S_{\bar{x}} + \gamma I\right)^{-1} S_{\bar{x}}^T.$$

He then goes on to create estimates to bound the error of this function reconstruction.

For the sake of time, these will not be included.

So, now, we would like to extend these results to frames and draw a connection to learning theory.

An Extension of this Result to Frames

$$\sum_{x\in\overline{x}} |< f, E_x >_{\mathcal{H}}|^2 \le B\|f\|^2_{\mathcal{H}}, \qquad \forall f \in \mathcal{H}.$$

We see that by the boundedness of the Sampling operator, the family of H.S. functions {E_x} is a Bessel sequence

$$f = \sum_{x\in\overline{x}} < f, (S^T_{\overline{x}} S_{\overline{x}})^{-1} E_x >_{\mathcal{H}} E_x, \qquad \forall f \in \mathcal{H}.$$

With an extra Assumption, we see that the Sampling Operator composed with its Adjoint, gives us a Frame Operator

$$\lambda_{\overline{x}} := \inf_{f\in\mathcal{H}} \|S_{\overline{x}} f\|_{\ell^2(\overline{x})} / \|f\|_{\mathcal{H}} > 0$$

$$\tilde{f} := \arg\min_{f\in\mathcal{H}} \left\{ \sum_{x\in\overline{x}} (< f, E_x >_{\mathcal{H}} - y_x)^2 + \gamma\|f\|^2_{\mathcal{H}} \right\}.$$

We now can now replace our point Evaluations
with functionals in our problem.

Reproducing Kernel Hilbert Spaces

We now assume that our space X is compact and we construct a Reproducing Kernel Hilbert Space (RKHS) with kernel K.

$$K_x := K(x, \cdot) \qquad < K_x, g >_K = g(x)$$

$$f_{\mathbf{z},\lambda} := \arg \min_{f \in \mathcal{H}_K} \left\{ \frac{1}{m} \sum_{i=1}^{m} (f(x_i) - y_i)^2 + \lambda \|f\|_K^2 \right\}$$

$$\mathcal{H}_{K,\mathbf{z}} := \operatorname{span}\{K_{x_i}\}_{i=1}^{m}$$

Using the RKHS property, we can view the minimizing function of this problem as a projection onto the subspace $\mathcal{H}_{K,\mathbf{z}}$

Smale observes that our original family of functions {E_x} becomes the family of function {K_x}

Reproducing Kernel Hilbert Space

Smale observes that our current problem we are solving becomes the original problem when you make the substitutions $f^* = f_\rho$. and $E_x = K_x$

$$f_{\mathbf{z},\lambda} := \arg \min_{f \in \mathcal{H}_K} \left\{ \frac{1}{m} \sum_{i=1}^{m} (f(x_i) - y_i)^2 + \lambda \|f\|_K^2 \right\}$$

Problem we are solving

$$\tilde{f} := \arg\min_{f \in \mathcal{H}} \left\{ \sum_{x \in \overline{x}} (f(x) - y_x)^2 + \gamma \|f\|_{\mathcal{H}}^2 \right\}$$

Problem from earlier

$$f_{\mathbf{z},\lambda} := \arg \min_{f \in \mathcal{H}_{K,\mathbf{z}}} \left\{ \sum_{x \in \overline{x}} (f(x) - y_x)^2 + \gamma \|f\|_K^2 \right\}$$

Problem with Substitutions

Note: There are some other technical assumptions that have been omitted for the sake of illustrating the main concept

Reproducing Kernel Hilbert Spaces

Now we find ourselves in our original context, except this time we have used a RKHS structure to bring a regression function into play.

We remember the original Theorem from the beginning:

Theorem 1. *If $S_{\bar{x}}^T S_{\bar{x}} + \gamma I$ is invertible, then $\widetilde{f}$ exists, is unique and*

$$\widetilde{f} = Ly, \qquad L := \left(S_{\bar{x}}^T S_{\bar{x}} + \gamma I\right)^{-1} S_{\bar{x}}^T.$$

And see that it can be applied now to solve our most recent problem

$$f_{\mathbf{z},\lambda} = \left(S_{\bar{x}}^T S_{\bar{x}} + m\lambda I\right)^{-1} S_{\bar{x}}^T y.$$

How Good is This Approximation

In Smale’s paper, what follows is a variety of technical proofs that eventually converge towards the following result

Corollary 5. *Let* $\mathbf{z}$ *be randomly drawn according to* ρ *satisfying* $|y| \leq M$ *almost everywhere. If* f_ρ *is in the range of* L_K*, then for any* $0 < \delta < 1$*, with confidence* $1 - \delta$ *we have*

$$\|f_{\mathbf{z},\lambda} - f_\rho\|_K \leq \widetilde{C}\left(\frac{(\log(4/\delta))^2}{m}\right)^{\frac{1}{6}} \quad \text{by taking} \quad \lambda = \left(\frac{(\log(4/\delta))^2}{m}\right)^{\frac{1}{3}} \tag{8.5}$$

and

$$\|f_{\mathbf{z},\lambda} - f_\rho\|_\rho \leq \widetilde{C}\left(\frac{(\log(4/\delta))^2}{m}\right)^{\frac{1}{4}} \quad \text{by taking} \quad \lambda = \left(\frac{(\log(4/\delta))^2}{m}\right)^{\frac{1}{4}}, \tag{8.6}$$

where $\widetilde{C}$ *is a constant independent of the dimension:*

$$\widetilde{C} := 30\kappa M + 2\kappa^2 M + \|L_K^{-1} f_\rho\|_\rho.$$

Conclusion

The key takeaway is that Smale's work gives an interesting characterization for solving a regression problem and extends his previous work to a characterization with frames.

$$f_{\mathbf{z},\lambda} = \left(S_{\bar{x}}^T S_{\bar{x}} + m\lambda I\right)^{-1} S_{\bar{x}}^T y.$$

We see in this paper that there is an interplay between RKHS and other fields, such as Learning Theory and Regression Analysis.

Furthermore, we see that dimension independent estimates were given for the error of our RKHS approximation.