A General Learning Algorithm for Functions Between Metric Spaces

Kerry M. Soileau

December 21, 2016

Abstract

In this paper we show how to approximate ("learn") a function $f : X \to Y$, where (X, σ) and (Y, ρ) are metric spaces.

Keywords. accuracy, algorithm, approximate, compact, confidence, convergence, distance, error, evaluation, event, function, learn, limit, metric, model, precision, probability, sample, sequence, space, uniformly, .

1 Introduction

Suppose we wish to develop an approximation of a function by repeatedly sampling its domain and evaluating it at the sample points. For definiteness, suppose (X, σ) and (Y, ρ) are given metric spaces, with X infinite and compact, and for any $x \in X$ we have a means of determining a value $f(x) \in Y$, although a convenient, explicit specification of $f : X \to Y$

may not be available. Our approach is to define a succession $f_n : X \to Y$ of approximations to $f : X \to Y$ in the following way: Each function $f_n(x)$ is determined by the first n domain samples, namely $\{x_1, x_2, x_3, \cdots, x_n\}$, together with their respective range values $\{y_1, y_2, y_3, \cdots, y_n\}$. Specifically, we define $f_1(x) = y_1$ for all $x \in X$, and for $n > 1$, for any $x \in X$ such that $\sigma(x, x_j(x)) \leqslant \min_{1 \leqslant i \leqslant n} \sigma(x, x_i)$ has solution $x_j(x) \in \{x_1, x_2, x_3, \cdots, x_n\}$, $f_n(x)$ is defined to equal $f_n(x_j(x))$. If the satisfying value of $x_j(x)$ is not unique, we take the satisfying value with least index j. This completely specifies the functions $f_n : X \to Y$ for $n = 1, 2, 3, \cdots$.

The compactness of X suggests that if f is continuous, we may expect pointwise convergence of f_n to f and thus may use this approach to approximate f pointwise with arbitrary precision by choosing a sufficiently large n. However, in practice it may be advantageous to sacrifice evaluation precision so as to keep n within reasonable limits. In this article we propose an operational tradeoff that provides evaluation accuracy within user-specified limits with a confidence above a user-specified probability. For example, using our technique, the user can stipulate that the evaluation accuracy will remain within .01 with a probability of .99.

To do so, it is obvious that the model must sometimes "forget" a previous evaluation. Our logic for forgetting is rather simple. Fix numbers $\varepsilon \in (0, \infty)$ and $q \in \left[\frac{1}{2}, 1\right)$. The former specifies the acceptable evaluation error, while the latter controls the confidence that the error will be acceptable. Let A_i denote the evaluation points "remembered" after i

evaluations, with A_1 defined to equal $\{x_1\}$. For every nonempty finite $A \subseteq X$, define $\wp(x, A) \equiv \{t \in A; \forall u \in A, \sigma(t, x) \leqslant \sigma(u, x)\}$, and let $u(S) \equiv$ a uniformly randomly chosen member of a finite set S. When a point x_n is evaluated, its image $f(x_n)$ and $f_{n-1}(x_n)$ are compared. If $\rho(f(x_n), f_{n-1}(x_n)) > \varepsilon$, then x_n is remembered, i.e. $A_n = A_{n-1} \cup \{x_n\}$. If $\rho(f(x_n), f_{n-1}(x_n)) \leqslant \varepsilon$, then either x_n is forgotten, i.e. $A_n = A_{n-1}$, or $u(\wp(x_n, A_{n-1}))$ is forgotten, i.e. $A_m = A_{m-1} \sim \{u(\wp(x_m, A_{m-1}))\}$. The probability of the former event is $2 - \frac{1}{q}$, that of the latter is $\frac{1}{q} - 1$. More explicitly, we may write:

Define $A_n \subseteq X$ according to $A_1 = \{x_1\}$ and

$$A_n = \begin{cases} A_{n-1} \cup \{x_n\} & \text{if } \rho(f(u(\wp(x_n, A_{n-1}))), f(x_n)) > \varepsilon \\ A_{n-1} \sim \{u(\wp(x_n, A_{n-1}))\} & \text{with probability } \frac{1}{q} - 1 \text{ if } \rho(f(u(\wp(x_n, A_{n-1}))), f(x_n)) \leqslant \varepsilon \\ A_{n-1} & \text{with probability } 2 - \frac{1}{q} \text{ if } \rho(f(u(\wp(x_n, A_{n-1}))), f(x_n)) \leqslant \varepsilon \end{cases} \tag{1.1}$$

for $n > 1$.

2 The Approximation Function

We now define a sequence of functions $\{f_n : X \to Y\}_{n=1}^{\infty}$. Let $f_n(x) \equiv f(u(\wp(x, A_n)))$ for $n = 1, 2, 3, \cdots$

A simple calculation shows the growth behavior of A_n:

$$|A_n| - |A_{n-1}| = \begin{cases} 1 & \text{if } \rho\left(f_{n-1}\left(x_n\right), f\left(x_n\right)\right) > \varepsilon \\ -1 & \text{with probability } \frac{1}{q} - 1 \text{ if } \rho\left(f_{n-1}\left(x_n\right), f\left(x_n\right)\right) \leqslant \varepsilon \\ 0 & \text{with probability } 2 - \frac{1}{q} \text{ if } \rho\left(f_{n-1}\left(x_n\right), f\left(x_n\right)\right) \leqslant \varepsilon \end{cases}$$

So

$$E\left(|A_n| - |A_{n-1}| \,|\rho\left(f_{n-1}\left(x_n\right), f\left(x_n\right)\right) > \varepsilon\right) = 1$$

$$E\left(|A_n| - |A_{n-1}| \,|\rho\left(f_{n-1}\left(x_n\right), f\left(x_n\right)\right) \leqslant \varepsilon\right) = 1 - \frac{1}{q}$$

Hence

$$E\left(|A_n| - |A_{n-1}|\right) = 1P\left(\rho\left(f_{n-1}\left(x_n\right), f\left(x_n\right)\right) > \varepsilon\right) + \left(1 - \tfrac{1}{q}\right) P\left(\rho\left(f_{n-1}\left(x_n\right), f\left(x_n\right)\right) \leqslant \varepsilon\right)$$
$$= 1 - \tfrac{1}{q} P\left(\rho\left(f_{n-1}\left(x_n\right), f\left(x_n\right)\right) \leqslant \varepsilon\right)$$

Proposition 2.1 *If* $\lim_{n\to\infty} E\left(|A_n| - |A_{n-1}|\right) = 0$, *then*
$\lim_{n\to\infty} P\left(\rho\left(f_{n-1}\left(x_n\right), f\left(x_n\right)\right) \leqslant \varepsilon\right) = q$.

Proof:

$$0 = \lim_{n\to\infty} E\left(|A_n| - |A_{n-1}|\right) = \lim_{n\to\infty} \left(1 - \tfrac{1}{q} P\left(\rho\left(f_{n-1}\left(x_n\right), f\left(x_n\right)\right) \leqslant \varepsilon\right)\right)$$
$$= 1 - \tfrac{1}{q} \lim_{n\to\infty} P\left(\rho\left(f_{n-1}\left(x_n\right), f\left(x_n\right)\right) \leqslant \varepsilon\right)$$

Hence $\lim_{n\to\infty} P\left(\rho\left(f_{n-1}\left(x_n\right), f\left(x_n\right)\right) \leqslant \varepsilon\right) = q$.■

This suggests that if we attain zero expected growth per point of $|A_n|$ in the limit, i.e. $\lim_{n\to\infty} E\left(|A_n| - |A_{n-1}|\right) = 0$, we must expect to get $\lim_{n\to\infty} P\left(\rho\left(f_{n-1}\left(x_n\right), f\left(x_n\right)\right) \leqslant \varepsilon\right) = q$. Therefore we may interpret the parameter q as the limiting "performance" of the approximation, i.e. the

limiting probability that the approximation will return a value within distance ε of the correct value.

3 Bibliography

[1] E.T. Whittaker and G.N. Watson, *A Course of Modern Analysis*, Cambridge University Press, London and New York, 1958.

---

email: kerry@kerrysoileau.com