Universal Piecewise Linear Least Squares Prediction

David Luengo
Dept. de Teoría de la Señal y Comunicaciones of Universidad Carlos III at Madrid email:luengod@ieee.org

Suleyman S. Kozat
ECE Dept. of University of Illinois at Urbana-Champaign e-mail: kozat@ifp.uiuc.edu

Andrew C. Singer [1]
ECE Dept. of University of Illinois at Urbana-Champaign e-mail: singer@ifp.uiuc.edu

Abstract — **We consider the problem of sequential prediction of real-valued sequences using piecewise linear models under the square-error loss function. In this context, we demonstrate a sequential algorithm for prediction whose accumulated squared error for every bounded sequence is asymptotically as small as that of the best fixed predictor for that sequence taken from the class of piecewise linear predictors. We also show that this predictor is optimal in certain settings in a particular min-max sense. This approach can also be applied to the class of piecewise constant predictors, for which a similar universal sequential algorithm can be derived with corresponding min-max optimality.**

I. Summary

In this paper, we consider the problem of predicting a sequence $x^n = \{x[t]\}_{t=1}^n$ as well as the best piecewise linear predictor out of a large, continuous class of piecewise linear predictors. The real-valued sequence x^n is assumed to be bounded, i.e. $|x[t]| \le A$ for some $A < \infty$, for all t. Rather than assuming a statistical ensemble of sequences, and attempting to achieve optimal performance according to some statistical criterion, our goal is to predict any sequence x^n as well as the best predictor out of a large class of predictors.

We first consider the class of fixed scalar piecewise linear predictors as our competition class. For a scalar piecewise linear predictor, the past observation space $x[t-1] \in [-A, A]$ is parsed into K disjoint regions $\mathcal{R}_j$ where $\bigcup_{j=1}^K \mathcal{R}_j = [-A, A]$. At each time t, the competing predictor forms its prediction as $\hat{x}_{w_j}[t] = w_j x[t-1]$, $w_j \in R$, when $x[t-1] \in \mathcal{R}_j$. We assume that the number of regions and the region boundaries are known. Here, we seek to minimize the following regret:

$$\sup_{x^n} \left\{ \sum_{t=1}^n (x[t] - \hat{x}_q[t])^2 - \inf_{\vec{w} \in R^K} l_n(x, \hat{x}_{\vec{w}}) \right\},$$

where $\hat{x}_q[t]$ is the prediction at time t of any sequential algorithm, $l_n(x, \hat{x}_{\vec{w}}) = \sum_{t=1}^n (x[t] - w_{s[t-1]} x[t-1])^2$, $s[t-1]$ is the region indicator variable taking integer values, i.e., $s[t-1] = j$, $j = 1, \ldots, K$, when $x[t-1] \in \mathcal{R}_j$. That is, we wish to obtain a sequential predictor that can predict every sequence x^n as well as the best fixed piecewise linear predictor for that sequence taken from the class of piecewise linear predictors with the given set of regions. Defining K time vectors (or index sequences) of length n_j, $t_j^{n_j} = \{t : s[t-1] = j\}$, with $j = 1, \ldots, K$, and two sequences $d_j^{n_j} = \{x[t_j[k]]\}_{k=1}^{n_j}$, and $x_j^{n_j} = \{x[t_j[k] - 1]\}_{k=1}^{n_j}$, we can define a universal predictor $\tilde{x}_u[n]$ by applying the universal predictor given in [1] for each region separately. The universal predictor is given by $\tilde{x}_u[n] = w_u^{(j)}[n-1] x[n-1]$, with $j = s[n-1]$, and $w_u^{(j)}[n] = R_{d_j x_j}^{n_j} / (R_{x_j x_j}^{n_j+1} + \delta_j)$. Here, n_j is the number of points of x^{n-1} that belong to $\mathcal{R}_j$, $R_{xy}^n = \sum_{t=1}^n x[t] y[t]$, and $\delta_j > 0$ is a constant.

Theorem 1: *Let x^n be a bounded, real-valued sequence, such that $|x[t]| < A$, for all t. Then $l_n(x, \tilde{x}_u) = \sum_{t=1}^n (x[t] - \tilde{x}_u[n])^2$ satisfies*

$$l_n(x, \tilde{x}_u) \le \min_{\vec{w}} \{ l_n(x, \hat{x}_{\vec{w}}) + \vec{w}^T \Delta \vec{w} \} + \sum_{j=1}^K h_j \ln\left(1 + \frac{n_j A_j^2}{\delta_j}\right)$$

with $h_j = \frac{1}{n_j} \sum_{k=1}^K n_{jk} A_k^2$. Here n_{jk} is the number of elements of region k that result from a transition from region j and $\Delta = diag(\delta_1, \ldots, \delta_K)$ is a diagonal matrix with positive entries.

We then obtain a lower bound on the prediction error in the worst-case where we consider a set of K regions such that $x[t] \in \mathcal{R}_j \Leftrightarrow A_{j-1} < |x[t]| < A_j$, i.e. we consider a set of K regions which are concentric around the origin. Note that the upper bound in this case continues to be valid, since we did not make any assumptions on the shape of the regions, other than that inside the j-th region $|x[t]| < A_j$.

Theorem 2: *Let x^n be a bounded, real-valued arbitrary sequence such that $|x[t]| < A$ for all t. Let $\hat{x}_q[t]$ be the predictions from any sequential prediction algorithm. Then for any $C_j > 0$*

$$\inf_{q \in \mathcal{Q}} \sup_{x^n} \left\{ l_n(x, \hat{x}_q) - \inf_{\vec{w} \in R^K} l_n(x, \hat{x}_{\vec{w}}) \right\} \ge \sum_{j=1}^K \frac{2C_j}{2C_j + 1} A_j^2 \ln\left(1 + \frac{n_j - 2}{2C_j}\right),$$

where $\mathcal{Q}$ is the class of all sequential predictors.

Similar results are derived for the class of fixed piecewise p_j-th order linear predictors in each region where the past observation space $[-A, A]^m$ is parsed into K disjoint regions $\mathcal{R}_j$ such that $\bigcup_{i=1}^K \mathcal{R}_j = [-A, A]^m$. At each time t the prediction is formed by $\hat{x}_{\vec{w}_j}[n] = \sum_{k=1}^{p_j} w_{j,k} x[n-k]$ if $\{x[n-1], \ldots, x[n-m]\} \in \mathcal{R}_j$ and $w_{j,k} \in R$ for $j = 1, \ldots, K$, $k = 1, \ldots, p_j$, $p_j \le m$. The universal predictor is given by $\tilde{x}_u[n] = \vec{w}_u^{(j)^T}[n-1] \vec{x}[n-1]$ where $\vec{w}_u^{(j)}[n] = (R_{\vec{x}_j \vec{x}_j}^{n_j} + \delta_j I)^{-1} R_{d_j \vec{x}_j}^{n_j}$, and $\vec{x}[n-1] = [x[n-1], \ldots, x[n-p_j]]^T$. The upper and lower bounds for the vector case are in the same form as the scalar case where each bound for the scalar case is scaled by the prediction order p_j as in [1]. We also observe that the derivation for scalar piecewise linear predictors can be applied to piecewise constant predictors where the competing class now forms predictions as constants such that $\hat{x}_{c_j}[n] = c_j$ when $\{x[t-1], \ldots, x[n-m]\} \in \mathcal{R}_j$ and $c_j \in R$.

References

[1] A. C. Singer, S. S. Kozat, M. Feder, " Universal linear least squares prediction: upper and lower bounds," *IT.*, vol. 48, pp. 2354-2362, Aug 2002;

[1] This work was supported in part by the National Science Foundation, under grants number CCR-0092598 (CAREER), CCR 99-79381, and ITR 00-85929, and Office Of Naval Research under Award No.: NO00140110117