Nonparametric prediction of time series

László Györfi
Department of Computer Science and Information Theory
Budapest University of Technology and Economics
Magyar tudósok körútja 2., Budapest, Hungary, H-1117
gyorfi@szit.bme.hu

Abstract: We study the problem of sequential prediction of a real valued sequence. At each time instant $t = 1, 2, \ldots$, the predictor is asked to guess the value of the next outcome Y_t of a sequence of real numbers $Y_1, Y_2, \ldots$ with knowledge of the pasts $Y_1^{t-1} = (Y_1, \ldots, Y_{t-1})$ (where Y_1^0 denotes the empty string) and the side information vectors $X_1^t = (X_1, \ldots, X_t)$, where $X_t \in \mathbb{R}^d$. Thus, the predictor's estimate, at time t, is based on the value of X_1^t and Y_1^{t-1}. A prediction strategy is a sequence $g = \{g_t\}_{t=1}^{\infty}$ of functions

$$g_t : \left(\mathbb{R}^d\right)^t \times \mathbb{R}^{t-1} \to \mathbb{R}$$

so that the prediction formed at time t is $g_t(X_1^t, Y_1^{t-1})$.

In this paper we assume that $(X_1, Y_1), (X_2, Y_2), \ldots$ is a stationary and ergodic process. After n time instants, the *normalized cumulative prediction error* is

$$L_n(g) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{t=1}^{n} (g_t(X_1^t, Y_1^{t-1}) - Y_t)^2.$$

We show a universally consistent prediction strategy g such that for any stationary ergodic process $\{(X_n, Y_n)\}_{-\infty}^{\infty}$ with $\mathbb{E}\{Y_0^4\} < \infty$,

$$\lim_{n\to\infty} L_n(g) = L^* \quad \text{almost surely,}$$

where

$$L^* \stackrel{\text{def}}{=} \mathbb{E}\left\{\left(Y_0 - \mathbb{E}\{Y_0 \mid X_{-\infty}^0, Y_{-\infty}^{-1}\}\right)^2\right\}$$

is the minimal mean squared error of any prediction for the value of Y_0 based on the infinite past $X_{-\infty}^0, Y_{-\infty}^{-1}$.

The previous prediction may result in universally consistent classification rule as follows. Here Y_i is binary valued, and the classifier formed at time t is $f_t(X_1^t, Y_1^{t-1})$. We show a classification strategy f such that for any stationary ergodic process $\{(X_n, Y_n)\}_{-\infty}^{\infty}$, the *normalized cumulative* $0-1$ *loss* converges to the minimal error probability:

$$\begin{aligned} R_n(f) &\stackrel{\text{def}}{=} \frac{1}{n} \sum_{t=1}^{n} I_{\{f_t(X_1^t, Y_1^{t-1}) \neq Y_t\}} \\ &\to R^* \stackrel{\text{def}}{=} \mathbf{E}\left\{\min\left(\mathbf{P}\{Y_0 = 1 | X_{-\infty}^0, Y_{-\infty}^{-1}\}, \mathbf{P}\{Y_0 = 0 | X_{-\infty}^0, Y_{-\infty}^{-1}\}\right)\right\}, \end{aligned}$$

where $I_{\{\cdot\}}$ denotes the indicator function.