CCIT Report No. 372

On Competitive Prediction and its Relation to Rate-Distortion Theory and to Channel Capacity Theory

Tsachy Weissman and Neri Merhav

ABSTRACT

Consider the normalized cumulative loss of a predictor *F* on the sequence $x^n = (x_1, \ldots, x_n)$, denoted $L_F(x^n)$. For a set of predictors $\mathcal{G}$, let $L(\mathcal{G}, x^n) = \min_{F \in \mathcal{G}} L_F(x^n)$ denote the loss of the best predictor in the class on x^n. Given the stochastic process $\boldsymbol{X} = X_1, X_2, \ldots$, we look at $EL(\mathcal{G}, X^n)$, termed the *competitive predictability of* $\mathcal{G}$ *on* X^n. Our interest is in the optimal predictor set of size *M*, i.e., the predictor set achieving $\min_{|\mathcal{G}| \le M} EL(\mathcal{G}, X^n)$. When *M* is sub-exponential in *n*, simple arguments show that $\min_{|\mathcal{G}| \le M} EL(\mathcal{G}, X^n)$ coincides, for large *n*, with the Bayesian envelope $\min_F EL_F(X^n)$. Our interest is in the behavior, for large *n*, of $\min_{|\mathcal{G}| \le e^{nR}} EL(\mathcal{G}, X^n)$, which we term the *competitive predictability of* **X** *at rate R*. It is shown that under difference loss functions, the competitive predictability of **X** is lower bounded by the Shannon lower bound (SLB) on the distortion-rate function of **X** and upper bounded by the distortion-rate function of any (not necessarily memoryless) innovation process through which the process **X** has an autoregressive representation. This precisely characterizes the competitive predictability whenever **X** can be autoregressively represented via an innovation process for which the SLB is tight (e.g., when **X** is a Gaussian process under squared error loss). We next derive lower and upper bounds on the error exponents, i.e., on the exponential behavior of $\min_{|\mathcal{G}| \le \exp(nR)} \Pr\left(L(G, X^n) > d\right)$, which are shown to be tight for many cases of interest. Finally, the universal setting is considered, where a predictor set is sought which minimizes its worst-case competitive predictability over all sources in a given family. The problem is shown to significantly diverge from its non-universal origin when the effective number of sources in the family grows exponentially with *n*. The optimal predictor set for this problem is shown to be related to the capacity-achieving code-book corresponding to the "channel" from the family of sources to their realizations.

Key words and phrases: Channel capacity, competitive prediction, error exponents, rate distortion theory, redundancy , scandiction, strong converse.