An Introduction to Information Theoretic Learning

Jose C. Principe
BellSouth Professor
Computational NeuroEngineering Laboratory
University of Florida

Abstract

Information Theoretic Learning (ITL) is a training paradigm that utilizes information theoretic measures of the samples (entropy and mutual information) to adapt the parameters of linear or nonlinear networks. Information theory has been applied previously for optimization, but it normally requires the knowledge (or a priori assumption) of the probability density function (pdf). However, ITL is nonparametric, i.e. it exploits directly the information contained in the samples.

In this paper we will show how entropy can be estimated directly from a set of samples using Renyi's quadratic entropy definition. We will define information potential and information force. We will show that the training of linear or nonlinear mappers can be effected with ITL by combining the criterion with the backpropagation (BP) algorithm.

We will also propose two new distance measures between pdfs, which are proxies for the Kullback Leibler divergence (KL). They seem to produce reasonable approximations to KL and can still be computed using the information potential.

Finally, we will present applications of ITL for both unsupervised and supervised learning problems.

Training a MLP Layer-by-Layer with Information Theoretic Learning

Dongxin Xu and Jose C. Principe
Computational NeuroEngineering Laboratory
University of Florida, Gainesville, FL 32611

Purpose: The backpropagation algorithm (BP) has been tied to the multilayer perceptron (MLP) topology since its introduction in 1987. Many authors even describe MLPs as backpropagation networks because there was no other known method to train MLP. This paper proposes a new criterion to train MLPs in a srictly feedforward fashion, i.e. layer-by-layer.

Method: This new training method is based on information theoretic learning (ITL) principles recently developed by our group [1]. The basic concept is associated with adapting each layer's weights such that the mutual information between the desired response and the layer output is maximized. Each layer is modeled as a communication channel where the goal is to transfer from the input to the output the maximum information about the desired response. So training progresses from the input to the output in a feedforward fashion, as opposed to BP where the errors are propagated back from the output to the input. In order to implement information theoretic learning, a new nonparametric estimator of mutual information is required. We propose the Euclidean distance metric coupled with Renyi's quadratic entropy to implement such a criterion [2]. The appeal of this criterion for learning is that it is based on the sample-by-sample metaphor and does not require a priori specification about data distributions. Hence it is quite general.

Results: The paper will briefly describe the Information Theoretic Criterion and will show two examples, one of classification (static data) and another of nonlinear filtering. In both cases ITL solves the problem.

New Aspect of Work: Training MLP layer-by-layer by information-theoretic learning, rather than back-propagating the error from the output layer.

Conclusions: This work shows that, contrary to common knowledge, the well establish method of error backpropagation is not the only way to train multilayered topologies.

[1] Xu D., Principe J., Fisher J., Wu H., A novel measure for independent component analysis, in Proc. ICASSP98, vol II 1161-1164, Seatlle, 1998.

[2]Xu D., Principe J., "Learning from examples with mutual information", Proc. IEEE Workshop on NNSP, 155-164, Cambridge, England, 1998.