A Statistical Analysis Of Neural Computation

John A. Cortese (†)
Rodney M. Goodman
Department Of Electrical Engineering, 116-81
California Institute Of Technology
Pasadena, CA. 91125 U.S.A.

Email : jcort@micro.caltech.edu

(† : Correspondence Author)

Abstract

This paper presents an architecture and learning algorithm for a feedforward neural network implementing a two pattern (image) classifier. By considering the input pixels to be random variables, a statistical binary hypothesis (likelihood ratio) test is implemented. A linear threshold separates $p[X|H_0]$ and $p[X|H_1]$, minimizing a risk function. In this manner, a single neuron is considered as a BSC with the pdf error tails under the threshold yielding the cross-over probability ε. A single layer of neurons is viewed as a parallel bank of independent BSC's, which is equivalent to a single *effective* BSC representing that layer's hypothesis testing performance. A multiple layer network is viewed as a cascade of BSC channels, and which again collapses into a single effective BSC.

The effective BSC channel capacity is examined as the information theoretic ability of the network to extract the single bit of information encoded in the multidimensional input vector and pass the resulting fractional information bit of information to the one dimensional output "decision" variable.

A linear threshold neuron is shown to implement a hard decision decoding scheme. The implementation of a soft decision decoding scheme leads to the introduction of non-linear sigmoid functions. For large input dimensionality, and weakly correlated pixel random variables, the optimal soft decision sigmoid is shown to be a tanh function. The optimum gain for the tanh function is analytically derived from the hypothesis pdf moments, which are available from a labeled training set.

Classical, difficult problems such as Parity, are shown to have hypothesis pdf's $p[X|H_i]$ which differ in higher moments. K-WTA lateral inhibition structures are shown to implement order statistic's on these pdf's. The lateral inhibition operation results in a removal of the lower moments of these pdf's. The transformed pdf's can now be presented to a succeeding layer of neurons, which will effectively implement a linear threshold test on the higher moments.

Throughout the talk, links are made to statistical detection and estimation theory. Connections between the information theory / BSC model and the statistical communication model are made via Shannon's relationship $C = B \log_2 (1 + SNR)$. Key points are illustrated with an example network operating on a NIST database of 32 by 32 binary pixel images of handwritten "0"'s and "1"'s.