Chapter 4.16

Learning Algorithms for RBF Functions and Subspace Based Functions

Lei Xu
Chinese University of Hong Kong & Beijing University, PR China

ABSTRACT

Among extensive studies on radial basis function (RBF), one stream consists of those on normalized RBF (NRBF) and extensions. Within a probability theoretic framework, NRBF networks relates to nonparametric studies for decades in the statistics literature, and then proceeds in the machine learning studies with further advances not only to mixture-of-experts and alternatives but also to subspace based functions (SBF) and temporal extensions. These studies are linked to theoretical results adopted from studies of nonparametric statistics, and further to a general statistical learning framework called Bayesian Ying Yang harmony learning, with a unified perspective that summarizes maximum likelihood (ML) learning with the EM algorithm, RPCL learning, and BYY learning with automatic model selection, as well as their extensions for temporal modeling. This chapter outlines these advances, with a unified elaboration of their corresponding algorithms, and a discussion on possible trends.

DOI: 10.4018/978-1-60960-818-7.ch4.16

BACKGROUND

The renaissance of neural network and then machine learning since the 1980's is featured by two streams of extensive studies, one on multilayer perceptron and the other on radial basis function networks or shortly RBF net. While a multilayer perceptron partitions data space by hyperplanes and then making subsequent processing via nonlinear transform, a RBF net partitions data space into modular regions via local structures, called radial or non-radial basis functions. After extensive studies on multilayer perceptron, studies have been turned to RBF net since the late 1980's and the early 1990's, with a wide range applications (Kardirkamanathan, Niranjan, & Fallside, 1991; Chang & Yang, 1997; Lee, 1999; Mai-Duy & Tran-Cong, 2001; Er, 2002; Reddy & Ganguli, 2003; Lin & Chen 2004; Isaksson, Wisell, & Ronnow, 2005; Sarimveis, Doganis, & Alexandridis, 2006; Guerra & Coelho, 2008; Karami & Mohammadi, 2008).

In the literature of machine learning, advances on RBF net can be roughly divided into two streams. One stems from the literature of mathematics on multivariate function interpolation and spline approximation, as well as Tikhonov type regularization for ill-posed problems, which were brought to neural networks learning by (Powell, 1987; Broomhead & Lowe, 1998; Poggio & Girosi, 1990; Yuille & Grzywacz, 1989) and others. Actually, it is shown that RBF net can be naturally derived from Tikhonov regularization theory, i.e. least square fitting subjected to a rotational and translational constraint term by a different stabilizing operator (Poggio & Girosi, 1990; Yuille & Grzywacz, 1989). Moreover, RBF net has also been shown to have not only the universal approximation ability possessed by multilayer perceptron (Hartman, 1989; Park & Sandberg,1993), but also the best approximation ability (Poggio & Girosi, 1990) and a good generalization property (Botros & Atkeson, 1991).

The other stream can be traced back to Parzen Window estimator proposed in the early 1960's. Studies of this stream progress via active interactions between the literature of statistics and the literature of machine learning. During 1970's-80's, Parzen window has been widely used for estimating probability density, especially for estimating the class densities that are used for Bayesian decision based classification in the literature of pattern recognition. Also, it has been introduced into the literature of neural networks under the name of probabilistic neural net (Specht, 1990). By that time, it has not yet been related to RBF net since it does not directly relates to the above discussed function approximation purpose.

In the literature of neural networks and machine learning, Moody & Darken (1989) made a popular work via Gaussian bases as follows:

$$f(x) = \sum_{j=1}^{k} w_j \varphi_j(x - c_j, \Sigma_j),$$
$$\varphi_j(x - c_j, \Sigma_j) = e^{-0.5(x-c_j)^T \Sigma_j^{-1}(x-c_j)} \Big/ \sum\nolimits_{j=1}^{k} e^{-0.5(x-c_j)^T \Sigma_j^{-1}(x-c_j)} . \tag{1}$$

This normalized type of RBF net is featured by

$$\sum_{j=1}^{k} \varphi_j(x - c_j, \Sigma_j) = 1, \tag{2}$$

and thus usually called Normalized RBF (RBF). Moody & Darken (1989) actually considered a special case $\Sigma_j = \sigma_j^2 I$. Unknowns are learned from a set of samples via a two stage implementation. First, a clustering algorithm (e.g., k-means) is used to estimate c_i as each cluster's center. Second, $\varphi_j(x)$ is calculated for every sample and unknowns w_j are estimated by a least square linear fitting. Nowlan (1990) proceeded along this line with Σ_j considered in its general form such that not only the receptive field of each base function can be elliptic instead of radial symmetrical, but also the well known EM algorithms (Redner & Walker, 1984) are adopted for estimating c_i and Σ_j with better performances than that by a clustering algorithm.

In the nonparametric statistics literature, extensive studies have also been made towards regression tasks under the name of kernel regression estimator, which is an extension of Parzen window estimator from density estimation to statistical regression (Devroye, 1981&87). In (Xu, Krzyzak, & Yuille, 1992&94), kernel regression estimators are shown to be special cases of NRBF