Radial Basis Function Techniques for Regression Analysis of Economic Trends

Marcus L. Roberts[1] and Steven C. Gustafson[1]

[1]Department of Electrical Engineering and Computer Science
Air Force Institute of Technology
Wright-Patterson Air Force Base, OH 45433

Abstract

Economic statistics often invite "good or bad?" and "rising or falling?" questions. Pattern recognition techniques can be applied to economic data to produce objective answers to such questions. This paper analyzes nine American leading economic indicators over the past 20 years. It investigates least squares (LS) fitting of linear models as a baseline and then uses more advanced techniques, such as Gaussian radial basis function (RBF) interpolation and principal component (PC) dimensionality reduction. Results indicate that while linear models provide a general approximation, the inherent variance in the data leads to better approximation by a modified Gaussian RBF, as measured by root mean square (RMS) error.

Keywords: radial basis function, regression analysis, econometrics, pattern recognition, neural network.

1. Introduction

Econometrics is "the application of statistical methods to the study of economic data and problems" [1]. This paper investigates the application of traditional *pattern recognition* methods to the study of economic data and trends. As such, it loosely falls into the realm of econometrics in addition to pattern recognition. Nine American leading economic indicators over the past 20 years are analyzed. These indicators are the Dow Jones Ind. Ave. (DJIA), the NASDAQ index, unemployment rates, interest rates, new housing starts, gross domestic product, consumer price index, producer price index, and oil prices. LS fitting of linear models is the baseline, and more advanced techniques like Gaussian RBF interpolation and PC dimensionality reduction are also employed.

2. Background

The LS approach, i.e., minimizing the sum-squared error between modeled and truth values, is used as a baseline in most regression problems. LS is attractive in its simplicity and low variance in regression problems. As such, it holds to *Occam's razor*, the principle whereby one should prefer a simple model to a complex model, and that this preference should be balanced against the extent to which the model fits the data [2]. For dynamic systems such as those involving economic data, it may be necessary to increase model complexity to produce better curve fitting estimates. But, an overly complex model may yield an unstable system or a system tailored to a specific data set.

The LS solution for an inconsistent system with m equations and n unknowns satisfies [3]

$$A^T A\bar{x} = A^T b \tag{1}$$

if the columns in A are linearly independent. Another technique, slightly more complicated but a better fit to the data, applies RBFs.

Radial basis functions are widely used for non-linear function modeling, and much research has reported on their application [2,4-6]. A typical multi-layer perceptron (MLP) neural network (NN) is based on units which compute a non-linear function of the scalar product of the input vector with a weight vector. However, in RBF, a different NN class, the activation of a hidden unit is determined by the distance between the input vector and a prototype vector [2].

For a regression problem with m data points, a MLP neural network with one hidden layer of n units produces a solution

$$y(x) = \frac{c_1}{1+e^{-(a_1+b_1x)}} + \cdots + \frac{c_n}{1+e^{-(a_n+b_nx)}} \tag{2}$$

and chooses a_i, b_i, and c_i to minimize

$$\left(y(x_1)-y_1\right)^2 + \cdots + \left(y(x_m)-y_m\right)^2, \tag{3}$$

where $y(x_i)$ is the NN output at x_i and y_i is the truth value.

For the same m-data-point regression problem, a RBF NN can be designed to *interpolate* the points y_1, y_2, ... y_m using basis functions that vary from the LS line. The RBF methodology introduces a set of N basis

functions which have the form $\phi\left(\left\|\bar{x}-\bar{x}_n\right\|\right)$, where $\phi(\cdot)$ is some non-linear function and N is the number of data points [2,5]. Hence, there is a basis function at each selected data point, and the n^{th} function depends on the distance (typically Euclidean) between $\bar{x}$ and $\bar{x}_n$. The result of this mapping is the linear combination of the basis functions [2]

$$h(\bar{x})=\sum_n w_n\phi\left(\left\|\bar{x}-\bar{x}_n\right\|\right). \quad (4)$$

Theoretical and empirical studies [5] show that many properties of the interpolating function (in the context of the exact interpolation problem) are relatively insensitive to the exact form of $\phi(\cdot)$ [2]. The most common form is the Gaussian [2]

$$\phi(x)=\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (5)$$

where σ is the parameter that controls the smoothness of the interpolating function, and μ is the mean at which the Gaussian is centered. This Gaussian RBF (GRBF) form is the basis function used herein.

When necessary, PC techniques are used to reduce the dimensionality of the data, since PC analysis (PCA) indicates directions of maximum variance in the data. Typically, pattern recognition systems employ three or less directions of maximum variance (for visualization purposes). By eliminating the other directions of data variance, the system essentially reduces the dimensionality of the data. In PCA, the covariance matrix of the data is found, and the data points are projected onto the one, two, or three eigenvectors corresponding to the largest one, two, or three eigenvalues. This procedure reduces the number of dimensions to one, two, or three, respectively.

3. Methodology

The analysis here involves LS fitting of linear models as well as GRBF interpolation and PCA. The following recipe describes the methodology.

First, collect raw data from the past 20 years for the nine leading economic indicators [7-9]. Economic indicators are modified so increases correspond to positive economic change. Second, create a table where the rows correspond to 81 data records consisting of raw averaged quarterly data from Jan. 1984 to Mar. 2004 and the columns correspond to the nine inputs or "features." Third, leave out one row and normalize each column of the reduced table to zero mean and unit variance. Fourth, use PC techniques to find a linear combination of the inputs in the reduced table that is "most important" in that it has maximum variance. Fifth, fit a LS line of the projection onto the PC axis versus time (81 quarterly values), and use it to predict the left-out output. This projection is stated below as an indicator of economic strength. Although not optimized, it gives some indication of economic progression. Sixth, repeat the steps above until each row is left out, and find the RMS prediction error. Seventh, repeat steps 3 to 6 for quadratic and cubic estimates of the projection onto the PC axis versus time.

Finally, repeat steps 3 to 6 using a GRBF NN interpolation from the LS line, with basis functions of identical variances at the input points. Try many values of variance, searching for the σ that minimizes error. Since there are many data records (81), it is unrealistic to interpolate every data point. Explore the merits of using cluster sizes of three, nine, and 13 points in a "modified" GRBF NN setup.

4. Results

Rather than pursuing an optimal set of economic indicators, this paper focuses on selecting a wide variety of indicators and examining their correlations. Normalized versions of nine variables are used, and the general trend is upwards.

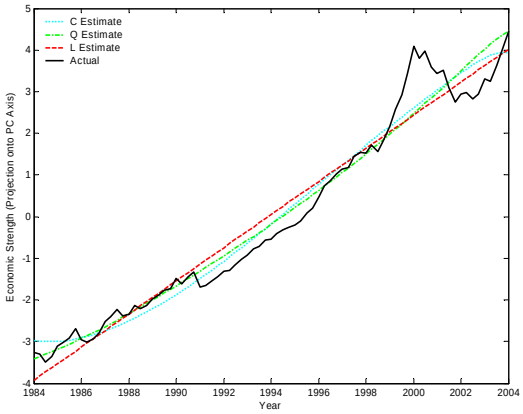

Fig. 1: Economic strength vs. time for the past 20 years. As the polynomial order increases, deviation from the true data (thick black line) decreases, but only slightly.

Least squares fitting of linear models is shown in Fig. 1. A LS line (L) and quadratic (Q) and cubic (C) polynomial approximations are shown to generalize the performance of economic strength (defined as the projection of the normalized economic data onto the PC axis). However, for these approximations, the variance in the data is too great for linear models. Higher order polynomials might approximate the data better, but they would also be more unstable. Although the bias would be lower (they would come closer to interpolating the data), the variance would be greater - an unsatisfactory trade off. Also, instability would

dramatically increase, that is, a small change in input could yield a large change in output. The result would be similar even if a fewer number (than nine) of features were chosen. The polynomial approximations might slightly improve, but they would always be inadequate to precisely model the data.

The RMS error (RMSE) of the approximations drops from 0.50 to 0.42 when moving from a least squares line to a cubic polynomial approximation, which is a slight, but not significant, decrease. The RMSE is calculated by leaving out a data point and squaring the distance between the model and truth at the left out point. Taking the square root of the mean of the squared errors yields the RMSE.

GRBF interpolation provides a smoother estimate of the raw averaged quarterly economic data (see Fig. 2). The number of data records (81) dictates that it is desirable to group the data records in clusters and then use GRBF to interpolate the clusters, essentially a "modified" GRBF. The first question is *how large should the clusters be*? Here, cluster sizes of three, nine, and 13 data points are examined. An odd number is chosen for the cluster size so that an exact integer cluster center exists. These cluster sizes are not optimal; rather, they illustrate different ways to use GRBFs for data interpolation. The next question with GRBF is *what variance should be used*? An excessively small variance produces a rough estimate with a spiked appearance. An excessively large variance interpolates well in the region of the data points but has large swings in the tails of the estimates. Note that the GRBF is designed in this analysis to return to the LS estimate outside the region of the data points (before 1984 and after 2004).

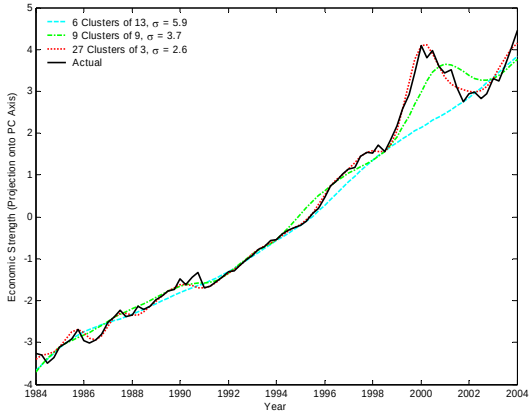

Fig. 2: Economic strength vs. time for the past 20 years. The RMSE values are 0.48, 0.28, and 0.13 for cluster sizes of 13, 9, and 3 points, respectively.

Fig. 2 shows the GRBF estimate of the data. Each cluster size is matched with an optimal variance such that this variance yields a minimum RMSE for the approximation. For this data projection, the optimal σ of each cluster size is listed in the figure legend. In general, as cluster sizes increase, so does the distance between cluster centers and the corresponding variance size required to minimize the RMSE.

Optimal standard deviations for each cluster size are found by plotting the GRBF interpolation of the cluster centers for many σ values. Small variance values produce a rough approximation (higher error) between cluster centers. Large variance values produce a smooth approximation with large swings in the tails (higher error). A plot of cluster RMSE versus σ has a bowl-shaped appearance, and the σ at which the minimum cluster RMSE is located is taken as the optimal standard deviation.

The RMSE values listed in the figure caption are calculated by taking the root-mean of the squared errors between the truth and the plotted RBF estimates. This method most closely approximates the LS RMSE estimates from the linear models above as they are calculated over *all* data points using a "leave-one-out" approach. Here, the "leave-one-out" approach cannot be used for comparison since there are not 81 cluster centers.

The cluster size of 13 appears to be too large to approximate the dynamic fluctuations over the 1999 to 2004 time period. Consequently, its RMSE of 0.48 is nearly the same as the LS RMSE from Fig. 1 (0.50). GRBF with cluster sizes of nine (0.28) and three points (0.13) yield much lower RMSE values than quadratic (0.45) and cubic (0.42) polynomial approximations. However, these RBF RMSE values are for different "optimal" σ values. The list below shows the RMSE for all analyzed regression techniques. GRBF interpolation significantly reduces the RMSE of the modeled output if the cluster size and variance are chosen appropriately.

- Least Squares Line, RMSE = 0.50
- Quadratic Polynomial, RMSE = 0.45
- Cubic Polynomial, RMSE = 0.42
- GRBF 13-Cluster, σ = 5.9, RMSE = 0.48
- GRBF 9-Cluster, σ = 3.7, RMSE = 0.28
- GRBF 3-Cluster, σ = 2.6, RMSE = 0.13

Figures 3 and 4 show the same information as Fig. 2 but with all cluster sizes using the same standard deviation at extremely low and high values. Notice that smaller σ values produce a spiked-looking output and larger σ values produce a smooth estimate with large swings at the tails.

In econometrics, a common goal is to produce an estimate that not only models the data well but can predict future results with some accuracy. The GRBFs here are designed to converge to the LS line, and it is interesting to examine this convergence. Fig. 5 shows an extrapolation of economic strength through 2010

using the LS estimate as a baseline. With optimal σ values, the GRBF estimates quickly and smoothly converge to the LS estimate. However, given sub-optimal σ values, future economic strength cannot be predicted with as much confidence.

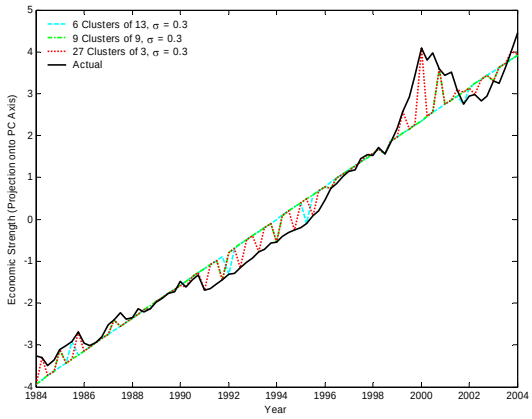

Fig. 3: Same information as Fig. 2 except σ values are equal (σ = 0.3). The RMSE values are 0.49, 0.48, and 0.40 for cluster sizes of 13, nine, and three points, respectively.

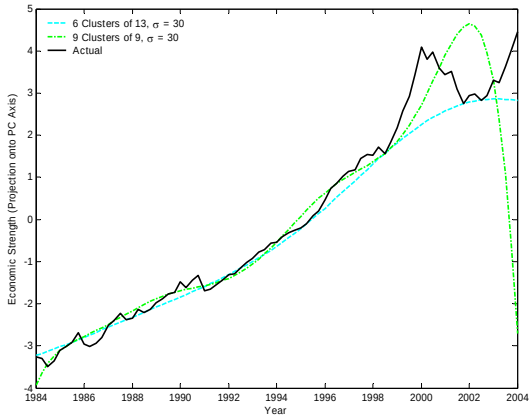

Fig. 4: Same information as Fig. 2 except σ values are equal (σ = 30). The RMSE values are 0.51 and 1.12 for cluster sizes of 13 and nine points, respectively.

Other variations of GRBF estimates can be used for extrapolation and prediction of economic strength, such as regularization, a technique used to control function smoothness [2]. Instead of directly adding the GRBF estimate to the LS solution, either part can be scaled by a regularization term, which is designed to penalize rough estimates. Because smoothness may be a desired attribute in economic prediction, a regularization term may be beneficial.

5. Conclusions

The analysis illustrates two primary results. First, from an economic standpoint, the trend of American economic health is an upward progression over time (regardless of market slumps, policy changes, etc.). Second, modified GRBF interpolation provides a more accurate estimate of raw quarterly data over the past 20 years, especially with small cluster sizes and standard deviations near three. Cluster sizes should be chosen small enough to estimate dynamic fluctuations in the market, yet large enough to avoid large swings of variance in the tails that are detrimental for predicting future growth. Linear models are very smooth and can provide rough estimates of economic health, but not with the accuracy of GRBFs.

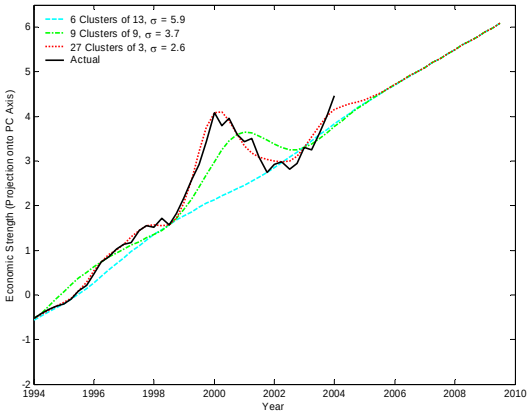

Fig. 5: Extrapolation of economic strength through 2010. GRBF σ values are the same as Fig. 1 (optimal), and thus the GRBF estimates converge quickly to the LS line.

6. References

[1] Merriam-Webster, *Webster's Ninth New Collegiate Dictionary*, 1991.

[2] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford Univ. Press, Oxford, UK, 1995.

[3] B. Strang, *Linear Algebra and Its Applications*, Third ed., Harcourt Brace Jovanovich, Fort Worth, TX 1988.

[4] M. Figueiredo, "On Gaussian Radial Basis Function Approximations: Interpretation, Extensions, and Learning Strategies," *15th International Conf. on Pattern Recognition*, Barcelona, Spain, Sep. 2000.

[5] M. Powell, "Radial Basis Functions for Multivariate Interpolation" in *Algorithms for Approximation*, J. Mathson and M. Cox, eds., Clarendon Press, Oxford, 1987, pp. 143-167.

[6] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Second ed., Prentice Hall, Upper Saddle River, NJ, 1999.

[7] http://www.forecasts.org.

[8] http://stats.bls.gov/bls.

[9] http://research.stlouisfed.org/fred2/categories.