Artificial Neural Network Approach as an Alternative to Ridge Regression on an Economic Problem

SENAY YOLACAN, BERNA YAZICI, EMBIYA AGAOGLU
Department of Statistics
Anadolu University
Eskisehir, 26470
TURKEY

Abstract: In this study ridge regression and artificial neural network (ANN) algorithm that does not require any assumption are applied on an economic data set with multicollinearity. The results are interpreted and compared.

Key-Words: Ridge regression, multicollinearity, artificial neural networks.

1 Introduction

The least squares estimators of the regression coefficients are the best linear unbiased estimators. That is, of all possible estimators that are both linear functions of the data and unbiased for the parameters being estimated, the least squares estimators have the smallest variance.

One of the assumptions of linear regression model is the independency of explanatory variables. If those variables are correlated with each other multicollinearity problem occurs. Multicollinearity is defined as the existence of nearly linear dependency among column vectors of the design matrix X in the linear model $Y = X\beta + u$. One way of explaining the structure of multicollinearity is to look over the eigenvalues and eigenvectors of the matrix $X'X$. One or more small eigenvalues show the existence of multicollinearity (Akdeniz and Erol, 2003).

In the presence of multicollinearity minimum variance may be unacceptably large. Severe multicolinearity makes the estimates so unstable that they are practically useless and the matrix $X'X$ becomes singular. To overcome this problem there are several biased estimation methods proposed by different authors including ridge regression (Hoerl and Kennard, 1970), Liu estimator (Liu, 1993).

In this study firstly ridge regression among biased estimation methods is taken into account, afterwards appropriate ANN algorithm is defined. The advantages and disadvantages of using that algorithm are mentioned. An application study is given about economical indicators. The last part of the study the conclusions and recommendations are given.

2 Ridge Regression

Ridge regression is a popular method for dealing with multicollinearity builds on the fact that a singular square matrix can be made nonsingular by adding a constant to the diagonal of the matrix. The model is as in Eq.1.

$$\hat{\beta} = (Z'Z + kI)^{-1} Z'Y \tag{1}$$

Where k is called shrinkage parameter, **I** is the identity matrix. **Z** is the standardized design matrix using following standardization:

$$\mathbf{Z} = \frac{X_{ij} - \overline{X}_j}{\sqrt{\sum \left(X_{ij} - \overline{X}_j\right)^2}} \tag{2}$$

k changes between 0 and 1 and the estimation of the parameters change with the value of k. When k=0 we have the ordinary least squares estimator of $\hat{\beta}$ of β. The goal is to choose k so as to minimize the mean squared error (MSE). However, MSE cannot be evaluated and therefore the choice of k is somewhat subjective. In order to overcome this problem an ANN algorithm is used to estimate regression parameter.

3 Radial Basis Functions Network

Optimal ridge regression parameters can be estimated for linear feed forward networks in supervised learning problems. With radial basis functions (RBF) because of their local nature ridge regression with multiple adjustable regularization parameters performs a kind of locally adaptive smoothing. The optimization is performed using either unbiased estimate of variance or generalized cross validation and needs an initial guess for the values of the parameters. Parameters in both the initial guesses and the final optimal estimates whose value is infinity correspond to pruned basis functions (Orr, 1997).

RBF networks have traditionally been associated with radial functions in a single-layer network as shown in Figure 1. The input layer is made of source nodes that connect the network to its environment. The second layer, the only hidden layer in the network, applies a non-linear transformation from the input space to the hidden space; one important point is the fact that the dimension of the hidden space is directly related to the capacity of the network to approximate a smooth input-output mapping (in most applications the hidden space is of high dimensionality). The output layer is linear, supplying the response of the network to the activation pattern applied to the input layer.

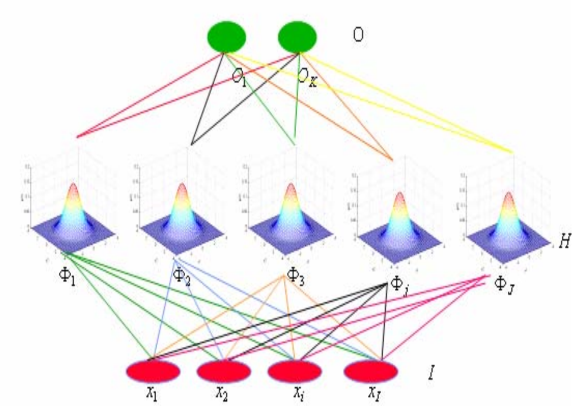

Figure 1. The traditional radial basis function network. Each of n components of the input vector x feeds forward to m basis functions whose outputs are linearly combined with weights $\{w_j\}_{j=1}^{m}$ into the network output f(x).

A RBF network, therefore, has a hidden layer of radial units, each actually modeling a Gaussian response surface. Since these functions are nonlinear, it is not actually necessary to have more than one hidden layer to model any shape of function: sufficient radial units will always be enough to model any function. (Haykin, 1999; Bishop, 1995)

In standard ridge regression for smoothing linear networks the cost function is made up of the usual sum-squared error plus a roughness penalty dependent on the squared length of the weight vector and a positive regularization parameter, k:

$$C(w,k) = \sum_{i=1}^{p} \left(f(x_i, w) - y_i \right)^2 + k \sum_{j=1}^{m} w_j^{\,2} \qquad (3)$$

where $f(x) = \sum_{j=1}^{m} w_j h_j(x)$ is the output of the network for input x and weight vector w and $\{(x_i, y_i)\}_{i=1}^{p}$ is the training set. A generalization of this to the case of multiple regularization parameters, $k = [k_1 k_2 \ldots k_m]^T$ (one for each basis function) is

$$C(w,k) = \sum_{i=1}^{p} \left(f(x_i, w) - y_i \right)^2 + \sum_{j=1}^{m} k_j w_j^{\,2} \qquad (4)$$

In the case of least squares applied to supervised learning with a linear model the function to be minimized is the cost function given in Eq. (4) used in ridge regression. Optimal weight vector is found as a result of this minimization as follows:

$$\hat{w} = \left(H^T H + kI_m \right)^{-1} H^T \hat{y} \qquad (5)$$

where H is design matrix, has the vectors $\{h_j\}_{j=1}^{m}$ as its columns, and has p rows, one for each pattern in the training set.

An important point is that if the training set has been used to estimate the regularization parameter(s), as in ridge regression, or to chose the basis functions, as in forward selection, the $\left(H^T H + kI_m \right)$=A is a stochastic variable and there is no longer a simple linear relationship between uncertainty in $\hat{w}$ and uncertainty in $\hat{y}$ (Orr, 1997).

Moreover RBF network has advantages: first, when we choose the ridge parameter k, we are able to choose a small number which is much more effective for smoothing the data. Further, by having the single parameter for adjusting the payoff between the two factors we can choose k appropriately so that we may vary any potential solution between the extremes of RBF and canonical correlation analysis depends on the needs of any particular data set (Gou and Fyfe, 2003).

4 Application

In this part of the study a data set on Turkey's GDP between 1988 and 2004 is analyzed both using ridge regression and ANN. The variables of the model are as follows:

Y: GDP deflator
X_1: Money Supply
X_2: Money Supply (previous term)
X_3: Inflation rate
X_4: Exchange rate
X_5: Unemployment Rate
X_6: Harmonized Indices of Consumer Prices (HCIP %)
X_7: Interest Rate
X_8: volume of currency issued
X_9: Gross Domestic Product (% increase)

Ridge regression parameters are estimated for different k values. The estimates are given in Table 1. The standard errors of parameter estimates are given in Table 2.

Table 1. Ridge estimates for different k values.

Parameter Estimates				
k	$\tilde{\beta}(k)_1$	$\tilde{\beta}(k)_2$	$\tilde{\beta}(k)_3$	$\tilde{\beta}(k)_4$
0	36.0440	1.4265	183.6293	55.5772
0.002	32.7395	1.0818	152.1507	50.8963
0.004	30.3096	0.8972	130.0135	47.5297
0.006	28.4300	0.8063	113.6534	44.9801
0.008	26.9217	0.7733	101.1102	42.9737
0.010	25.6772	0.7770	91.2168	41.3470
0.020	21.6463	1.0232	62.6306	36.2588
0.030	19.3836	1.3475	49.3235	33.4773
0.040	17.9204	1.6370	41.8730	31.6432
0.050	16.9036	1.8765	37.2211	30.2992
0.100	14.6205	2.5410	27.9927	26.4663
0.200	13.7463	2.9087	23.6772	22.9382
0.300	13.6555	3.0235	22.0564	20.9602
0.400	13.6226	3.0962	21.0008	19.5927
0.500	13.5537	3.1560	20.1692	18.5506
0.600	13.4441	3.2076	19.4619	17.7088
0.700	13.3032	3.2517	18.8377	17.0016
0.800	13.1408	3.2888	18.2755	16.3910
0.900	12.9644	3.3192	17.7622	15.8528
1.000	12.7798	3.3435	17.2894	15.3709

Parameter Estimates					
k	$\tilde{\beta}(k)_5$	$\tilde{\beta}(k)_6$	$\tilde{\beta}(k)_7$	$\tilde{\beta}(k)_8$	$\tilde{\beta}(k)_9$
0	122.8257	-74.8638	155.4209	141.3032	114.6210
0.002	108.6168	-47.9658	143.0157	133.2527	102.1725
0.004	98.1327	-29.2660	133.5448	126.9637	93.0473
0.006	89.9824	-15.6267	125.9370	121.8024	86.0014
0.008	83.3981	-5.3227	119.5989	117.4173	80.3483
0.010	77.9212	2.6724	114.1741	113.5963	75.6779
0.020	59.6520	24.5894	94.8608	99.4297	60.3714
0.030	48.6540	33.5118	82.1992	89.6311	51.4099
0.040	40.9302	37.6413	72.8346	82.1133	45.2439
0.050	35.0602	39.5994	65.4820	76.0430	40.6250
0.100	18.1645	39.9180	43.4616	56.8270	27.5353
0.200	4.7813	34.9715	25.5089	39.2898	16.9158
0.300	-0.8582	31.2533	17.8781	30.6399	11.9835
0.400	-3.8449	28.5931	13.8012	25.3098	9.0393
0.500	-5.6174	26.5814	11.3423	21.6206	7.0558
0.600	-6.7416	24.9863	9.7399	18.8792	5.6209
0.700	-7.4840	23.6755	8.6370	16.7429	4.5331
0.800	-7.9855	22.5686	7.8455	15.0211	3.6808
0.900	-8.3275	21.6139	7.2580	13.5980	2.9966
1.000	-8.5594	20.7768	6.8096	12.3988	2.4367

Table 2. The standard errors of parameter estimates

Standard Errors				
k	$s[\tilde{\beta}(k)_1]$	$s[\tilde{\beta}(k)_2]$	$s[\tilde{\beta}(k)_3]$	$s[\tilde{\beta}(k)_4]$
0	2.0535	1.7590	7.0552	2.5022
0.002	2.0007	1.7380	5.9322	2.4106
0.004	1.9597	1.7183	5.1411	2.3416
0.006	1.9255	1.6994	4.5547	2.2852
0.008	1.8955	1.6812	4.1031	2.2368
0.010	1.8684	1.6637	3.7449	2.1938
0.020	1.7575	1.5840	2.6866	2.0223
0.030	1.6686	1.5145	2.1621	1.8887
0.040	1.5927	1.4530	1.8433	1.7765
0.050	1.2720	1.3980	1.6253	1.6796
0.100	1.2792	1.1893	1.0858	1.3345
0.200	0.9923	0.9430	0.7093	0.9695
0.300	0.8222	0.7969	0.5498	0.7751
0.400	0.7070	0.6977	0.4596	0.6534
0.500	0.6231	0.6248	0.4010	0.5695
0.600	0.5587	0.5685	0.3596	0.5079
0.700	0.5078	0.5234	0.3285	0.4607
0.800	0.4664	0.4861	0.3041	0.4231
0.900	0.4320	0.4548	0.2844	0.3924
1.000	0.4029	0.4279	0.2681	0.3669

Standard Errors					
k	$s[\tilde{\beta}(k)_5]$	$s[\tilde{\beta}(k)_6]$	$s[\tilde{\beta}(k)_7]$	$s[\tilde{\beta}(k)_8]$	$s[\tilde{\beta}(k)_9]$
0	3.4315	6.3244	3.1761	2.4281	2.8743
0.002	3.0718	5.3511	2.9215	2.2962	2.5359
0.004	2.8255	4.6681	2.7416	2.2001	2.3039
0.006	2.6461	4.1636	2.6052	2.1246	2.1354
0.008	2.5089	3.7764	2.4961	2.0622	2.0076
0.010	2.4000	3.4700	2.4053	2.0085	1.9072
0.020	2.0633	2.5672	2.0920	1.8103	1.6090
0.030	1.8707	2.1176	1.8866	1.6705	1.4522
0.040	1.7338	1.8406	1.7314	1.5611	1.3487
0.050	1.6263	1.6478	1.6070	1.4715	1.5260
0.100	1.2849	1.1494	1.2191	1.1825	1.0498
0.200	0.9501	0.7752	0.8795	0.9095	0.8493
0.300	0.7707	0.6079	0.7193	0.7669	0.7399
0.400	0.6551	0.51	0.6225	0.6739	0.6649
0.500	0.5734	0.4449	0.5556	0.6064	0.6082
0.600	0.5121	0.3980	0.5056	0.5540	0.5628
0.700	0.4642	0.3624	0.4662	0.5117	0.5252
0.800	0.4257	0.3342	0.4339	0.4766	0.4933
0.900	0.3941	0.3114	0.4068	0.4467	0.4656
1.000	0.3674	0.2924	0.3837	0.4208	0.4414

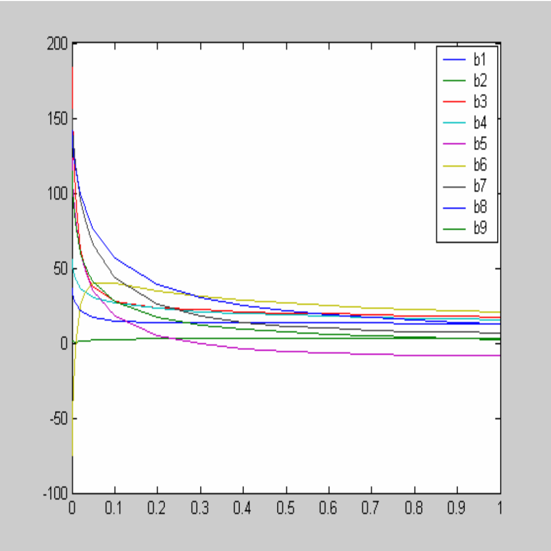

Figure 2. Ridge trace for Turkey's GDP data

k is chosen taken account the ridge trace given in Figure 2. k is defined as 0.3. Ridge estimates and related mean absolute error (MAE) value are given in Table 4.

Table 3. The weights of RBFN for Turkey's GDP data

Network Weights

	2.1	2.2	2.3	2.4	2.5
Thresh	1.0000	1.0000	1.0000	1.0000	1.0000
1.1	1.0000	0.3457	0.7064	0.6777	0.7019
1.2	0.1527	0.2798	0.6883	1.0000	0.3192
1.3	1.0000	0.5855	0.6309	0.7876	0.5669
1.4	1.0000	0.3552	0.3331	0.4930	0.2596
1.5	0.3846	0.3385	0.2154	0.0000	0.4769
1.6	1.0000	0.6091	0.5529	0.7775	0.5356
1.7	0.7056	0.5307	1.0000	1.0000	0.5215
1.8	0.4167	0.5833	0.4167	0.0000	0.8333
1.9	0.0000	0.8507	0.0373	0.9701	0.5000

Network weights

	2.6	2.7	2.8	2.9	3.1
Thresh	1.0000	1.0000	1.0000	1.0000	0.0036
1.1	0.3299	0.1867	0.0000	0.0025	
1.2	0.1466	0.6260	0.1450	0.0000	
1.3	0.6370	0.5267	0.2101	0.0000	
1.4	0.3276	0.1004	0.0992	0.0000	
1.5	0.2462	0.4615	0.9076	1.0000	
1.6	0.5648	0.5032	0.3531	0.0000	
1.7	0.5307	0.4617	0.7056	-0.000	
1.8	0.6667	0.7500	0.1667	1.0000	
1.9	0.4925	1.0000	0.9776	0.7836	
2.1					0.0749
2.2					0.0113
2.3					-0.046
2.4					0.0354
2.5					0.0677
2.6					-0.0517
2.7					-0.0017
2.8					-0.0086
2.9					-0.0711

RBFN results of Turkey's GDP data are calculated and weights are given in Table 3 and estimates of Y and MAE are given in Table 4.

Table 4. Ridge estimates of Y for k=0.3 and RBF estimates of Y

Y	$\hat{Y}$ (Ridge)	$\hat{Y}$ (RBFN)
69.7	71.43	59.69
75.5	69.23	75.50
57.6	60.41	57.60
59.2	65.29	59.20
63.5	67.37	63.50
67.4	64.31	59.19
107.3	100.57	107.30
81.9	76.97	79.59
78.1	70.99	79.33
81.2	79.58	81.20
75.3	73.52	75.80
55.8	68.22	55.80
50.9	52.67	57.95
55.3	55.76	55.97
44.4	35.82	44.40
22.5	20.85	31.36
11.9	24.48	11.90

MAE(Ridge)=4.91
MAE(RBF)= 2.28

As can be seen in Table 4 the MAE value for RBFN is quite small than the MAE value for ridge regression.

5 Conclusions and Recommendations

Multicollinearity is a very common problem in multiple linear regression. Ridge regression is one of the biased methods that are used in order to overcome that problem. But choosing k in ridge regression is somewhat subjective. Hence in this study one of the algorithms of ANN, RBFN is used.

As a criterion of comparison MAE for each model is calculated. It is seen that the RBFN result is better than ridge regression result.

Another advantage of using RBFN is the researcher does not need to define a k value to make the best estimation.

For the further studies also Liu estimators can be used and all results can be compared using MAE values.

References

[1] Akdeniz F. and Erol H, Mean Squared Error Matrix Comparisons of Some Biased Estimators in Linear Regression, *Communication in Statistics, Theory and Methods*, Vol. 32, No.12, 2003, pp.2389-2413.

[2] Hoerl, A. E. and Kennard, R., Ridge Regression: Biased Estimation for Non-orthogonal Problems, *Technometrics*, 12, 1970, pp.55-67.

[3] Gou Z. and Fyfe C., A Canonical Correlation Neural Network For Multicollinearity and Functional Data, *Neural Networks*, 17, 2004, pp. 285-293.

[4] Liu, Ke Jian, 1993, A new class of biased estimate in linear regression. *Communications in Statistics- Theory and Methods*, 22(2):393-402.

[5] Orr, M., MATLAB Routines for Subset Selection and Ridge Regression in Linear Neural Networks, Scotland, 1997, UK.

[6] Haykin, S. (1999), Neural Networks: A Comprehensive Foundation, Prentice Hall Pub. Co.

[7] Orr, M., 1995, Local smoothing of radial basis function networks, *Symposium on Artificial Neural Networks*, Hsinchu, Taiwan.

Appendix

Turkey's GDP data

Years	**Y**	X_1	X_2	X_3	X_4
1988	69.7	65.3	0	75.4	66
1989	75.5	95.7	65.3	64.3	49
1990	57.6	48.2	95.7	60.4	23
1991	59.2	61.4	48.2	71.1	60.1
1992	63.5	62.86	61.4	66.1	64.6
1993	67.4	48.8	62.86	71.1	60.5
1994	107.3	123.18	48.8	106.3	169.9
1995	81.9	99.36	123.18	88	53.2
1996	78.1	132.76	99.36	80.4	77.8
1997	81.2	93.47	132.76	85.7	87.1
1998	75.3	101.87	93.47	90.7	71.6
1999	55.8	96.11	101.87	70.5	61
2000	50.9	42.45	96.11	39.1	48.4
2001	55.3	48.04	42.45	68.5	96.4
2002	44.4	30.99	48.04	29.7	22.8
2003	22.5	33.67	30.99	18.4	-0.6
2004	11.9	31.22	33.67	9.32	6.6

Years	X_5	X_6	X_7	X_8	X_9
1988	8	73.7	54	3.5	2.1
1989	9	63.3	54	3.6	1.2
1990	8.9	60.3	50.75	3.5	7.9
1991	7.5	66	54.50	3.4	1.1
1992	8.1	70.1	54.50	3.3	5.9
1993	7.5	66.1	54.50	3.2	8
1994	8.4	106.3	64	3.1	-5.5
1995	7.2	93.6	57	2.9	7.2
1996	6.3	80.4	57	2.6	7
1997	5.9	85.7	80	2.6	7.5
1998	6.4	84.6	80	2.5	3.1
1999	7.3	64.9	80	3.1	-5
2000	8.3	54.9	70	3	7.4
2001	8.6	54.4	70	3	-7.6
2002	11.8	46.4	64	2.8	7.6
2003	12.3	25.3	48	3	5.8
2004	12.4	13.7	25.66	3.8	5