Prctica 1: Prediction with Back-Propagation and Linear Regression

Pulak Roy

Universitat Rovira i Virgili,
Av. Països Catalans 26,
E-43007 Tarragona, Catalonia, Spain
e-mail {pulakroy.pulak}@estudiants.urv.cat

1 Introduction

The goal of any supervised learning algorithm is to find a function that best maps a set of inputs to its correct output.Back-propagation is a common method for training a neural network.The goal and motivation for developing the backpropagation algorithm was to find a way to train a multi-layered neural network such that it can learn the appropriate internal representations to allow it learn any arbitrary mapping of input to output.Alongside Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable x is associated with a value of the dependent y.

2 Objective

Objective: Prediction of the power of the turbine of a hydro electrical plant, using the following algorithms:
1. Back-Propagation (BP)
2. Multiple Linear Regression (MLR)

3 Back-Propagation (BP)

Backpropagation is a supervised learning algorithm and is mainly used by Multi-Layer-Perceptrons to change the weights connected to the net's hidden neuron layer(s).The back-propagation algorithm uses a computed output error to change the weight values in backward direction. To get this net error, a forward-propagation phase must have been done before.While propagating in forward direction, the neurons are being activated using

the sigmoid activation function.The formula of sigmoid activation is: f(x) = $1/(1+e^{-x})$.The algorithm works as follows:

```
Randomly initialize weights and thresholds
for epoch 1 to total number of epoch
    for pattern 1 to total number of pattern
        choose random pattern.
        feed-forward
        back-propagation of error
        calculation of weights and thresholds updates
        update weights and thresholds
    end
end
```

Implementation decisions:

In this current implementation we can choose number of hidden layers and number of neurons per layer. For our data it is better to use two hidden layers. On the other hand to choose I have chosen thirteen nodes in first hidden layer and 5 at second hidden layer.Because this network architecture gives best prediction with error rate 0.57%
Along side number of epoch and learning rate plays an important role in the prediction. In this case I preferred epoch = 10000000 and Learning rate = 0.1

Language: Java

Tool: Eclipse JavaEE mars

Result:

real	prediction	error	total error
3698.29	3705.107	6.82	0.57%
1935.13	1947.488	12.36	
4509.67	4509.252	0.42	
5708.99	5701.963	7.03	
2024.05	2030.317	6.27	
3981.74	3990.353	8.61	
6171.22	6180.054	8.83	
4427.36	4420.862	6.5	
2117.9	2098.754	19.15	
2692.87	2689.051	3.82	
3237.11	3260.684	23.57	
3558.33	3590.829	32.5	
2327.54	2313.536	14	
4063.35	4089.213	25.86	
3838.09	3852.186	14.1	
6098.9	6114.126	15.23	
3432.48	3454.15	21.67	
2042.36	2038.665	3.69	
2841.89	2841.965	0.08	
4572.62	4540.121	32.5	
4384.97	4377.646	7.32	

4384.97	4377.646	7.32
6297.76	6345.418	47.66
2589.43	2585.458	3.97
2569.39	2563.981	5.41
4192.01	4187.853	4.16
3692.94	3702.912	9.97
2081.66	2090.112	8.45
4591.21	4731.523	140.31
2590.7	2614.435	23.74
4102.62	4126.668	24.05
2865.92	2877.225	11.3
3090.96	3142.397	51.44
4096.06	4023.985	72.08
1938.47	1941.835	3.37
4801.54	4758.522	43.02
3299.93	3317.742	17.81
3670.99	3693.869	22.88
3743.56	3741.261	2.3
3128.97	3147.344	18.37
1776.96	1839.691	62.73
5153.46	5160.462	7
5016.03	4947.968	68.06

5016.03	4947.968	68.06
3618.87	3636.821	17.95
5632.13	5647.897	15.77
5283.22	5282.003	1.22
2337.68	2311.808	25.87
6451.24	6505.081	53.84
4806.83	4814.235	7.4
4450.58	4478.64	28.06
2893.53	2905.953	12.42

Evaluation of the predictions(Error and Plot):
Error: 0.57%

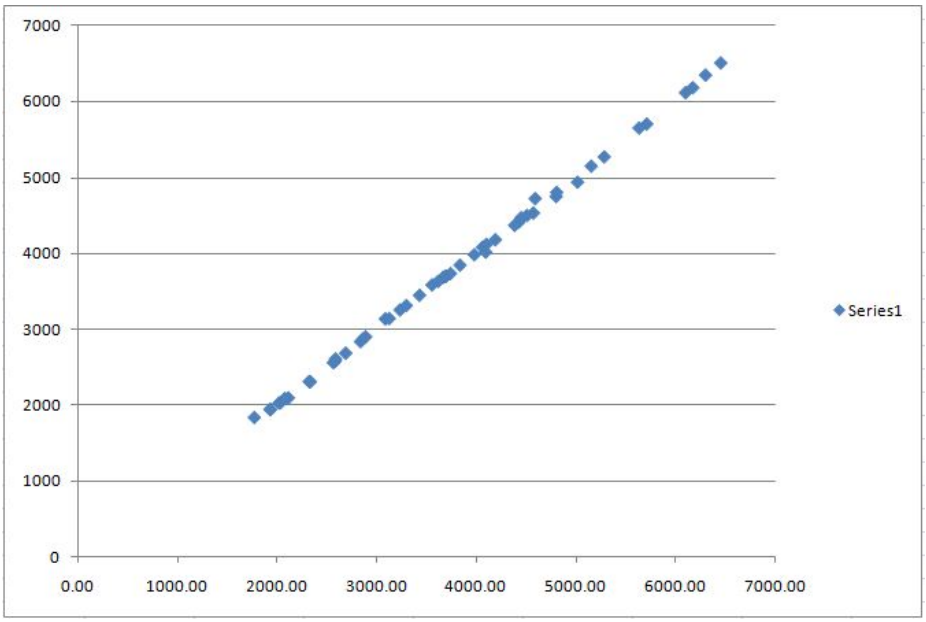

Figure1: Back-propagation prediction plot

4 Multiple Linear Regression

Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable x is associated with a value of the dependent variable y. Formally, the model for multiple linear regression, given n observations, is $y = b_0 + b_1x_1 + b_2x_2+.....+b_nx_n$ for i = 1,2.....n

Implementation decisions:

Language: Using software

Tool: Microsoft office Excel 2007

Result:

Real	Prediction	Error	Total Error
3698.29	3810.2275	111.9375	4.97%
1935.13	1552.841207	382.288793	
4509.67	4649.645717	139.975717	
5708.99	5585.837291	123.152709	
2024.05	1739.008274	285.041726	
3981.74	3993.76927	12.02927	
6171.22	5913.775515	257.444485	
4427.36	4540.332976	112.972976	
2117.9	2037.832114	80.067886	
2692.87	2640.237134	52.632866	
3237.11	3132.144469	104.965531	
3558.33	3912.385969	354.055969	
2327.54	2183.41351	144.12649	
4063.35	4132.439131	69.089131	
3838.09	3829.800158	8.289842	
6098.9	6083.748464	15.151536	
3432.48	3693.760487	261.280487	
2042.36	2288.678167	246.318167	
2841.89	3015.686084	173.796084	
4572.62	4973.168076	400.548076	
4384.97	4485.676605	100.706605	

4102.62	4058.612081	44.007919
2865.92	2729.977216	135.942784
3090.96	3562.24979	471.28979
4096.06	4505.587007	409.527007
1938.47	1960.739944	22.269944
4801.54	5082.480817	280.940817
3299.93	3475.135005	175.205005
3670.99	3703.774163	32.784163
3743.56	3913.813648	170.253648
3128.97	3142.330909	13.360909
1776.96	1578.14535	198.81465
5153.46	5591.841129	438.381129
5016.03	5288.229172	272.199172
3618.87	3524.925503	93.944497
5632.13	5670.823766	38.693766
5283.22	5312.555438	29.335438
2337.68	2099.404912	238.275088
6451.24	6302.373946	148.866054
4806.83	5039.273585	232.443585
4450.58	4401.429853	49.150147
2893.53	2873.914928	19.615072

Evaluation of the predictions(Error and Plot): Error: 4.97%

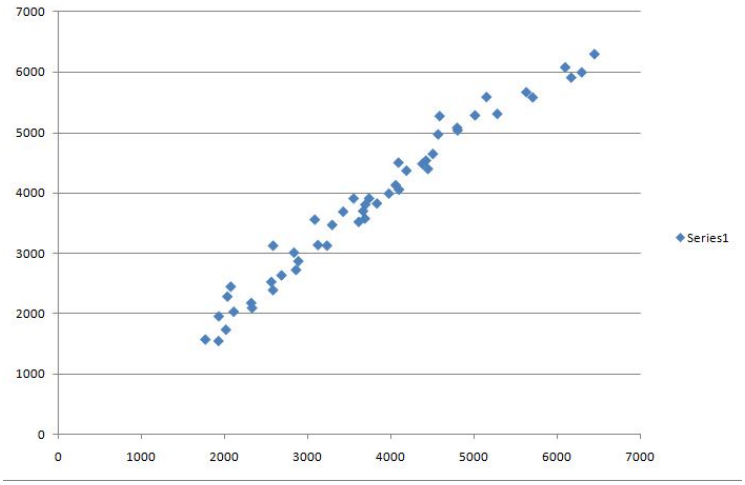

Figure2: Multi linear regression prediction plot

5 Conclusion

Our experiment suggests that ANN-BP technology holds great promise for the prediction of turbine data set. This research has provided some evidence that is very promising from ANN-BP rather than multi-linear regression.

References

1. "Backpropagation", [Online]. Available: , `https://en.wikipedia.org/wiki/Backpropagation` [Accesed: Jun. 5, 2016]

2. "Multiple Linear Regression ", [Online]. Available: , `http://www.stat.yale.edu/Courses/1997-98/101/linmult.htm` [Accesed: Jun. 5, 2016]

3. "Comparison of a Back Propagation Artificial Neural Network Model With a Linear Regression Model for Personnel Selection", [Online]. Available: , `http://www.stat.yale.edu/Courses/1997-98/101/linmult.htm` [Accesed: Jun. 5, 2016]