4.3 Price prediction model based on BP neural network regression.

4.3.1 The Foundation of Model

Back Propagation Neural Network (BP Neural Network) is a common artificial neural network with strong learning ability and adaptability. Its basic principle is to iteratively calculate the connection weights between neurons by using input and output data, in order to achieve approximation and prediction of nonlinear patterns.

The basic structure of BP neural network consists of input layer, hidden layer and output layer. The input layer accepts external input data, the hidden layer is an intermediate layer composed of multiple neurons, and the output layer outputs the final result. During the training process, the network uses forward propagation to pass the input data from the input layer to the output layer, and then calculates the output error using the back propagation algorithm, and adjusts the connection weights between neurons according to the error. This process is repeated many times until the network error reaches the preset threshold or the maximum number of iterations is reached.

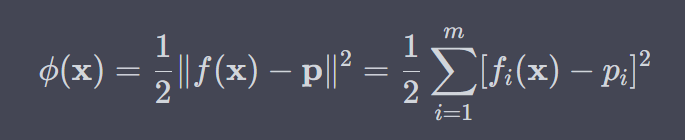
（插入神经网络输入输出隐藏层的圈圈图）

4.3.2 Algorithms used in the model

We used the Levenberg-Marquardt (LM) algorithm to train the neural network.

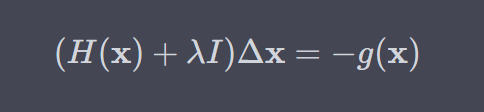
LM algorithm (Levenberg-Marquardt algorithm) is a commonly used nonlinear least squares optimization algorithm, mainly used in the training process of neural networks.

The LM algorithm is based on the following objective function:



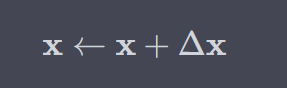
where denotes the Euclidean norm, m is the number of residuals, and  is the predicted value of the ith input of function f at x.

In the LM algorithm, we first use the Gauss-Newton method to obtain an initial value for x. Then, we calculate the gradient g(x) and the Hessian matrix H(x) at x. Next, we use the following formula to calculate the update for x:



where I is the identity matrix and is a parameter that controls the step size. If  is small, the update is close to the Gauss-Newton method, and if  is large, the update  is close to the gradient descent method. Therefore, the value of  should be balanced between the two.

Finally, update the value of x:



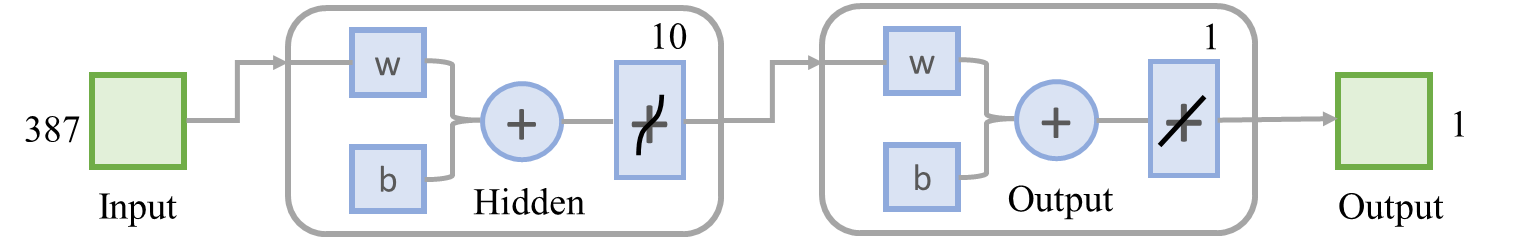
This process iterates until the error reaches the predetermined threshold or the maximum number of iterations is reached.

4.3.3 Training Process

1.我们使用用因子分析和相关性分析得到的降维后的具有强相关性的数据作为输入矩阵矩阵的列变量，不同的品牌及变种的帆船种数作为矩阵的行变量，构成输入矩阵X；以不同的品牌及变种的帆船作为矩阵的行变量，给定的价格作为矩阵的列变量，构成输入矩阵Y。

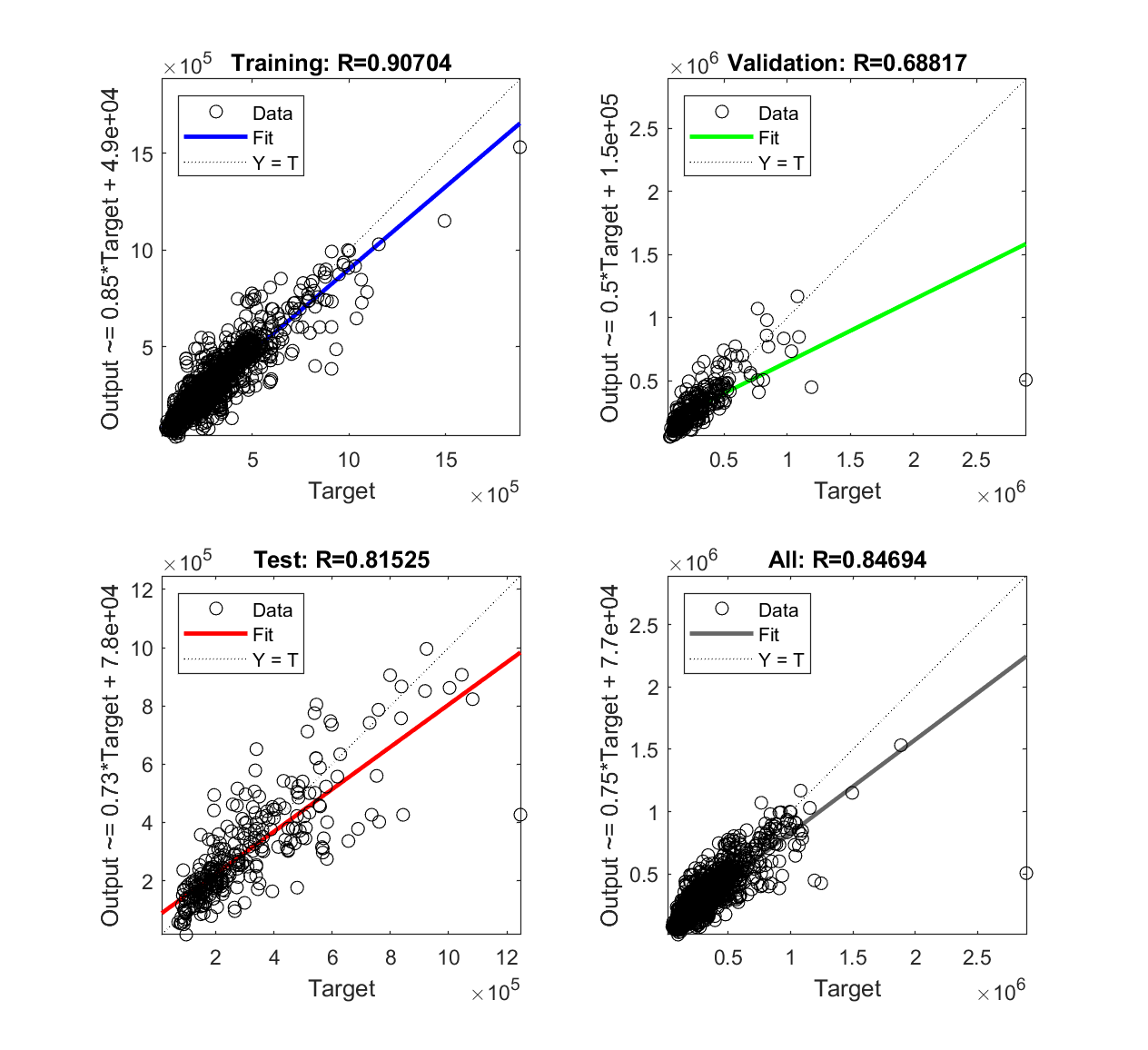
2.将补充收集的32984条数据进行划分，70%作为训练集，15%作为Validation Data，15%作为Test Data，构建含有1层输入层，10层隐藏层和1层输出层的神经网络。

3.我们使用MATLAB软件来编程实现上述操作。Training will continue until a pre-defined stopping condition is met (Epoch greater than or equal to 1000 or the validation error of the iteration continues to increase).



4.3.4 Analysis of the Result

Linear regression plots were generated to display the corresponding (target) network predictions (output) for the training set, validation set, and test set. If the fit is perfect, the data should fall along a 45 degree line where the network output is equal to the response. It can be seen from the figure below that the dataset fits very well. To obtain more accurate results, the dataset can be trained again with different initial weights and biases for the network, and the network can be improved after retraining.



The results of the mean square error (MSE) and R-value indicate that the network has high accuracy in predicting the prices of used sailboats, reflecting the efficiency of the network model.

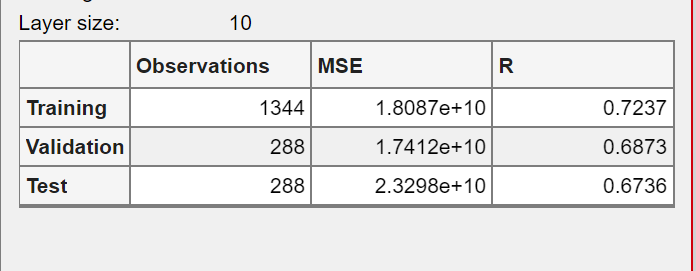
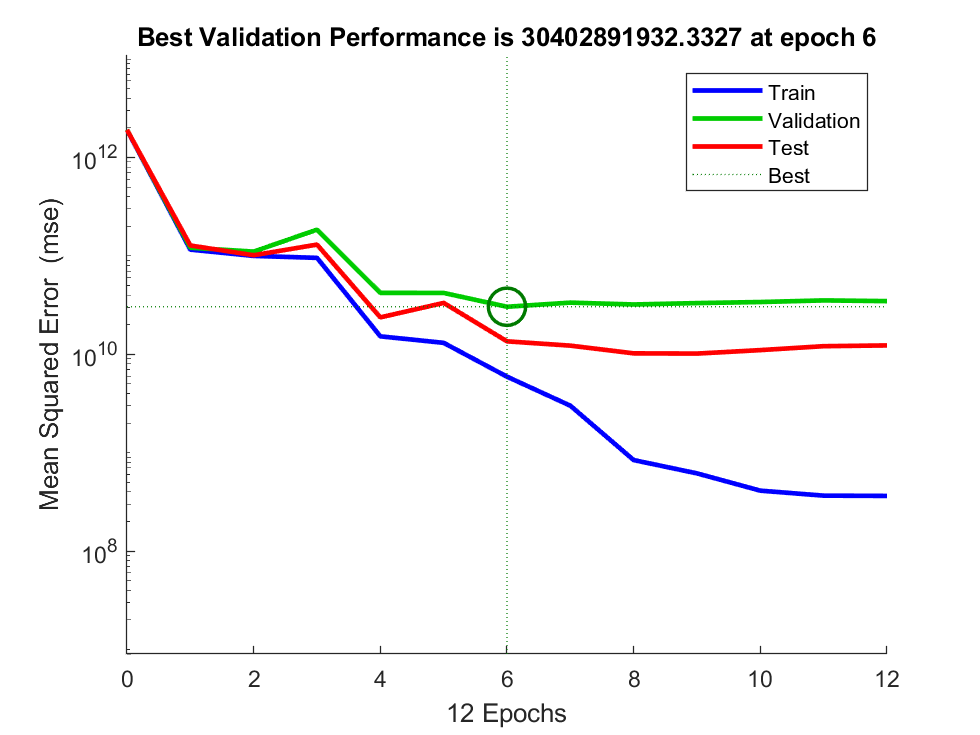


Figure : Parameter Index Chart

（画三线表，Observation数据改一下，从上到下23088，4948,4948）

This training used conditional stop training, which stopped at 6 epochs.



Optimal state chart

View the error histogram for additional verification of network performance.

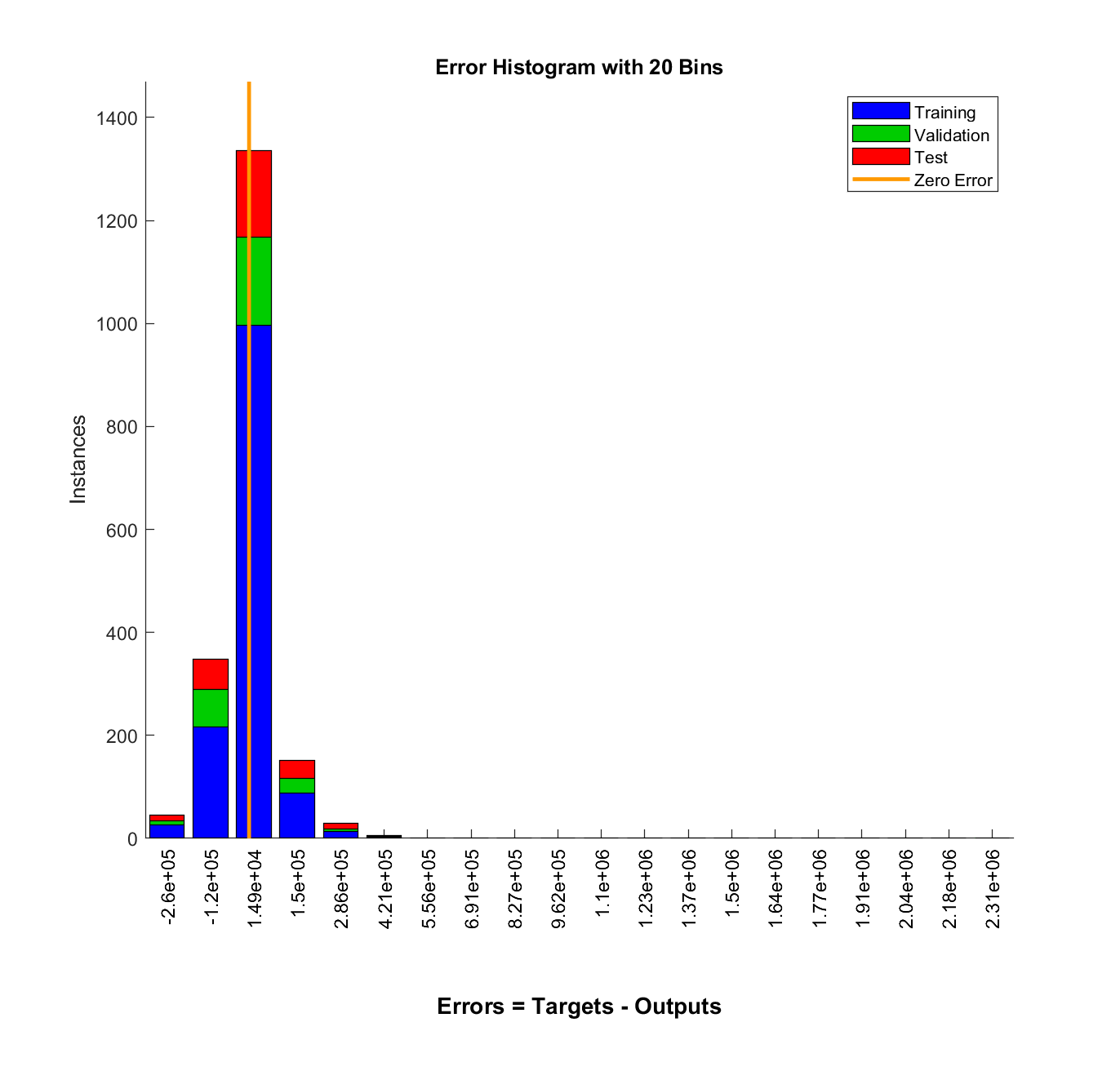


Figure : Error Histogram

The blue bars indicate training data, the green bars indicate validation data, and the red bars indicate test data. The histogram can indicate outliers, which are data points that fit significantly worse than the majority of the data. The outliers are examined to determine if the data are poor within the error range or if any of these data points differ from the rest of the data set, and the data points with valid outliers are interpolated using the network.