Topic: Predict Daily Performance of Nasdaq Composite Index by Relative Stocks

Member name: Ke Gao

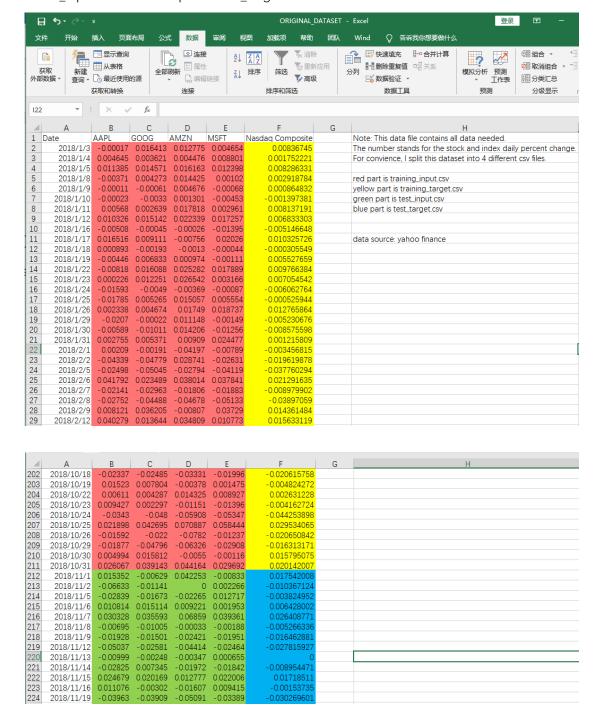
Date: 2018/12/23

1. Our goal

Our goal is to predict Nasdaq Composite Index (IXIC) daily performance by four different stocks daily performance. These four stocks are AAPL (Apple.Inc), GOOG (Google), AMZN (Amazon) and MSFT (Microsoft).

2. Our dataset:

Our data source is yahoo finance. All data needed is in ORIGINAL_DATASET.xls. The number stands for stocks and index daily percent change. For convenience, I split this dataset into 4 different csv files. Red part is training_input.csv, yellow part is training_target.csv, green part is test_input.csv and blue part is test_target.csv



I select data from 2018/01/03 to 2018/10/31 as training set (210 days), and data from 2018/11/01 to 2018/12/21 as test set (35 days).

Explanation of data									
color	Red	Yellow	Green	Blue					
filename	training_input.csv	training_target.csv	test_input.csv	test_target.csv					
Counts	840	210	140	35					

3.Our neural network

We use 3 different topologies of neural networks.

Type 1: (4,4,1)

(4 neurons in input layer, 4 neurons in hidden layer, 1 neuron in output layer)

Type 2: (4,4,3,1)

(4 neurons in input layer, 4 neurons in hidden layer I, 3 neurons in hidden layer II, 1 neuron in output layer)

Type 3: (4,4,3,2,1)

(4 neurons in input layer, 4 neurons in hidden layer I, 3 neurons in hidden layer II, 2 neurons in hidden layer III, 1 neuron in output layer)

Our activation function is bipolar sigmoid function:

$$f(x) = \frac{2}{1 + e^{-x}} - 1$$
$$f'(x) = 0.5 * (1 + f(x)) * (1 - f(x))$$

Our error function is

$$Error = 0.5 * (output - target)^2$$

4. Our programming

Our program contains class Neuron, Layer, NeuralNetwork and Matrix.

a).vector<int> topology stands for the topology of neural network. If topology is (4,4,3,2,1), which means neural network have 4 neurons in input layer, 4 neurons in hidden layer I, 3 neurons in hidden layer II, 2 neurons in hidden layer III, 1 neuron in output layer.

b). vector_of_inputs and vector_of_targets stand for the input and target data. They read data from input and target.csv which mentioned above.

```
vector<vector<double>> vector_of_inputs;
vector<vector<double>> vector_of_targets;
//read training file from csv
vector_of_inputs = ReadCSV("training_input.csv");
vector_of_targets = ReadCSV("training_target.csv");
```

c). default parameters are initialized here.

d). Training part.

If cycle > max_cycle or average_error < tolerance_error, the while loop breaks.

```
//training
while (cycle< max_cycle && average_error > tolerance_error) {
    average_error = 0;
    for (int i = 0; i < vector_of_inputs.size(); i++) {
        nn.SetInputValue(vector_of_inputs[i]);
        nn.ForwardPropagation();
        nn.BackPropagation();
        nn.BackPropagation();
        error = nn.GetTotalError();
        average_error += error / vector_of_inputs.size();
    }
    cycle += 1;
    cout << "cycle = " << cycle << " average error = " << average_error << endl;
}</pre>
```

e). Test part.

First, we read test_input.csv and test_target.csv.

Then we only do forward propagation to calculate predict_value and derived error = (predict_value -target_value)

In the end, we output result to txt file.

```
//testing
vector of inputs = ReadCSV("test_input.csv");
vector_of_targets = ReadCSV("test_target.csv");
cout << "Predict= \tTarget= \tPredict - Target = "<< endl;
output_file << "Predict= \tTarget= \tPredict - Target = " << endl;
for (int i = 0; i < vector_of_inputs.size(); i++) {
    nn.SetInputValue(vector_of_inputs[i]);
    nn.SetTargetValue(vector_of_targets[i]);
    nn.ForwardPropagation();
    cout.width(8);
    cout.setf(ios::showpoint);
    cout << setprecision(6);
    cout << nn.GetPredictValue(0) << " \t" << nn.GetTargetValue(0) << " \t" << nn.GetTotalDerivedError() << endl;
    output_file << nn.GetPredictValue(0) << " \t" << nn.GetTargetValue(0) << " \t" << nn.GetTotalDerivedError() << endl;
}
output_file.close();
return 0;
}
```

f). ForwardPropagation()

We take bias into account in forward propagation.

g). SetErrors()

We calculate errors given the result value of forward propagation.

```
oid NeuralNetwork::SetErrors() {
if (target value.size() == 0){
  cout << "No target value for neural network!" << endl;
if \ (target\_value.size() \ != \ layers[layers.size()-1].GetLayerSize()) \ \{\\
  cout << "Target size do not match with output layer size!" << endl;</pre>
total_error = 0.0;
total_derived_error = 0.0;
int output_layer_index = layers.size() - 1;
vector < Neuron > \ output\_neurons = layers[output\_layer\_index]. GetNeuronLayer();
vector<double> result_errors;
vector<double> result_derived_errors;
for (int i = 0; i < target_value.size(); i++) {
  double temp_error = 0.5*(output_neurons[i].GetActivatedValue() - target_value[i])*(output_neurons[i].GetActivatedValue() - target_value[i]);
  double temp_derived_error = (output_neurons[i].GetActivatedValue() - target_value[i]);
  result_errors.push_back(temp_error);
  result_derived_errors.push_back(temp_derived_error);
  total_error += temp_error;
  total derived error += temp derived error;
errors = result errors;
derived errors = result derived errors:
total_error = total_error / topology[topology.size() - 1];
total_derived_error = total_derived_error / topology[topology.size() - 1];
historical_total_errors.push_back(total_error);
```

h). BackPropagation()

We take bias into account in back propagation.

```
int output_layer_index = layers.size() - 1;
int last_hidden_layer_index = output_layer_index - 1;
Matrix derived_output_to_hidden(layers[output_layer_index].MatrixifyDerivedValues()); //create derived values matrix
Matrix error_output_to_hidden(1, errors.size(), derived_errors);
Matrix gradient_output_to_hidden(derived_output_to_hidden.PointwiseProduct(error_output_to_hidden)); // gradient matrix
Matrix delta_output_to_hidden((layers[last_hidden_layer_index].MatrixifyActivatedValuesPlusBias().TransposeMatrix())*gradient_output_to_hidden);
weight_matrices[last_hidden_layer_index] = weight_matrices[last_hidden_layer_index] - delta_output_to_hidden*learning_rate;
Matrix gradient hidden(gradient output to hidden);
for (int i = last_hidden_layer_index; i > 0; i--) {
  {\bf Matrix}\ derived\_hidden(layers[i]. Matrix if y Derived Values());
  Matrix weight matrix b;
  weight\_matrix\_b = weight\_matrices[i].EraseLastRow();
  gradient_hidden = derived_hidden.PointwiseProduct(gradient_hidden*(weight_matrix_b.TransposeMatrix()));
  Matrix neuron_value_a(layers[i - 1].MatrixifyActivatedValuesPlusBias());
    neuron_value_a = layers[0].MatrixifyValues();
  Matrix\ delta\_hidden(neuron\_value\_a. Transpose Matrix()*gradient\_hidden);
  weight\_matrices[i-1] = weight\_matrices[i-1] - delta\_hidden*learning\_rate;
```

5.Our Output:

We take Type 3: (4,4,3,2,1) neuron network as a demo to show how does the program works. (This result is different from the final result I submitted, because initial weight matrices are randomly generated, my final result is in folder *Prediction&WeightMatrices Results*)

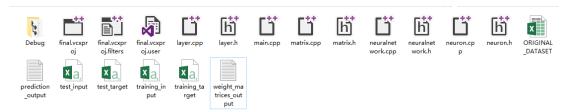
PS: Weight matrices has one more row due to bias value.

```
■ C:\Users\HP\Desktop\final with bias\Debug\final.exe
                                                                           Before training
·***********BEGIN*********
NeuronValue =
0 0
eightMatrix =
           0.69782
0.323501
0.843951
0.529642
                                  0. 260041
0. 63615
0. 447886
0. 803365
 ******
NeuronActivatedValue=
0 0 0
NeuronActivatedValue=
NeuronActivatedValue=
|
| VeightMatrix (nrow + 1 because of bias value ) =
). 0266079
). 518386
 *********
Layer = 4
NeuronActivatedValue=
Total Error = 0
*******************
```

```
■ C:\Users\HP\Desktop\final with bias\Debug\final.exe
After trainining
************BEGIN********
Layer = 0
NeuronValue =
                0.0391427
                                 0.0441643
                                                  0.0296924
0.0260666
WeightMatrix =
                                 0.896998
0.905356
0.632145
0.536509
0. 792998
1. 02485
                0. 714343
0. 343508
0. 859722
                                                  0. 287421
0. 669327
                                                  0. 473843
0. 848529
0.281717
                0.55674
0.498622
***********
Layer = 1
NeuronActivatedValue=
                0.0432561
0.0439881
                                 0.0512891
                                                  0.0398854
0. 373355
0. 32835
0. 159847
0. 669706
                                 0. 0718626
0. 789772
1. 01383
                0. 935865
0. 822235
0. 868823
0.630842
                0.0310458
                                 0.243647
Layer = 2
NeuronActivatedValue=
                                 0.183496
0.334768
                0.087414
******
Layer = 3
NeuronActivatedValue=
                0.492389
0. 175128
WeightMatrix (nrow + 1 because of bias value ) =
0.531258
0.557531
-0.313456
*******************************
Layer = 4
NeuronActivatedValue=
0.0270454
Total Error = 2.38286e-05
********************************
Predict=
0.00870872
-0.0119144
                                 Predict - Target = -0.00883329 -0.00154728
                Target=
0.0175420
-0.0103671
```

n nnaaa495

Our program will generate two txt files, prediction_output.txt, which contains prediction and error calculation of testset.



prediction_output - 记事本

文件(F) 编辑(E) 格式(O)	查看(V) 帮助(H)	
Predict=	Target=	Predict - Target =
0.00870872	0.017542	-0.00883329
-0.0119144	-0.0103671	-0.00154728
-0.00765358	-0.00382495	-0.00382863
0.00928079	0.006428	0.00285279
0.0322549	0.0264088	0.0058461
-0.00147189	-0.00526634	0.00379444
-0.0116476	-0.0164629	0.00481524
-0.0237878	-0.0278159	0.00402816
-0.000425519	0 -0.000425	5519
-0.00756597	-0.00895447	0.0013885
0.0168966	0.0171851	-0.000288527
0.00295761	-0.00153735	0.00449496
-0.0274256	-0.0302696	0.00284404
-0.0119187	-0.017025	0.0051063
0.00920047	0.00918104	1.94304e-05
-0.00671224	-0.00477178	-0.00194047
0.0235665	0.0205895	0.00297696
0.00221823	0.000120039	0.00209819
0.0330875	0.0294929	0.00359453
-0.00052222	-0.00253851	0.00201629
0.00520117	0.00790036	-0.0026992
0.0201717	0.015138	0.00503365
-0.0315174	-0.0380406	0.00652326
0.00784895	0.00416706	0.00368189
-0.0243287	-0.0304677	0.006139
0.00994567	0.0073566	0.00258907
0.00567267	0.001611	0.00406167
0.00794866	0.00945415	-0.00150549
0.00411747	-0.00394178	0.00805925
-0 0193735	-0 0225831	0 00320956

And Weight_matrices_output.txt contains the topology of neural network and weight matrices.

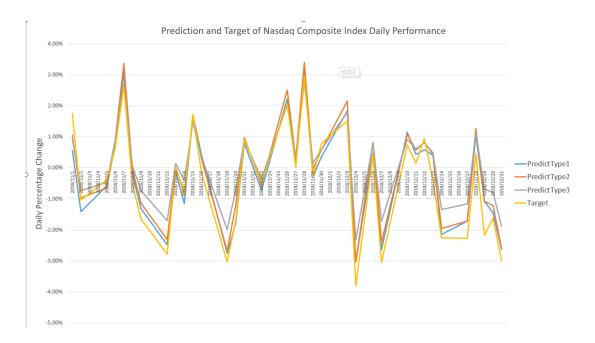
```
🥘 weight_matrices_output - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)
**********BEGIN*******
Layer = 0
NeuronValue =
           0.0260666
WeightMatrix =
0.792998 0.714343 0.896998 0.287421
0.281717  0.859722  0.632145  0.473843
0.498622 0.55674 0.536509 0.848529
*********
Layer = 1
NeuronActivatedValue=
0.0439881
             0.0432561 0.0512891
                                           0.0398854
WeightMatrix (nrow + 1 because of bias value ) =
0.373355 0.612044 0.988833
0.32835 0.935865 0.0718626
0.159847  0.822235  0.789772
0.669706 0.868823 1.01383
0.630842 0.0310458
                 0.243647
*********
Layer = 2
NeuronActivatedValue=
0.334768 0.087414 0.183496
```

6.Result Analyze

This is the excel file merged 3 txt files (prediction_output_type1.txt, prediction_output_type2.txt, prediction_output_type3.txt)

B 6 € € €							Result a	analyze -	EXCEI					
文件 开始	插入 页面布	词 公式 数据	审阅 视	图 加載项	豆 帮助 团队	Wind O	告诉我你想要做	a 什么						
	Jan 2 3 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	2.30 3031				y							m (←)	∑ 自动求利
00 34 WI	等线	+ 11 + A	_ =	= = 8	・ き 自动接行	常规	*	F		常规	差	^ ==	· 활 🖺	
温 复制・	B I U -	- B - A - 0	én - ≡	= = =	重 合并后居	p - 👺 - % :	• 0.0 .00 00 ÷.0	条件格式	套用	好	适中	- 插入		▼ 填充 →
→ 💞 格式刷	_							* #	格格式・				* *	◆ 清除 →
剪贴板。		字体	6	797	作方式	5 数字	E 6				样式		单元格	
35 +	× ×	f _x												
(A	В	С	D	E	F	G	н	1.0		1	K	1 1		
Type 1	D	C	U		Type 2	G	п			J	Type 3	L	, K	1
	Target=	Predict - Target =				Target=	Predict - Tar	raet =			Predict=	Target=	Predict - Tai	roet =
0.00556738	0.017542	-0.0119746			0.0105802	0.017542					0.0093056			0.00823641
-0.0141354		-0.00376825			-0.00969313	-0.0103671					-0.00732009			0.00304703
-0.0057172	-0.003825	-0.00189226			-0.00642355	-0.00382495					-0.00463411			.000809155
0.00704169	0.006428	0.000613691			0.00901731	0.006428	0.00258				0.00910698			0.00267898
0.0308507	0.0264088	0.00444197			0.0337688	0.0264088	0.007359				0.0270253			.000616528
-0.0021883		0.00307807			-0.00046627	-0.00526634	0.004800				0.00071819			0.00598453
-0.0130871		0.00337576			-0.0113662	-0.0164629	0.005096				-0.00727897			0.00918391
-0.0247906		0.00302531			-0.0231475	-0.0278159					-0.0169841			0.0108318
-0.0014491	0.02.10100	-0.00144908			0.000135808	0.0270200					0.00145618			0.00145618
-0.0014491 -0.0113933		-0.00243881			-0.00823311	-0.00895447					-0.00399029			0.00496418
0.0169606	0.0171851	-0.000224507			0.0167128	0.0171851					0.014685			0.00250014
0.00440235		0.0059397			0.00290704	-0.00153735	0.004444				0.00368464			0.00522199
-0.0275343		0.0027353			-0.0266293	-0.0302696					-0.0198617			0.0104079
-0.0177808	-0.017025	-0.000755776			-0.0120145	-0.017025	0.005010				-0.00704579			0.00997925
0.00792158		-0.00125947			0.00975484	0.00918104					0.00896551			.000215532
-0.0071447		-0.00237294			-0.00551828	-0.00477178	-0.00074				-0.00348436			0.00128742
0.0222506	0.0205895	0.00261234			0.0251058	0.0205895	0.004516				0.0202749			000314573
0.00207101		0.00195097			0.00297854	0.000120039	0.00285				0.00344964			0.0033296
0.0317878	0.0294929	0.00229485			0.0339313	0.0294929	0.004438				0.0275651			0.00192789
	-0.0025385	-0.00225466			-0.00039402	-0.0253851	0.002144				0.00153913			0.00407764
0.00361748		-0.000000031			0.00583988	0.00200031					0.00100310			0.00401104
0.0183378	0.015138	0.00319983			0.0214605	0.015138	0.006322				0.01788			0.00274197
	-0.0380406	0.00793252			-0.0302303	-0.0380406	0.007810				-0.0232701			0.0147706
0.00463658		0.000469526			0.00825969	0.00416706					0.00813958			0.00397253
-0.0263255		0.00414222			-0.0238981	-0.0304677	0.006569				-0.0171531			0.0133146
0.0115035	0.0073566	0.00414222			0.010861	0.0073566					0.00920546			0.00133146
0.0013035	0.0073500	0.00272869			0.010861	0.0073500	0.003304				0.00920340			0.00164660
0.00585713		-0.00359702			0.00371041	0.00945415					0.00810488			0.0043113
0.00383713		0.0080258			0.0043532	-0.00394178	0.001203				0.00488432			0.00134927
-0.0213668		0.00121625			-0.0194655	-0.00394178	0.008294				-0.0133726			0.00082011
		0.00121625			-0.0194655	-0.0227084	0.005117				-0.0135726			0.00921053
	0.00446867	0.00537589			0.0126135	0.00446867	0.005575				0.0113441			0.0111546
		0.00623693			-0.0120135	-0.0216807	0.008144				-0.00690661			0.00687543
-0.0107683 -0.0144833		0.0109125			-0.0107972	-0.0216807	0.01088				-0.00784308			0.0147741
	-0.0163361	0.0016526			-0.0121451	-0.0163361	0.004191				-0.00784308			0.00849302
-0.020303	0.0233023	0.0000293			-0.023040	-0.0233323	0.004004	-20			-0.0107301	-0.023932		0.0111702
Type 1 Topolo	ogy is (4.4.1)													
Type 2 Topolo														
Type 3 Topolo														

This is the prediction result by 3 different models and actual daily performance of Nasdaq composite index.(Target is the actual Nasdaq Composite Index performance)



We can find that modelType1 has the smallest average standard error = 0.00353092, which is the best model among 3 models. The modelType3 has the largest average standard error = 0.005773115, which is the worst.

A	В	С	D	E	F	G	Н	1]	K	L	M	N	0
	Type 1					Type 2					Type 3			
Date	Predict=	Target=	Predict - Target =	standard error		Predict=	Target=	Predict - Target =	square erro	r	Predict=	Target=	Predict - Target =	square erro
2018/11/	1 0.00556738	0.017542	-0.0119746	0.0119746		0.0105802	0.017542	-0.00696179	0.006962		0.0093056	0.017542	-0.00823641	0.008236
2018/11/	2 -0.0141354	-0.0103671	-0.00376825	0.00376825		-0.00969313	-0.0103671	0.000673998	0.000674		-0.00732009	-0.0103671	0.00304703	0.003047
2018/11/	5 -0.0057172	-0.00382495	-0.00189226	0.00189226		-0.00642355	-0.00382495	-0.0025986	0.002599		-0.00463411	-0.00382495	-0.000809155	0.0008091
2018/11/	6 0.00704169	0.006428	0.000613691	0.00061369		0.00901731	0.006428	0.0025893	0.002589		0.00910698	0.006428	0.00267898	0.002678
2018/11/	7 0.0308507	0.0264088	0.00444197	0.00444197		0.0337688	0.0264088	0.00735998	0.00736		0.0270253	0.0264088	0.000616528	0.0006165
2018/11/	8 -0.0021883	-0.00526634	0.00307807	0.00307807		-0.000466272	-0.00526634	0.00480006	0.0048		0.00071819	-0.00526634	0.00598453	0.005984
2018/11/	9 -0.0130871	-0.0164629	0.00337576	0.00337576		-0.0113662	-0.0164629	0.00509669	0.005097		-0.00727897	-0.0164629	0.00918391	0.00918
2018/11/1	2 -0.0247906	-0.0278159	0.00302531	0.00302531		-0.0231475	-0.0278159	0.00466845	0.004668		-0.0169841	-0.0278159	0.0108318	0.0108
2018/11/1	3 -0.0014491	0	-0.00144908	0.00144908		0.000135808	0	0.000135808	0.000136		0.00145618	0	0.00145618	0.001456
2018/11/1	4 -0.0113933	-0.00895447	-0.00243881	0.00243881		-0.00823311	-0.00895447	0.000721362	0.000721		-0.00399029	-0.00895447	0.00496418	0.00496
2018/11/1	5 0.0169606	0.0171851	-0.000224507	0.00022451		0.0167128	0.0171851	-0.000472302	0.000472		0.014685	0.0171851	-0.00250014	0.002500
	6 0.00440235					0.00290704					0.00368464			
	9 -0.0275343			0.0027353		-0.0266293	-0.0302696				-0.0198617			
	0 -0.0177808					-0.0120145	-0.017025				-0.00704579			
	1 0.00792158					0.00975484	0.00918104				0.00896551			
	3 -0.0071447					-0.00551828					-0.00348436			
2018/11/2				0.00166112		0.0251058	0.0205895				0.0202749			
	7 0.00207101			0.00195097		0.00297854			0.002859		0.0202749			0.000314
2018/11/2				0.00229485		0.0339313	0.0294929				0.0275651	0.0294929		
2018/11/2						-0.000394017		0.00214449			0.0275651			
	0 0.00361748					0.00583988					0.00193913			
							0.00790036							
2018/12/				0.00319983		0.0214605					0.01788			
	4 -0.0301081	-0.0380406		0.00793252		-0.0302303	-0.0380406				-0.0232701			0.0147
	6 0.00463658					0.00825969	0.00416706				0.00813958			
	7 -0.0263255			0.00414222		-0.0238981	-0.0304677				-0.0171531			
2018/12/1				0.00414689		0.010861	0.0073566				0.00920546			
	1 0.00433969			0.00272869		0.00571641	0.001611				0.0061223			0.0045
	2 0.00585713					0.00819078					0.00810488			0.001349
	3 0.00408402			0.0080258		0.0043532					0.00488432			0.008826
2018/12/1	4 -0.0213668	-0.0225831	0.00121625	0.00121625		-0.0194655	-0.0225831	0.00311757	0.003118		-0.0133726			0.009210
2018/12/1	7 -0.0171325	-0.0227084	0.00557589	0.00557589		-0.0171329	-0.0227084		0.005575		-0.0115538			0.0111
2018/12/1	8 0.0107056	0.00446867	0.00623693	0.00623693		0.0126135	0.00446867	0.00814484	0.008145		0.0113441	0.00446867	0.00687543	0.00687
2018/12/1	9 -0.0107683	-0.0216807	0.0109125	0.0109125		-0.0107972	-0.0216807	0.0108836	0.010884		-0.00690661	-0.0216807	0.0147741	0.0147
2018/12/2	0 -0.0144833	-0.0163361	0.0018528	0.0018528		-0.0121451	-0.0163361	0.00419103	0.004191		-0.00784308	-0.0163361	0.00849302	0.008493
2018/12/2	1 -0.026303	-0.0299323		0.0036293		-0.025848	-0.0299323	0.00408425	0.004084		-0.0187561	-0.0299323	0.0111762	0.0111
			average standard erro	0.00353092				average standard error	0.004128				average standard error	0.005773
			Av											
			f											
	Type 1 Topo	lony is (4.4.1)												
		logy is (4,4,3,1)												
		logy is (4,4,3,1)												
	Type 3 T000	rougy is (4,4,3,2,1)												

Neural network	Type1	Type2	Type3
Average Standard	0.00353092	0.004128	0.005773115
Error			

7.Conclusion

- 1.We find that neural network type1 has the best prediction performance among three neural networks.
- 2. More hidden layers do not necessarily give better prediction result.
- 3. We need to find a learning rate that is low enough that the network converges, but high enough that we don't have to spend too much time training it.

8.Reference

https://finance.yahoo.com/quote/AAPL

https://finance.yahoo.com/quote/GOOG

https://finance.yahoo.com/quote/AMZN

https://finance.yahoo.com/quote/MSFT

https://finance.yahoo.com/quote/%5EIXIC

Professional C++, Second Edition