CS7015: Assignment 1 Report

Team 5
Abhik Debnath – MA17M001
Sushant Kumar Seet – MA17M012

Task 1: Function Approximation

Data:
Energy efficiency Dataset
Data Set Characteristics: Multivariate
Number of Instances: 768
Attribute Characteristics: Real
Number of Attributes: 8
The dataset consists of 8 attributes (features) and 2 Target variable (last two column).
The target variables are Heating Load and Cooling Load

Case 1: Multi-layer Feedforward Network with **1 hidden layer**

Baseline System

Features: Un-normalized
Mode of Learning: Pattern
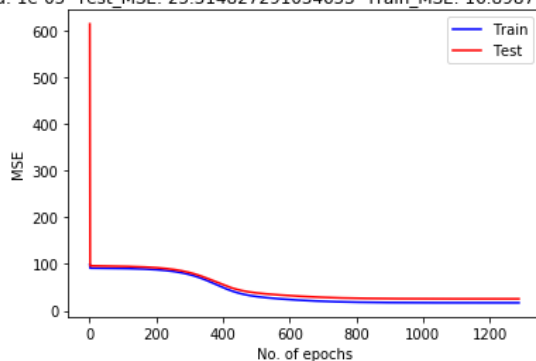Number of hidden nodes: 15
Activation Function: Logistic
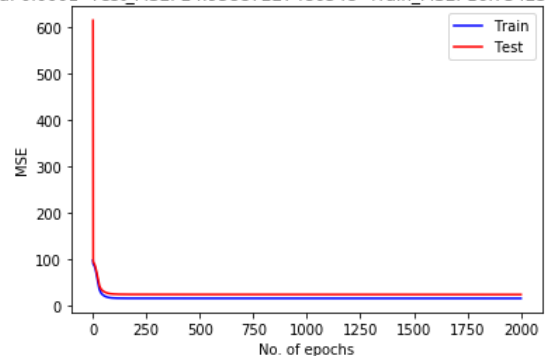Slope, β: [0.00001, 0.0001, 0.001, 0.01, 0.1, 0.2, 0.5, 1]
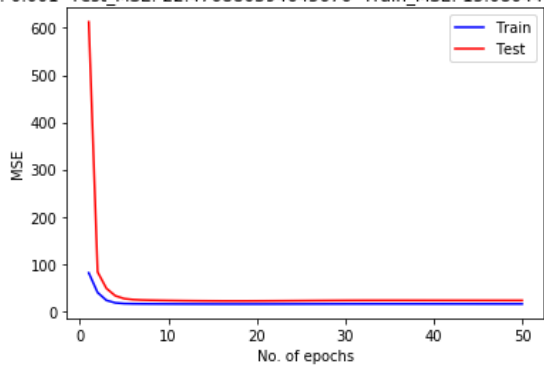Weight Update Rule: Delta
Learning Rate: 0.01

Number of epochs for convergence = 1288

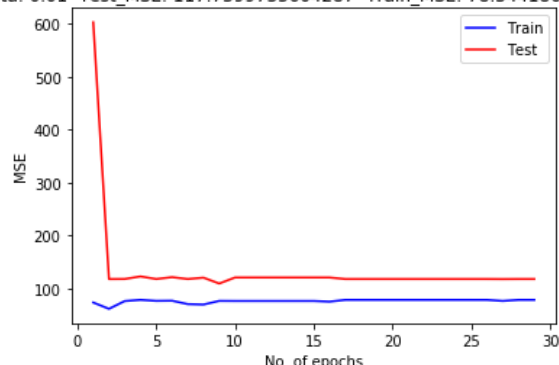Number of epochs for convergence > 2000

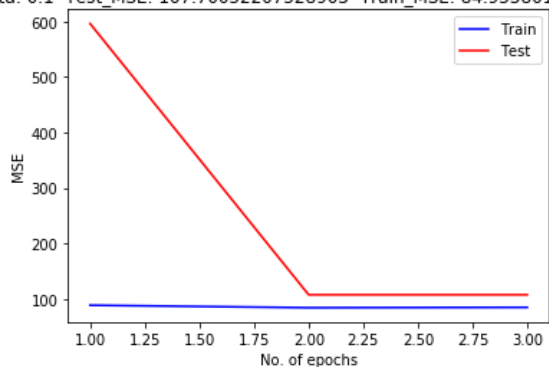Beta: 0.001 Test_MSE: 22.476880394645676 Train_MSE: 15.080447009051353

Number of epochs for convergence = 50

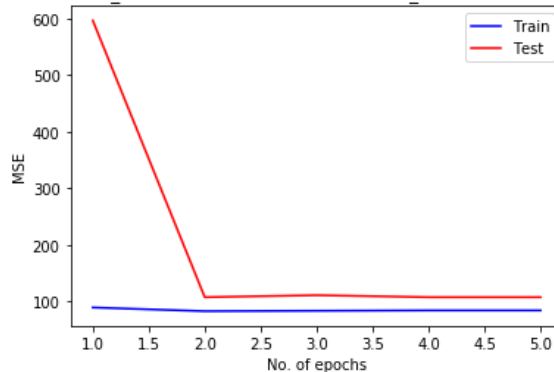Beta: 0.01 Test_MSE: 117.7399735604287 Train_MSE: 78.34418699317453

Number of epochs for convergence = 29

Beta: 0.1 Test_MSE: 107.70052207528903 Train_MSE: 84.95380165218478
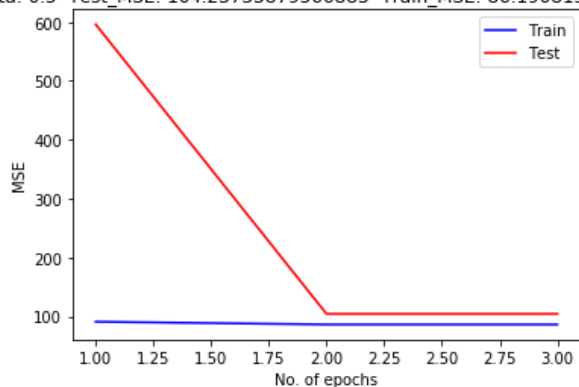
Number of epochs for convergence = 3

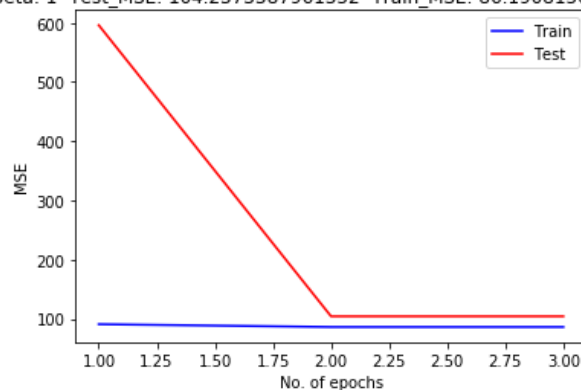Beta: 0.2 Test_MSE: 107.7005566653813 Train_MSE: 84.34191202489097

Number of epochs for convergence = 5

Beta: 0.5 Test_MSE: 104.23733879566883 Train_MSE: 86.19081568663387
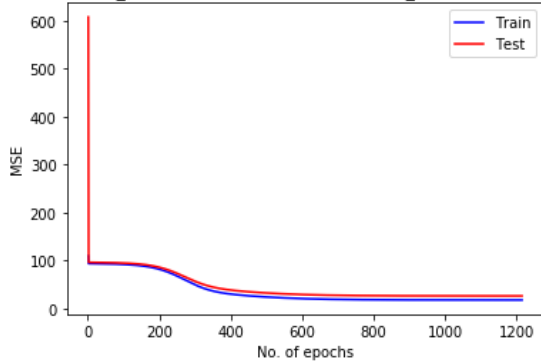
Number of epochs for convergence = 3

Beta: 1 Test_MSE: 104.2373387961352 Train_MSE: 86.19081568694521

Number of epochs for convergence = 3

Features: Un-normalized
Mode of Learning: Pattern
Number of hidden nodes: 5
Activation Function: Logistic
Slope, β: [0.00001, 0.0001, 0.001, 0.01, 0.1, 0.2]
Weight Update Rule: Delta
Learning Rate: 0.01

Number of epochs for convergence = 1216

Number of epochs for convergence > 2000

Number of epochs for convergence = 1137

Number of epochs for convergence = 12

Number of epochs for convergence = 3

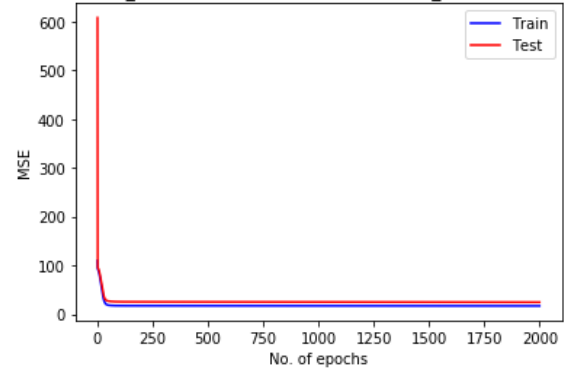
Number of epochs for convergence = 3

Features: Un-normalized
Mode of Learning: Pattern
Number of hidden nodes: 25
Activation Function: Logistic
Slope, β: [0.00001, 0.0001, 0.001, 0.01, 0.1, 0.2]
Weight Update Rule: Delta
Learning Rate: 0.01

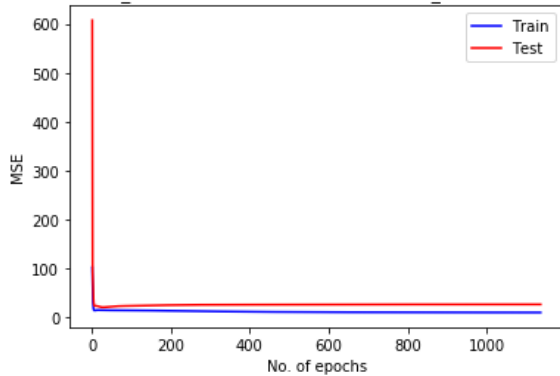Beta: 1e-05 Test_MSE: 26.049398878710026 Train_MSE: 15.404214807979052
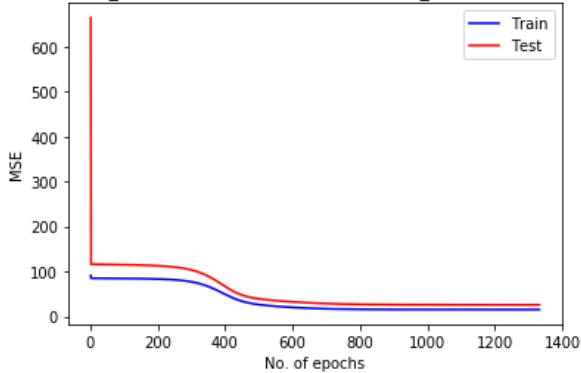
Number of epochs for convergence = 1332

Beta: 0.0001 Test_MSE: 26.319825614410856 Train_MSE: 15.504144678934605

Number of epochs for convergence = 237

Beta: 0.001 Test_MSE: 24.865639222720574 Train_MSE: 14.786633323736904

Number of epochs for convergence = 41

Beta: 0.01 Test_MSE: 132.1460633251683 Train_MSE: 77.01641665452725
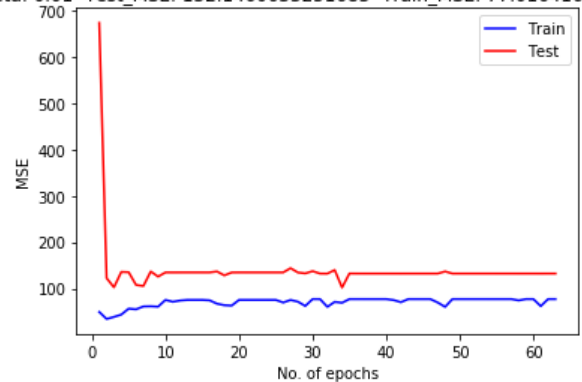
Number of epochs for convergence = 63

Beta: 0.1 Test_MSE: 134.57122290314274 Train_MSE: 75.18361777081006
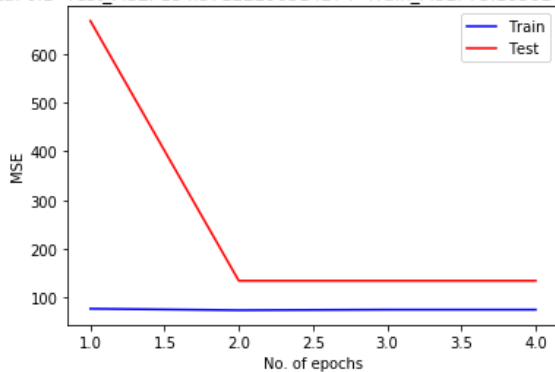
Number of epochs for convergence = 4

Beta: 0.2 Test_MSE: 129.34428808751966 Train_MSE: 77.31215310942483
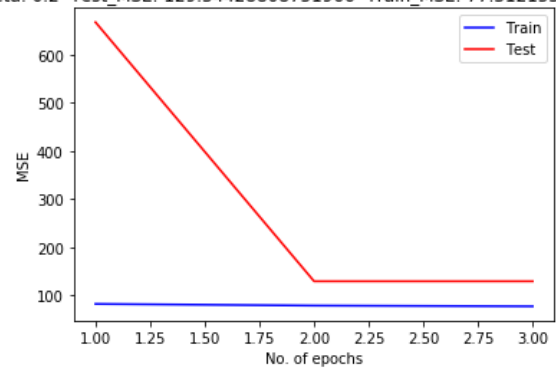
Number of epochs for convergence = 3

Features: Normalized
Mode of Learning: Pattern
Number of hidden nodes: 15
Activation Function: Logistic
Slope, β: [0.01, 0.1, 0.2, 0.5, 1, 10]
Weight Update Rule: Delta
Learning Rate: 0.01

Beta: 0.01 Test_MSE: 12.10549519556796 Train_MSE: 9.195317259890887
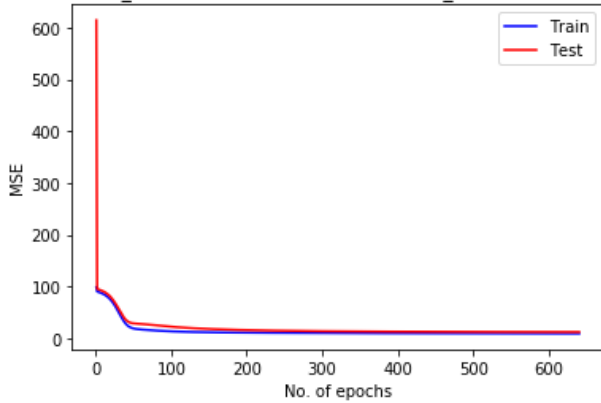
Number of epochs for convergence = 640

Beta: 0.1 Test_MSE: 11.119381589065705 Train_MSE: 8.265800442905766

Number of epochs for convergence = 252

Beta: 0.2 Test_MSE: 2.164838677268394 Train_MSE: 1.9745067762491872
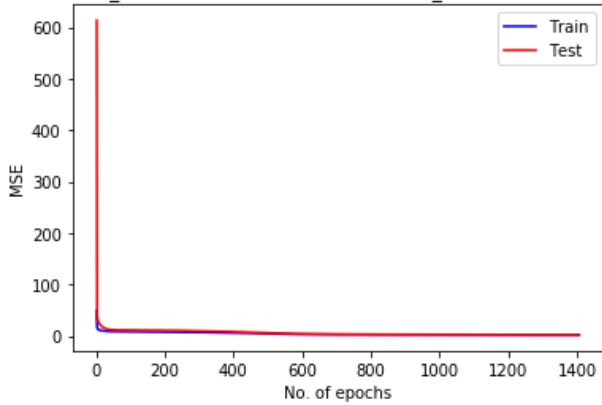
Number of epochs for convergence = 1406

Beta: 0.5 Test_MSE: 1.7631492685423635 Train_MSE: 1.5810078098447673
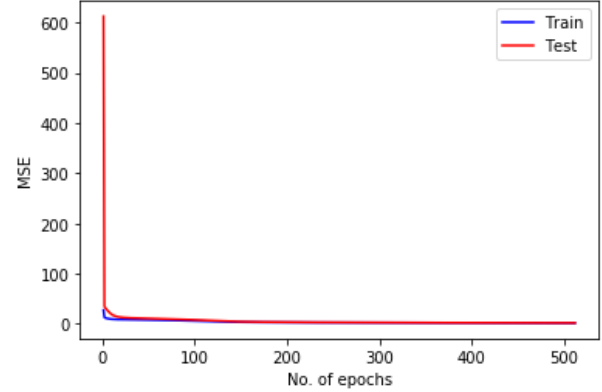
Number of epochs for convergence = 512

Beta: 1 Test_MSE: 1.7801269344006068 Train_MSE: 1.4851147072771427
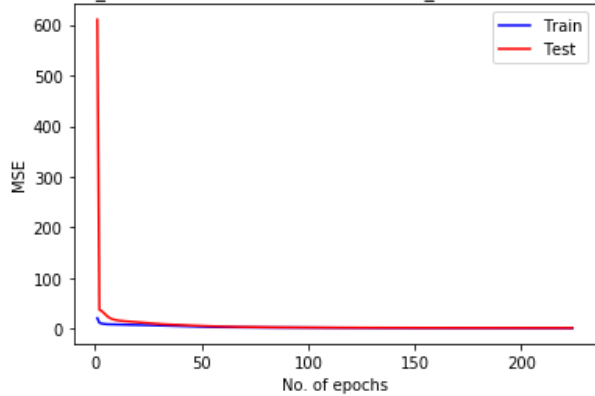
Number of epochs for convergence = 224

Beta: 10 Test_MSE: 31.850798577144854 Train_MSE: 14.59874017842085
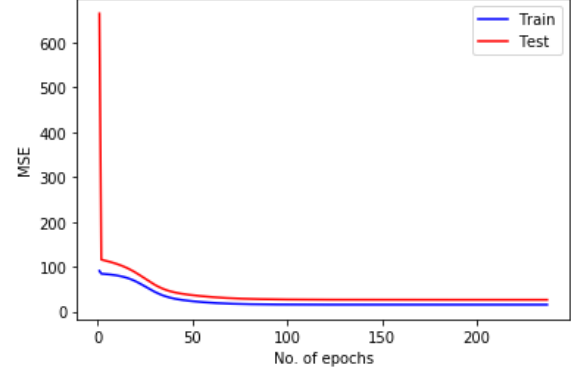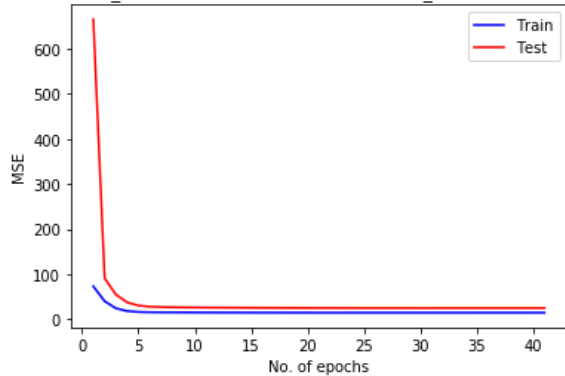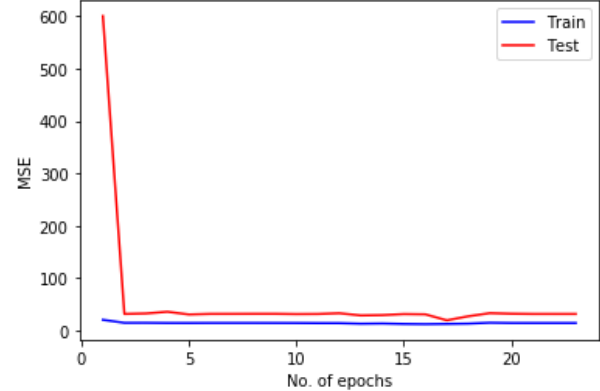

Number of epochs for convergence = 23

Features: Normalized
Mode of Learning: Pattern
Number of hidden nodes: 5
Activation Function: Logistic
Slope, β: [0.01, 0.1, 0.2, 0.5, 1, 10]
Weight Update Rule: Delta
Learning Rate: 0.01

Beta: 0.01 Test_MSE: 12.008233364825486 Train_MSE: 9.269160965802111
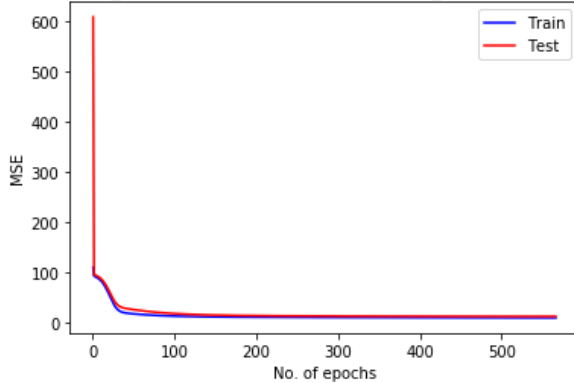
Number of epochs for convergence = 566

Beta: 0.1 Test_MSE: 11.212627239213813 Train_MSE: 8.289594227226313

Number of epochs for convergence = 102

Beta: 0.2 Test_MSE: 3.2285810516031095 Train_MSE: 2.58895922285768988
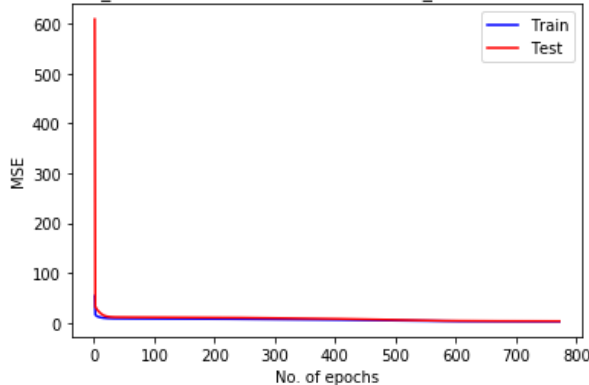
Number of epochs for convergence = 772

Beta: 0.5 Test_MSE: 3.836920063956168 Train_MSE: 3.1228283673031254
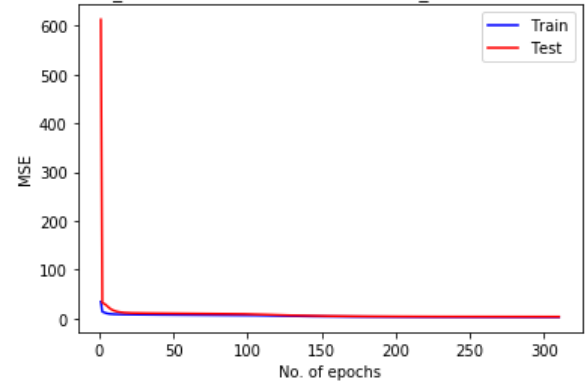
Number of epochs for convergence = 310

Beta: 1 Test_MSE: 3.641374689556929 Train_MSE: 2.9036874409994646
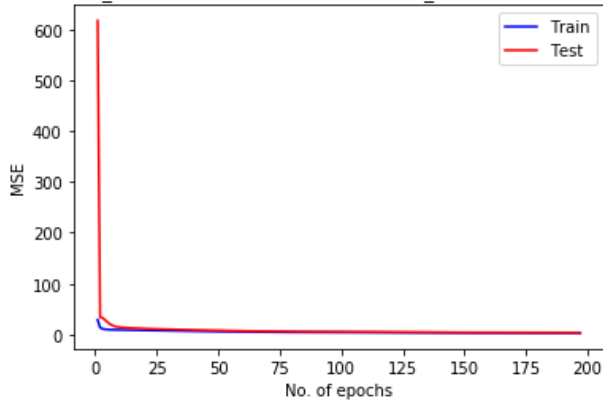
Number of epochs for convergence = 197

Beta: 10 Test_MSE: 30.551948019309055 Train_MSE: 16.638078492173324
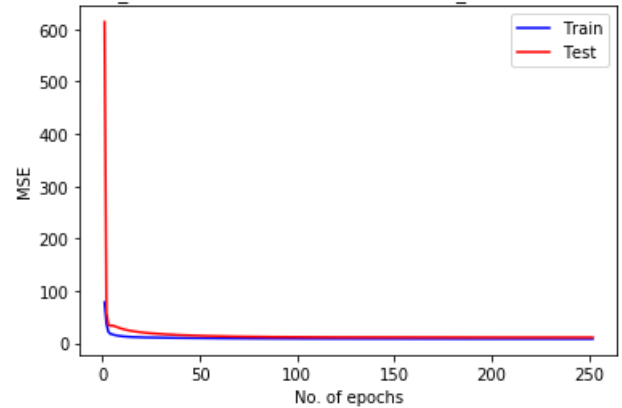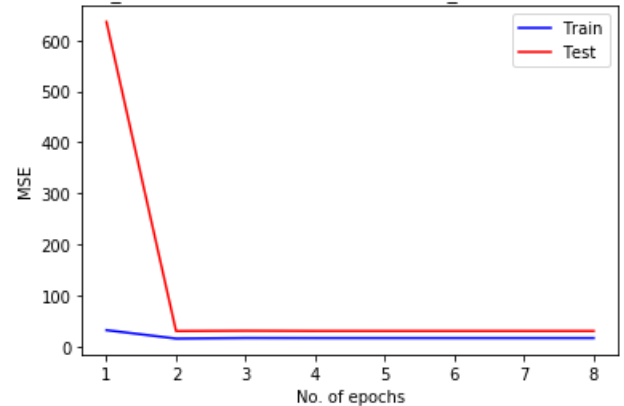

Number of epochs for convergence = 8

Features: Normalized
Mode of Learning: Pattern
Number of hidden nodes: 25
Activation Function: Logistic
Slope, β: [0.01, 0.1, 0.2, 0.5, 1, 10]
Weight Update Rule: Delta
Learning Rate: 0.01

Beta: 0.01 Test_MSE: 11.316306691402293 Train_MSE: 9.549562987125416
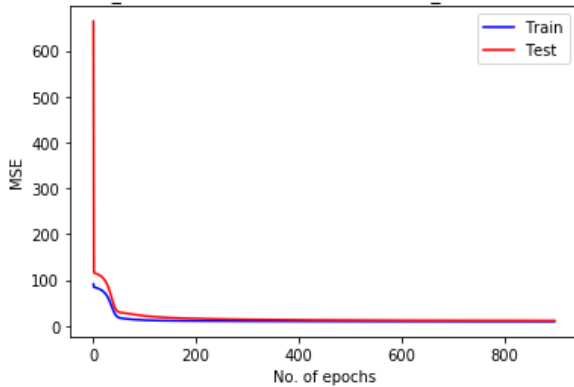
Number of epochs for convergence = 898

Beta: 0.1 Test_MSE: 2.5441083448570367 Train_MSE: 2.062015963722465

Number of epochs for convergence > 2000

Beta: 0.2 Test_MSE: 2.0206417615939207 Train_MSE: 1.7910144048087315
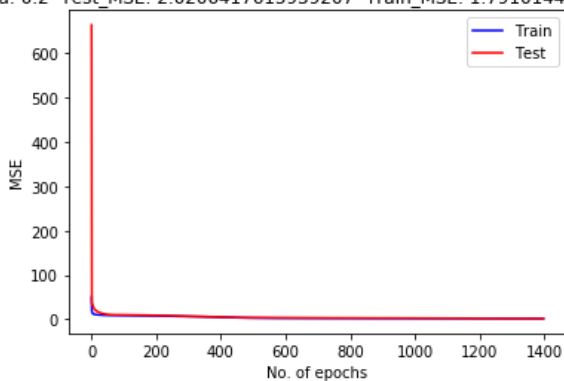
Number of epochs for convergence = 1397

Beta: 0.5 Test_MSE: 1.7523341897877323 Train_MSE: 1.4985474818375282
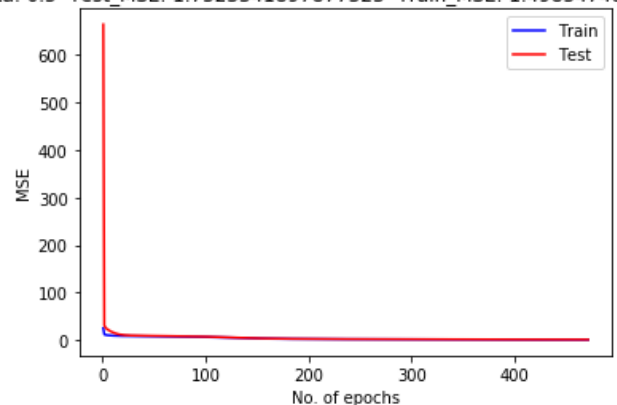
Number of epochs for convergence = 471

Beta: 1 Test_MSE: 1.9255194499964245 Train_MSE: 1.5183826380649805
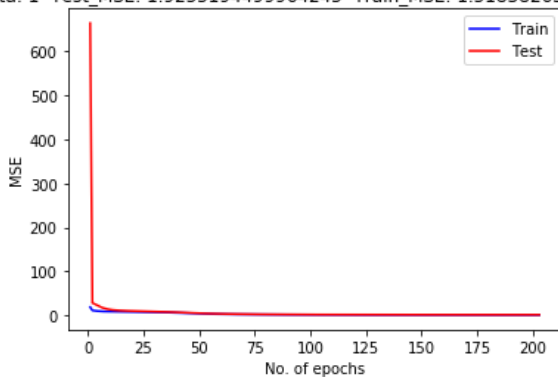
Number of epochs for convergence = 203

Beta: 10 Test_MSE: 18.397974899867343 Train_MSE: 10.165903282135618
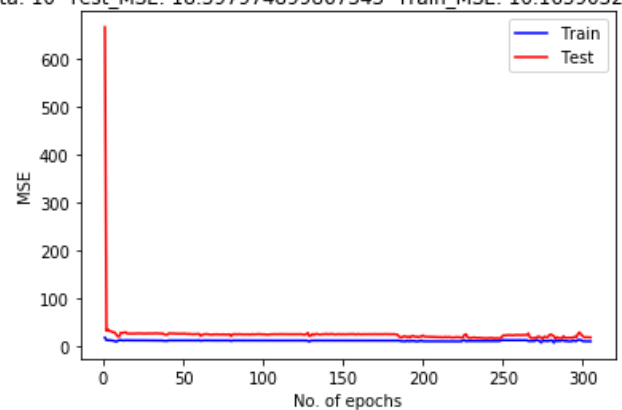
Number of epochs for convergence = 305

Case 2: Multi-layer Feedforward Network with **2 hidden layers**

Baseline System

Features: Un-normalized
Mode of Learning: Pattern
Number of hidden nodes: 15
Activation Function: Logistic
Slope, β: [0.005, 0.007, 0.01]
Weight Update Rule: Delta
Learning Rate: 0.01

Number of epochs for convergence = 1850

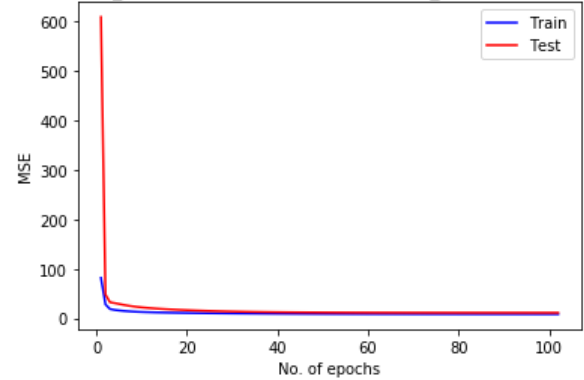Number of epochs for convergence = 815

Number of epochs for convergence = 64
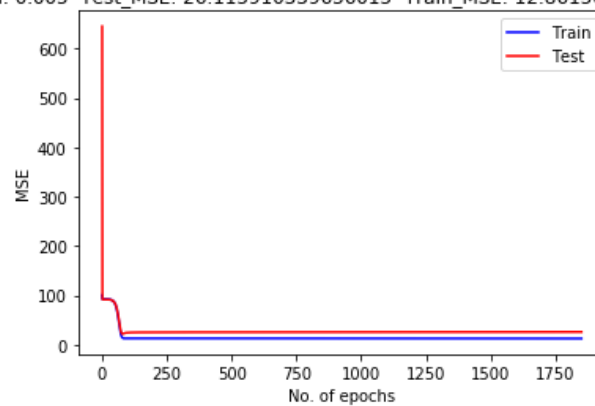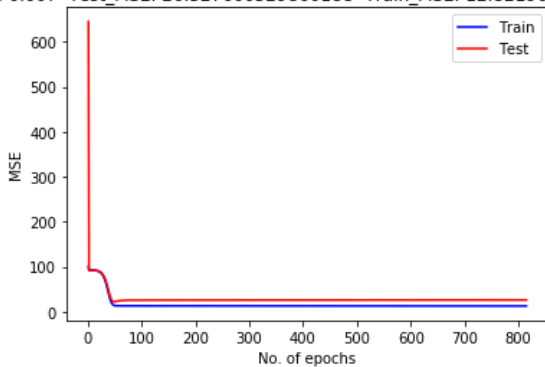
Features: Un-normalized
Mode of Learning: Pattern
Number of hidden nodes: 5
Activation Function: Logistic
Slope, β: [0.005, 0.007, 0.01]
Weight Update Rule: Delta
Learning Rate: 0.001

Beta: 0.005 Test_MSE: 92.43515129491695 Train_MSE: 98.08490216396117
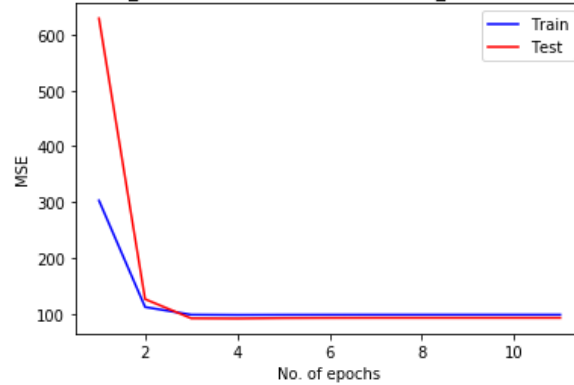
Number of epochs for convergence = 11

Beta: 0.007 Test_MSE: 21.245380796256313 Train_MSE: 17.326810464983705
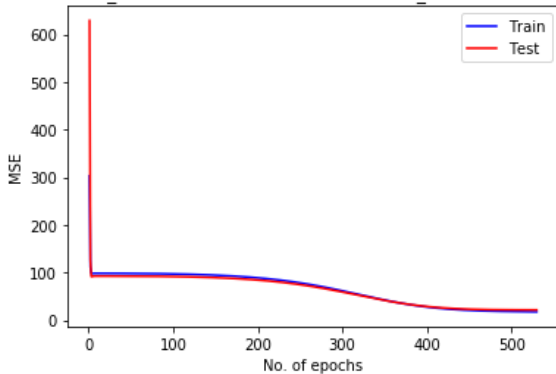
Number of epochs for convergence = 530

Beta: 0.01 Test_MSE: 21.714359672593122 Train_MSE: 15.798654237291672
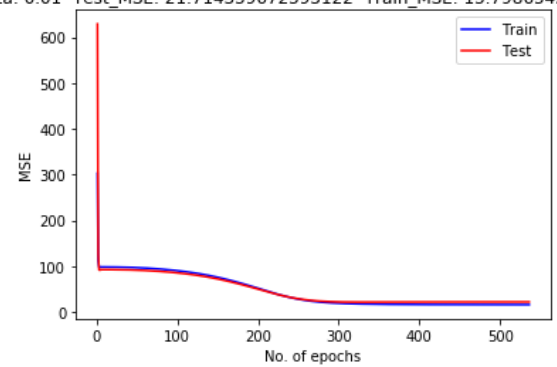
Number of epochs for convergence = 536

Features: Un-normalized
Mode of Learning: Pattern
Number of hidden nodes: 25
Activation Function: Logistic
Slope, β: [0.005, 0.007, 0.01]
Weight Update Rule: Delta
Learning Rate: 0.01

Number of epochs for convergence = 869

Number of epochs for convergence = 374

Number of epochs for convergence = 437
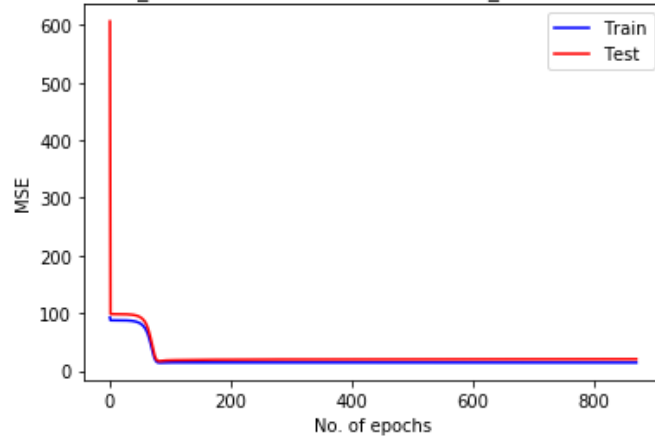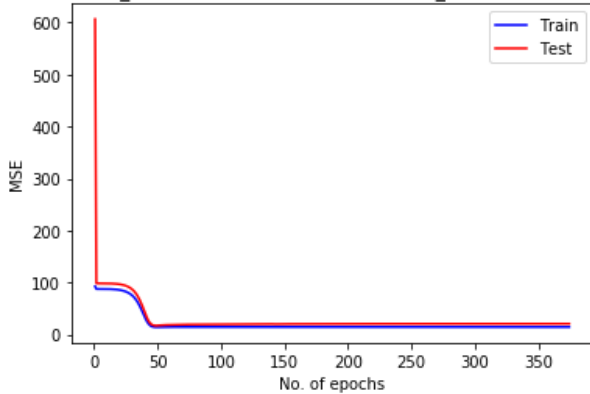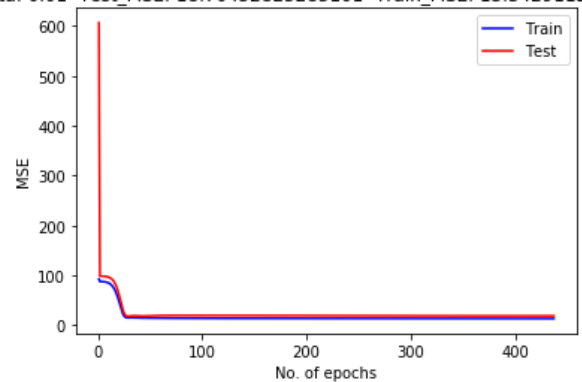
Features: Normalized
Mode of Learning: Pattern
Number of hidden nodes: 15
Activation Function: Logistic
Slope, β: [0.2, 0.5, 1]
Weight Update Rule: Delta
Learning Rate: 0.01

Beta: 0.2 Test_MSE: 1.5896910261922144 Train_MSE: 1.5453406446369695

Number of epochs for convergence = 1405
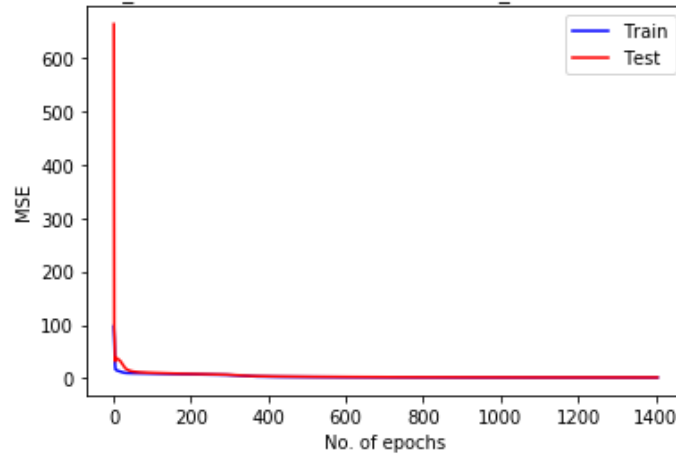
Beta: 0.5 Test_MSE: 1.810637879388642 Train_MSE: 1.569858458553743

Number of epochs for convergence = 497
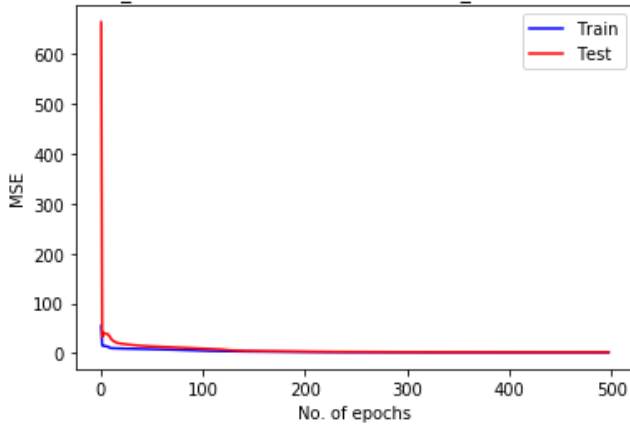
Beta: 1 Test_MSE: 1.8325073121271998 Train_MSE: 1.3767417237082213
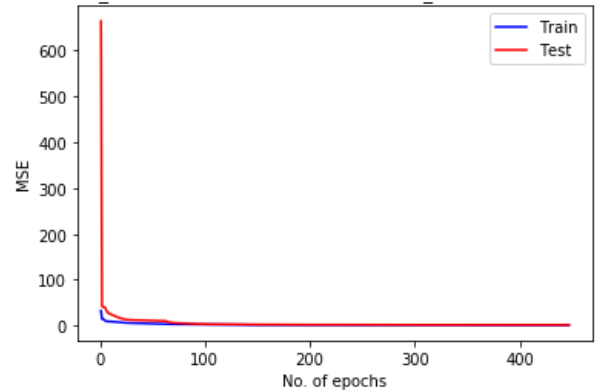
Number of epochs for convergence = 447
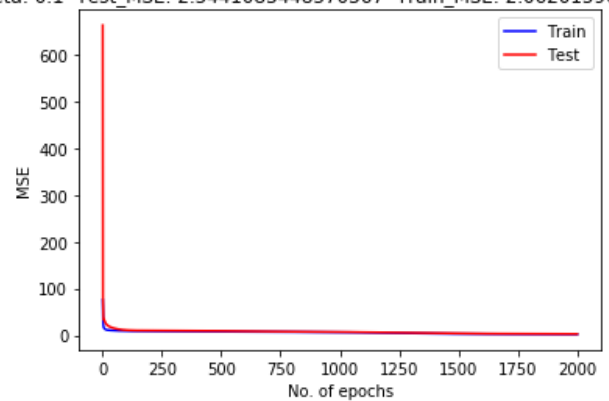
Features: Normalized
Mode of Learning: Pattern
Number of hidden nodes: 5
Activation Function: Logistic
Slope, β: [0.2, 0.5, 1]
Weight Update Rule: Delta
Learning Rate: 0.01

Beta: 0.2 Test_MSE: 1.7719554342463841 Train_MSE: 1.7365975392104114
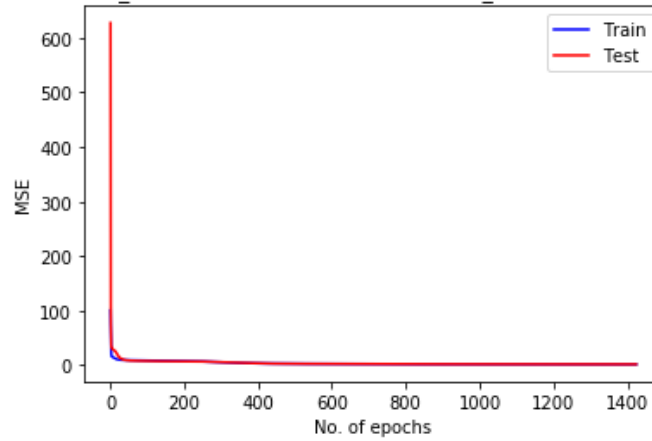
Number of epochs for convergence = 1424

Beta: 0.5 Test_MSE: 2.761140648473305 Train_MSE: 2.6345334380841208
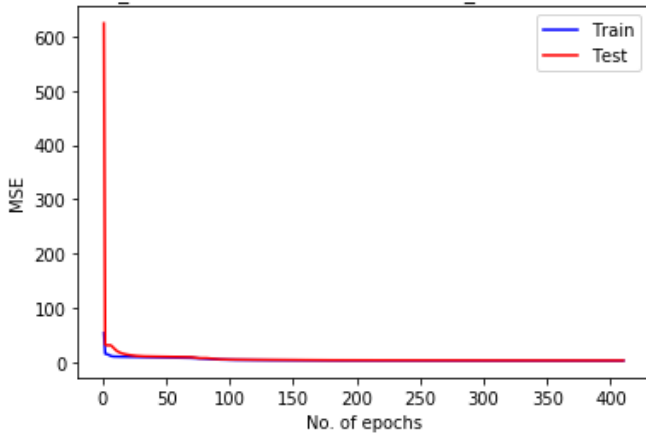
Number of epochs for convergence = 410

Beta: 1 Test_MSE: 3.9272918527492577 Train_MSE: 2.38270308830567
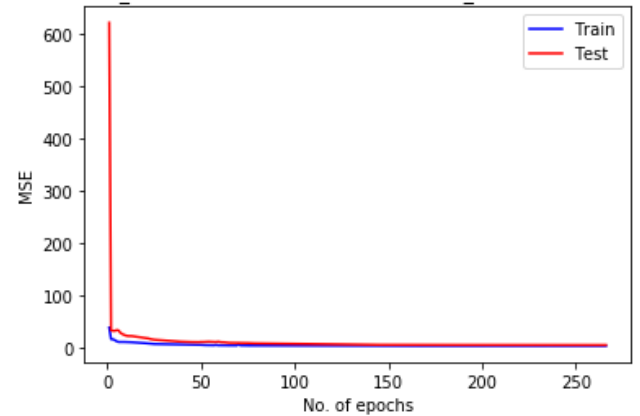
Number of epochs for convergence = 266

Features: Normalized
Mode of Learning: Pattern
Number of hidden nodes: 25
Activation Function: Logistic
Slope, β: [0.2, 0.5, 1]
Weight Update Rule: Delta
Learning Rate: 0.01

Beta: 0.2 Test_MSE: 1.5416375607620312 Train_MSE: 1.5502342387264088
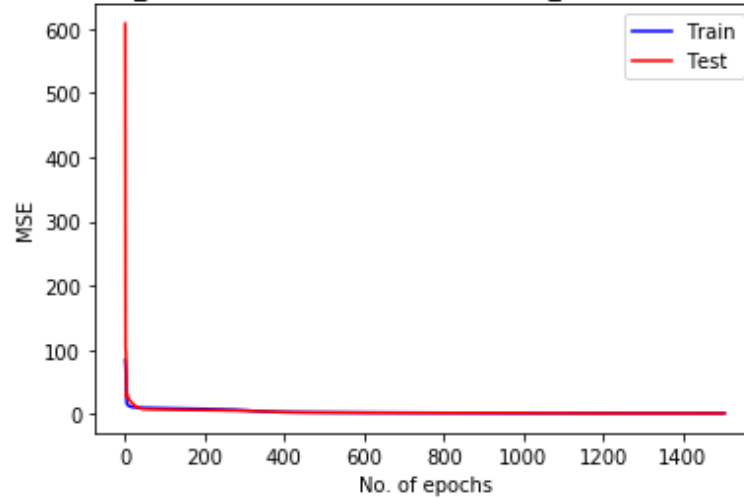
Number of epochs for convergence = 1503

Beta: 0.5 Test_MSE: 1.6676454502994698 Train_MSE: 1.659877032795916
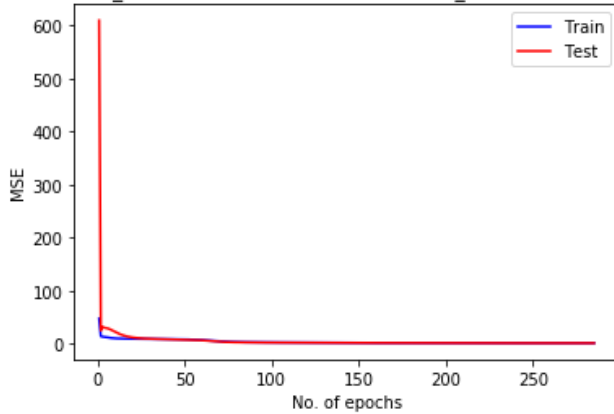
Number of epochs for convergence = 285

Beta: 1 Test_MSE: 1.5750917034289804 Train_MSE: 1.5566104255709974
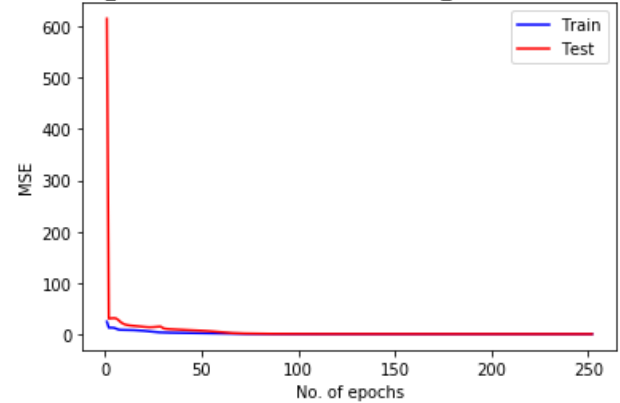
Number of epochs for convergence = 252

Observations:

Features: Un-Normalized
Number of Hidden layers: 1
Mode of Learning: Pattern
Activation Function: Logistic
Weight Update Rule: Delta
Learning Rate: 0.01

No. of hidden nodes	Slope, β	Test_MSE	Train_MSE	Epochs for convergence
15	0.00001	25.31482	16.89877	1288
	0.0001	24.95887	16.73424	>2000
	0.001	22.47688	15.08044	50
	0.01	117.73997	78.34418	29
	0.1	107.70052	84.85580	3
	0.2	107.70055	84034191	5
	0.5	104.23733	86.19081	3
	1	104.23734	86.19081	3
5	0.00001	26.42026	17.88151	1216
	0.0001	24.44379	16.84945	>2000
	0.001	25.92683	9.33028	1137
	0.01	95.24688	92.02015	12
	0.1	95.37836	92.05308	3
	0.2	95.37835	93.05307	3
25	0.00001	26.04939	26.31982	1332
	0.0001	26.31982	15.50414	237
	0.001	24.86564	14.78663	41
	0.01	132.14606	77.01641	63
	0.1	134.57122	75.18362	4
	0.2	129.34428	77.31215	3

Features: Normalized
Number of Hidden layers: 1
Mode of Learning: Pattern
Activation Function: Logistic
Weight Update Rule: Delta
Learning Rate: 0.01

No. of hidden nodes	Slope, β	Test_MSE	Train_MSE	Epochs for convergence
15	0.01	12.10549	9.19532	640
	0.1	11.11938	8.26580	252
	0.2	2.16483	1.97451	1406
	0.5	1.76315	1.58100	512
	1.0	1.78012	1.48511	224
	10	31.85079	14.59874	23
5	0.01	12.00823	9.26916	566
	0.1	11.21262	8.25959	102
	0.2	3.22858	2.58895	772
	0.5	3.83692	3.12282	310
	1.0	3.64137	2.90368	197
	10	30.55194	16.63807	8
25	0.01	11.31630	9.54956	898
	0.1	2.54410	2.06201	>2000
	0.2	2.02064	1.79101	1397
	0.5	1.75233	1.49854	471
	1.0	1.92552	1.51838	203
	10	18.39797	10.16590	305

Features: Un-Normalized
Number of Hidden layers: 2
Mode of Learning: Pattern
Activation Function: Logistic
Weight Update Rule: Delta
Learning Rate: 0.01

No. of hidden nodes	Slope, β	Test_MSE	Train_MSE	Epochs for convergence
15	0.005	26.11591	12.86156	1850
	0.007	26.32760	12.82196	815
	0.001	26.67501	13.16194	64
5	0.005	92.43515	98.08492	11
	0.007	21.24538	17.32681	530
	0.001	21.71435	15.79865	536
25	0.005	20.16343	14.24824	869
	0.007	20.67294	14.42876	374
	0.001	18.70452	13.34291	437

Features: Normalized
Number of Hidden layers: 2
Mode of Learning: Pattern
Activation Function: Logistic
Weight Update Rule: Delta
Learning Rate: 0.01

No. of hidden nodes	Slope, β	Test_MSE	Train_MSE	Epochs for convergence
15	0.2	1.58969	1.54534	1405
	0.5	1.81063	1.569858	497
	1	1.83251	1.37674	447
5	0.2	1.77195	1.73659	1424
	0.5	2.76114	2.63453	410
	1	3.92729	2.38270	266
25	0.2	1.54163	1.55023	1503
	0.5	1.66764	1.65987	285
	1	1.57509	1.55661	252

- The data is highly un-normalized.
- The rule of thumb used for selecting the upper bound on number of hidden neurons:

$$N_h = \frac{N_s}{\alpha * (N_i + N_o)}$$

N_i = Number of input neurons

N_o = Number of output neurons

N_s = Number of training data samples

α = arbitrary scaling factor (here 2)

- From the baseline system, we choose the best β, number of hidden nodes and number of hidden layers respectively which are to be used for further analysis with un-normalized data.

  β = 0.001, Number of hidden nodes = 15, Number of hidden layer = 1

- From the baseline system, we choose the best β, number of hidden nodes and number of hidden layers respectively which are to be used for further analysis with normalized data.

  β = 1, Number of hidden nodes = 15, Number of hidden layer = 1


```
y_train_model = NN_batch.predict(X_train)
ScatterPlot(y_train,y_train_model)
```

```
y_test_model = NN_batch.predict(X_test)
ScatterPlot(y_test,y_test_model)
```
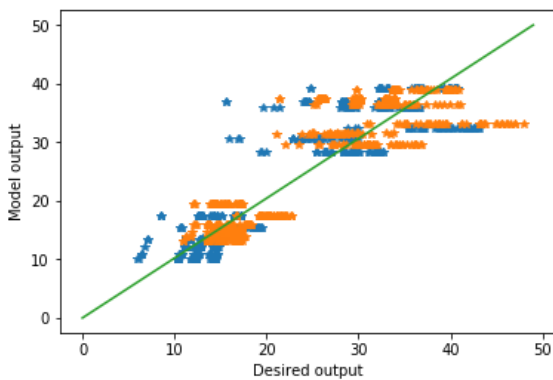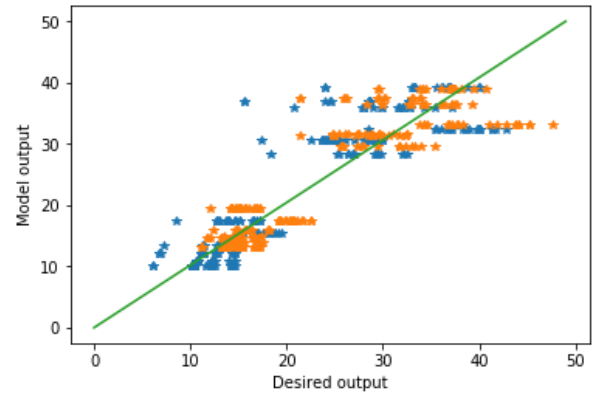
Scatterplots for un-normalized Baseline System


```
y_train_model = NN_batch.predict(X_train)
ScatterPlot(y_train,y_train_model)
```

```
y_test_model = NN_batch.predict(X_test)
ScatterPlot(y_test,y_test_model)
```
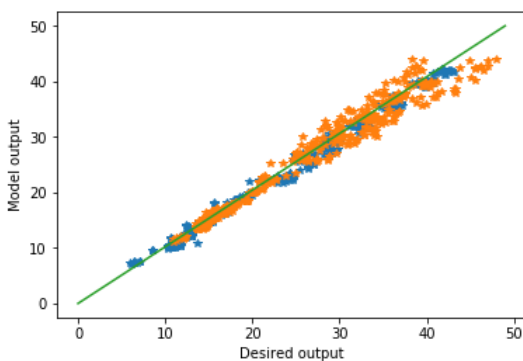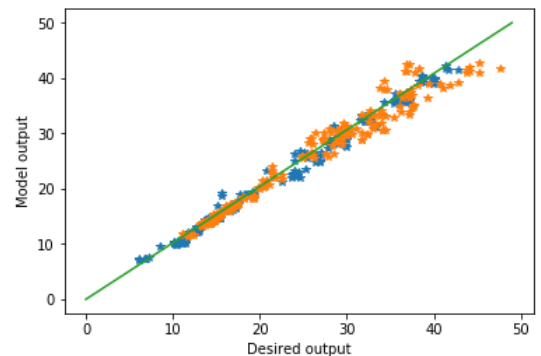
Scatterplots for normalized Baseline System

Experiments

Features: Un-Normalized
Mode of Learning: Pattern
Slope, β: 0.001
Number of Hidden Layers: 1
Number of Hidden Nodes: 15

Activation Function	Weight Update Rule	Learning Rate	Test_MSE	Train_MSE	Epochs for convergence
Logistic	Delta	0.01	23.72292	13.39906	167
	Generalized Delta	0.01	23.80066	12.82381	23
	AdaGrad	0.01	15.89439	16.52993	674
	RMSprop	0.01	22.67976	12.25360	80
	Adam	0.01	9.04521	6.70844	>2000
Tanh	Delta	0.01	29.15239	10.40841	229
	Generalized Delta	0.01	29.83644	10.30764	206
	AdaGrad	0.01	15.76964	15.06833	>2000
	RMSprop	0.01	15.42368	8.40982	829
	Adam	0.01	12.71977	7.65808	176
ReLU	Delta	0.01	98.67681	95.42270	6
	Generalized Delta	0.01	98.64000	95.42161	4
	AdaGrad	0.01	95.52193	96.39897	>2000
	RMSprop	0.01	96.31490	96.50874	21
	Adam	0.01	95.79110	96.77534	16

Features: Un-Normalized
Mode of Learning: Pattern
Activation Function: Logistic
Slope, β: 0.001
Number of Hidden Nodes: 15
Weight Update Rule: Adam
Learning Rate: 0.01

Scatterplot for Training Data

Scatterplot for Test Data

Features: Un-Normalized
Mode of Learning: Pattern
Activation Function: Tanh
Slope, β: 0.001
Number of Hidden Nodes: 15
Weight Update Rule: Adam
Learning Rate: 0.01

Scatterplot for Training Data

Scatterplot for Test Data

Features: Un-Normalized
Mode of Learning: Pattern
Activation Function: ReLU
Slope, β: 0.001
Number of Hidden Nodes: 15
Weight Update Rule: Adam
Learning Rate: 0.01

Optimizer: Adam Test_MSE: 95.79110292576414 Train_MSE: 96.7753495410517

Scatterplot for Training Data

Scatterplot for Test Data

Features: Un-Normalized
Mode of Learning: Batch
Slope, β: 0.001
Number of Hidden Layers: 1
Number of Hidden Nodes: 15

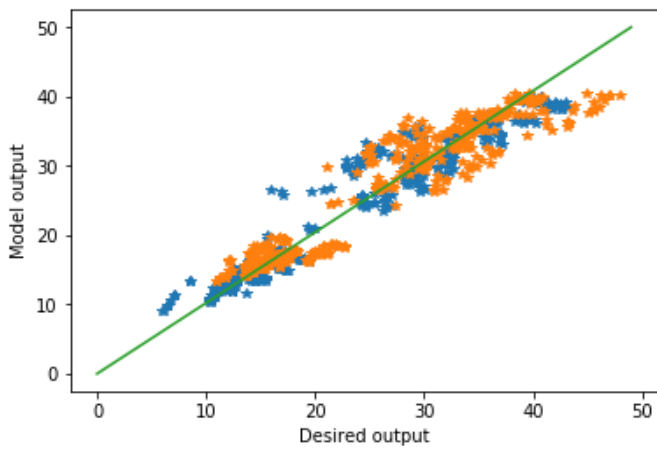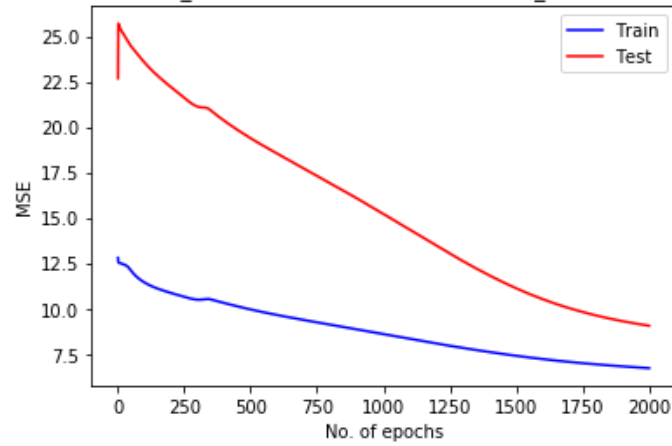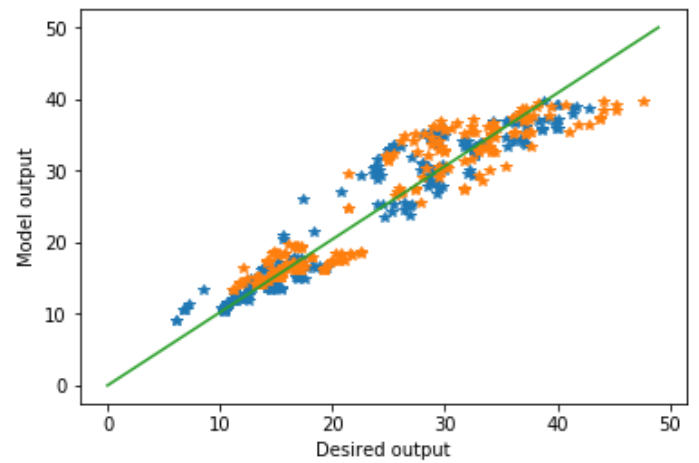| Activation Function	Weight Update Rule	Learning Rate	Test_MSE	Train_MSE	Epochs for convergence
Logistic	Delta	0.001	19.80627	21.12154	>2000
	Generalized Delta	0.001	19.66937	20.94365	870
	AdaGrad	0.001	19.61643	20.87256	249
	RMSprop	0.001	19.19128	20.21709	481
	Adam	0.001	18.16231	19.08416	>2000
Tanh	Delta	0.001	21.47732	21.95194	>2000
	Generalized Delta	0.001	21.08249	21.58669	1774
	AdaGrad	0.001	21.04316	21.54511	195
	RMSprop	0.001	21.00795	21.50677	24
	Adam	0.001	19.96502	20.46571	1299
ReLU	Delta	0.001	119.07414	126.18053	1968
	Generalized Delta	0.001	95.01602	96.76497	>2000
	AdaGrad	0.001	95.01569	96.76212	4
	RMSprop	0.001	95.01246	96.72883	20
	Adam	0.001	95.01228	96.72646	3

Features: Un-Normalized
Mode of Learning: Batch
Activation Function: Logistic
Slope, β: 0.001
Number of Hidden Nodes: 15
Weight Update Rule: Adam
Learning Rate: 0.001

Optimizer: Adam Test_MSE: 18.162313260168 Train_MSE: 19.084167774364705

Scatterplot for Training Data

Scatterplot for Test Data

Features: Un-Normalized
Mode of Learning: Batch
Activation Function: Tanh
Slope, β: 0.001
Number of Hidden Nodes: 15
Weight Update Rule: Adam
Learning Rate: 0.001

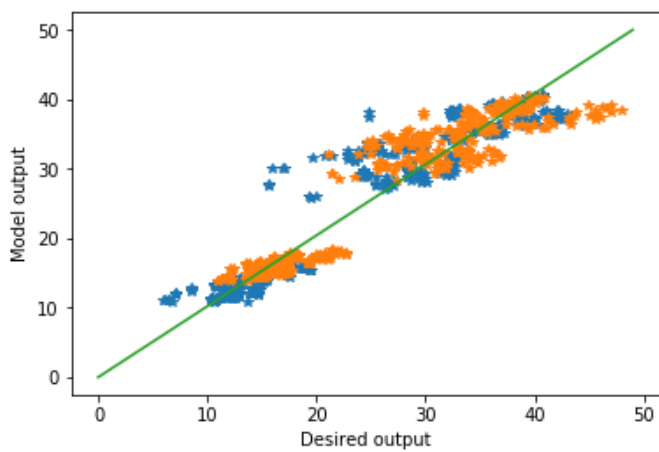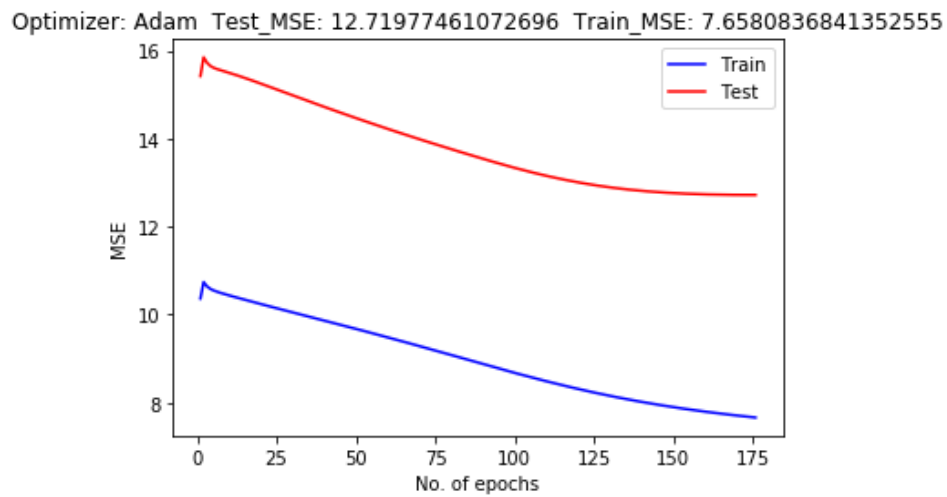Optimizer: Adam Test_MSE: 19.965026827191892 Train_MSE: 20.4657156726093
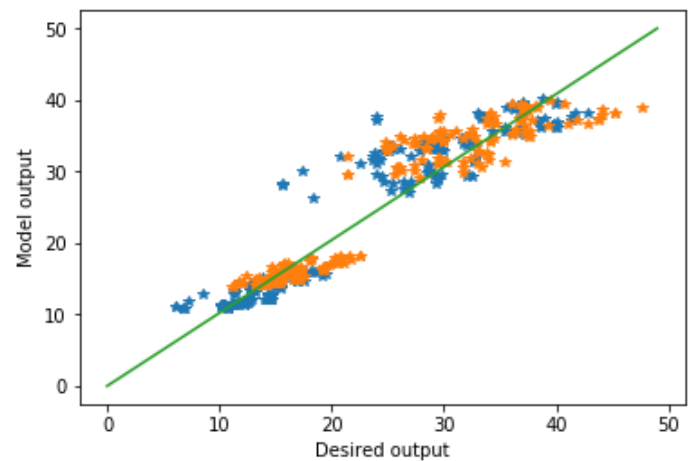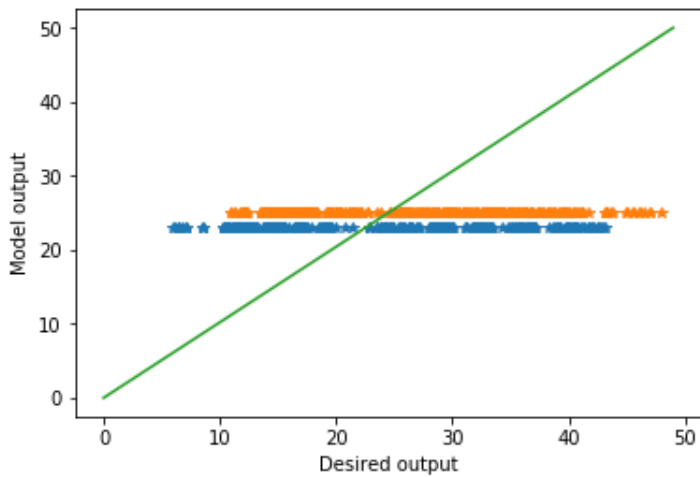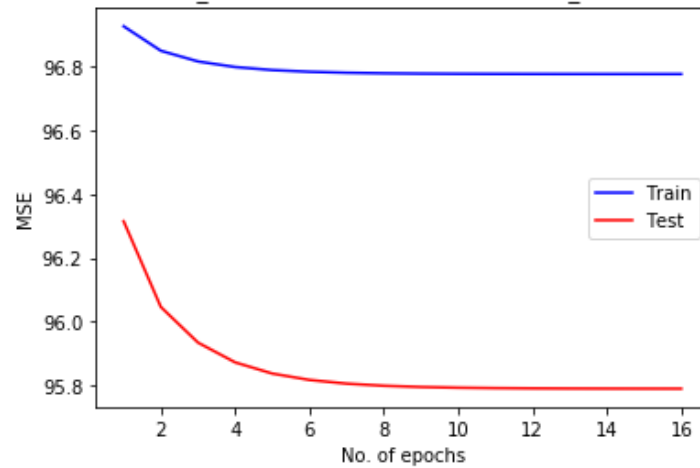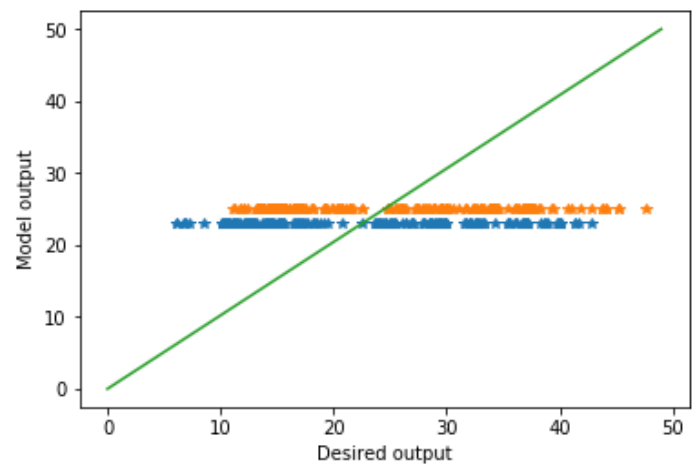

Scatterplot for Training Data

Scatterplot for Test Data

Features: Un-Normalized
Mode of Learning: Batch
Activation Function: ReLU
Slope, β: 0.001
Number of Hidden Nodes: 15
Weight Update Rule: Adam
Learning Rate: 0.001

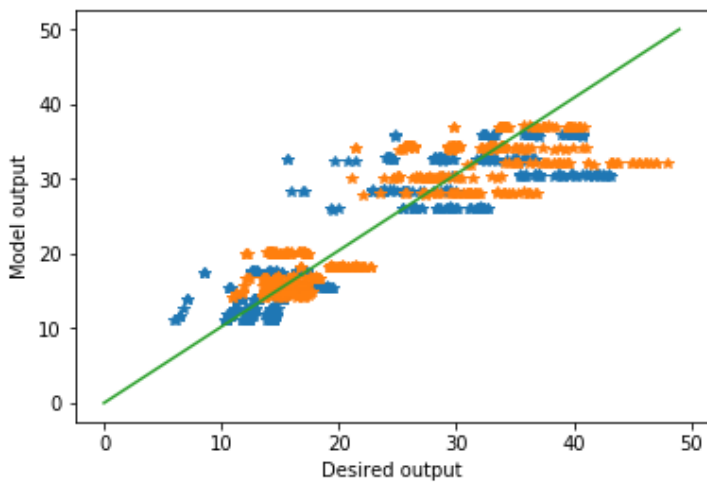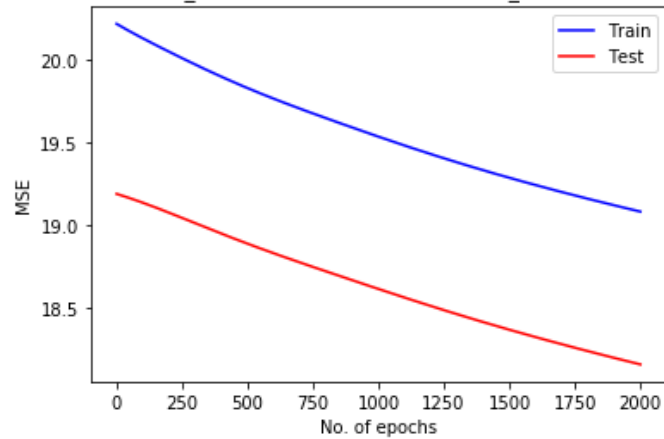Scatterplot for Training Data   Scatterplot for Test Data

Features: Normalized
Mode of Learning: Pattern
Slope, β: 1
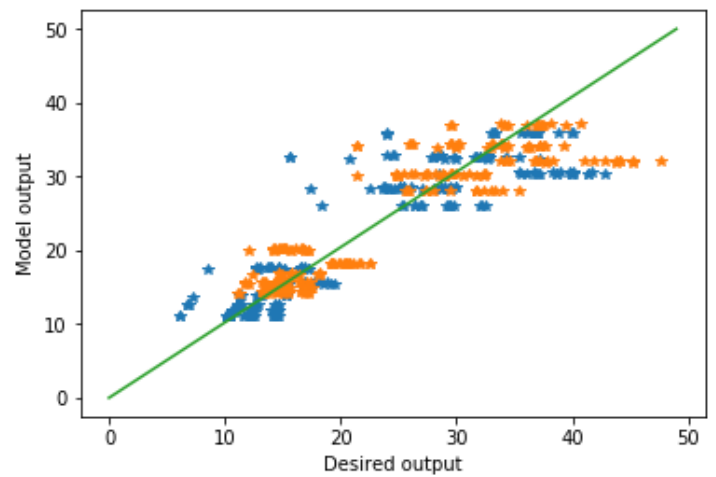Number of Hidden Layers: 1
Number of Hidden Nodes: 15

Activation Function	Weight Update Rule	Learning Rate	Test_MSE	Train_MSE	Epochs for convergence
Logistic	Delta	0.001	1.52653	1.23406	486
	Generalized Delta	0.001	1.52706	1.23434	4
	AdaGrad	0.001	1.52083	1.22170	15
	RMSprop	0.001	2.44552	1.79768	463
	Adam	0.001	1.64223	1.22669	817
Tanh	Delta	0.001	1.54061	1.19708	345
	Generalized Delta	0.001	1.54324	1.20103	7
	AdaGrad	0.001	1.52409	1.15712	10
	RMSprop	0.001	1.70952	1.28794	54
	Adam	0.001	1.47626	1.12536	134
ReLU	Delta	0.001	5.70967	3.98845	381
	Generalized Delta	0.001	5.77647	4.13417	69
	AdaGrad	0.001	4.77012	4.14011	335
	RMSprop	0.001	7.68138	4.17770	103
	Adam	0.001	4.74609	3.69558	95

Features: Normalized
Mode of Learning: Pattern
Activation Function: Logistic
Slope, β: 1
Number of Hidden Nodes: 15
Weight Update Rule: Delta
Learning Rate: 0.001

Optimizer: Delta_rule Test_MSE: 1.5265343044835764 Train_MSE: 1.2340369531341278

Scatterplot for Training Data
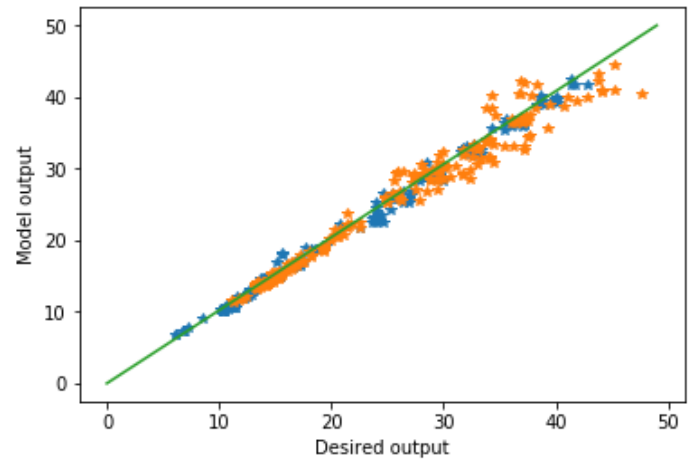
Scatterplot for Test Data

Features: Normalized
Mode of Learning: Pattern
Activation Function: Tanh
Slope, β: 1
Number of Hidden Nodes: 15
Weight Update Rule: Adam
Learning Rate: 0.001

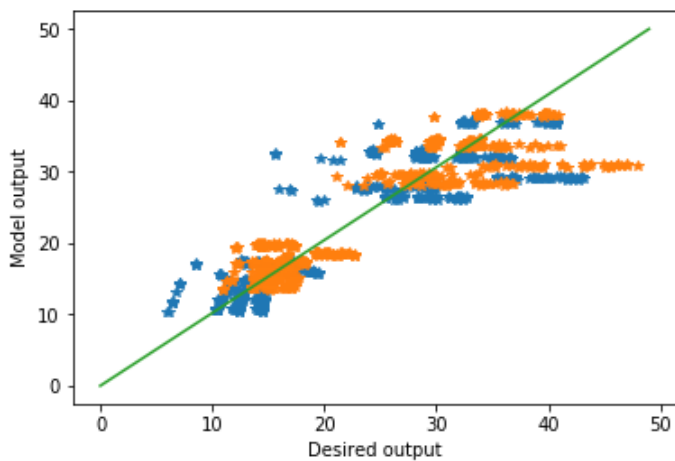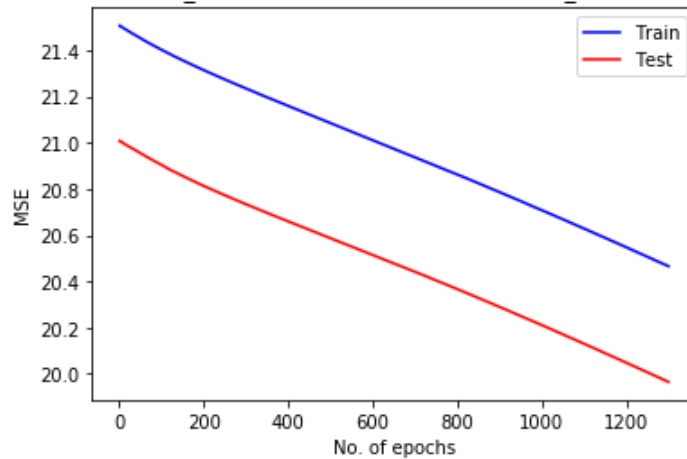Optimizer: Adam Test_MSE: 1.4762693389563957 Train_MSE: 1.125365436659699
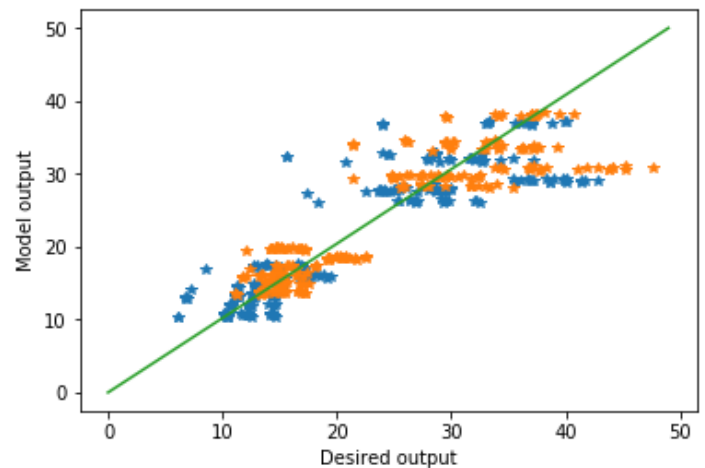

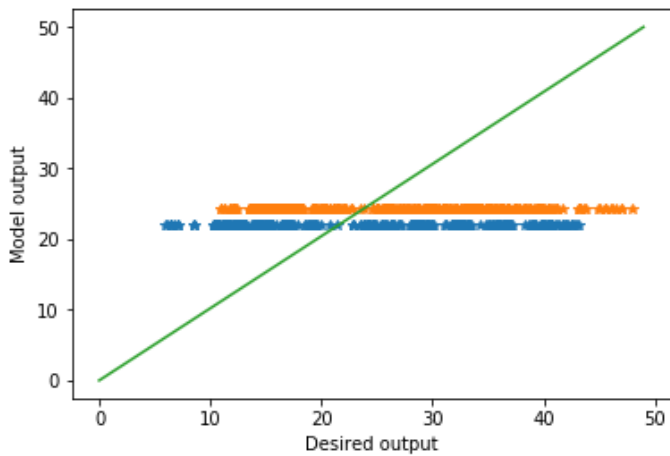Scatterplot for Training Data

Scatterplot for Test Data

Features: Normalized
Mode of Learning: Pattern
Activation Function: ReLU
Slope, β: 1
Number of Hidden Nodes: 15
Weight Update Rule: Adam
Learning Rate: 0.001

Scatterplot for Training Data

Scatterplot for Test Data

Features: Normalized
Mode of Learning: Batch
Slope, β: 1
Number of Hidden Layers: 1
Number of Hidden Nodes: 15

Activation Function	Weight Update Rule	Learning Rate	Test_MSE	Train_MSE	Epochs for convergence
Logistic	Delta	0.001	1.67183	1.37254	135
	Generalized Delta	0.001	1.67080	1.37103	11
	AdaGrad	0.001	1.65733	1.34646	82
	RMSprop	0.001	1.58755	1.28685	136
	Adam	0.001	1.56269	1.25986	129
Tanh	Delta	0.001	7.07882	5.71257	>2000
	Generalized Delta	0.001	4.41389	3.15328	>2000
	AdaGrad	0.001	3.88255	2.78382	>2000
	RMSprop	0.001	1.86238	1.4539	906
	Adam	0.001	1.77044	1.36297	180
ReLU	Delta	0.001	19.36051	12.99163	>2000
	Generalized Delta	0.001	13.76302	9.11586	>2000
	AdaGrad	0.001	13.02809	8.68157	>2000
	RMSprop	0.001	9.99247	6.56297	>2000
	Adam	0.001	9.87704	6.47658	106

Features: Normalized
Mode of Learning: Batch
Activation Function: Logistic
Slope, β: 1
Number of Hidden Nodes: 15
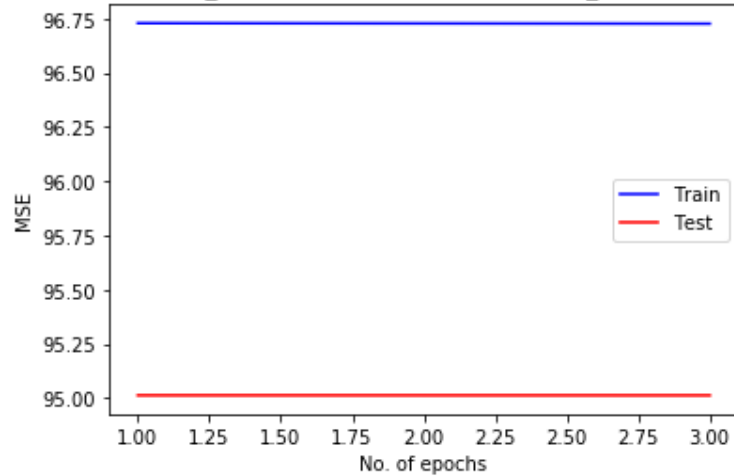Weight Update Rule: Adam
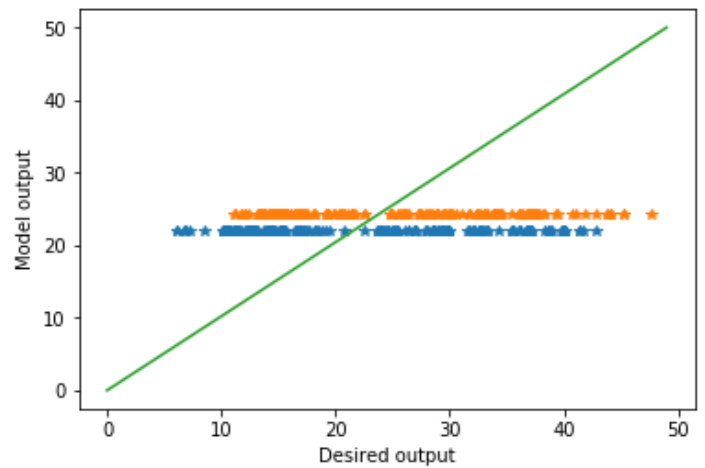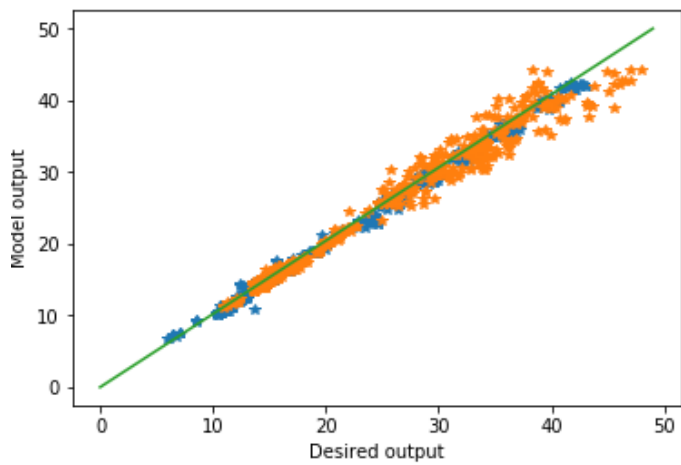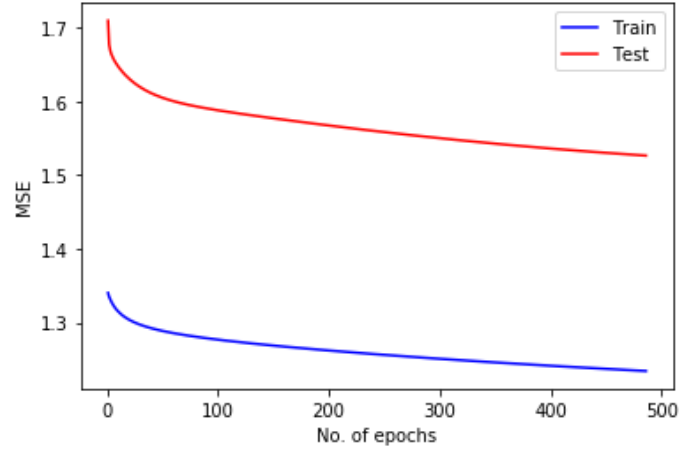Learning Rate: 0.001

Scatterplot for Training Data
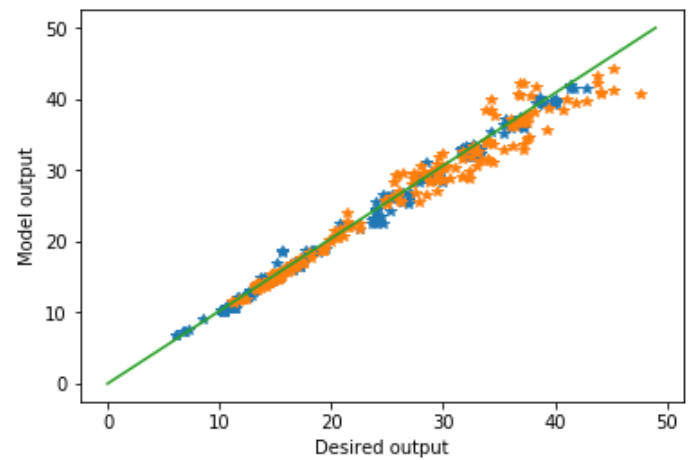
Scatterplot for Test Data

Features: Normalized
Mode of Learning: Batch
Activation Function: Tanh
Slope, β: 1
Number of Hidden Nodes: 15
Weight Update Rule: Adam
Learning Rate: 0.001

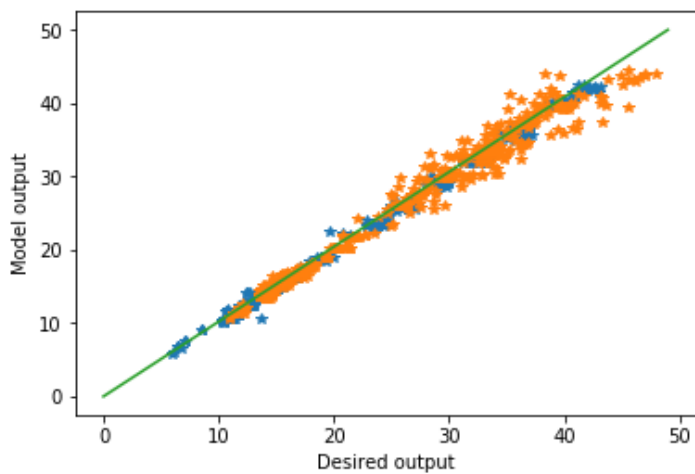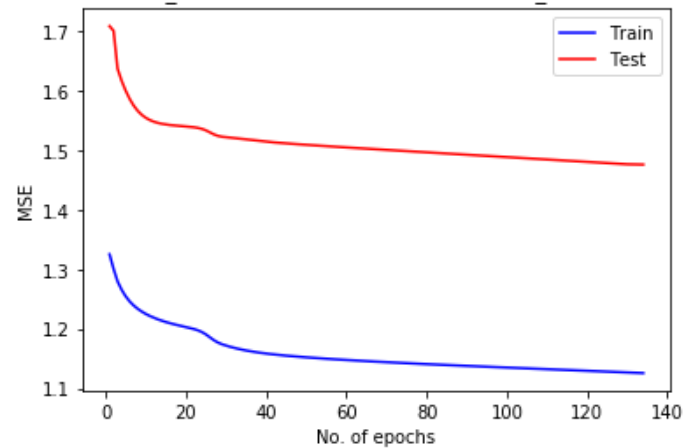Optimizer: Adam Test_MSE: 1.77044214692125 Train_MSE: 1.362970727939631
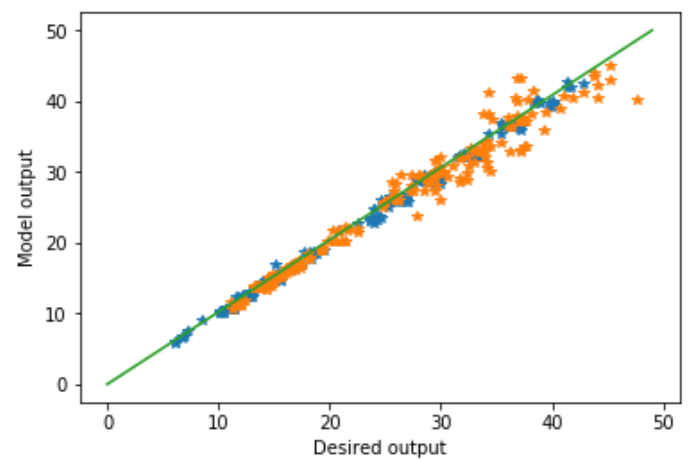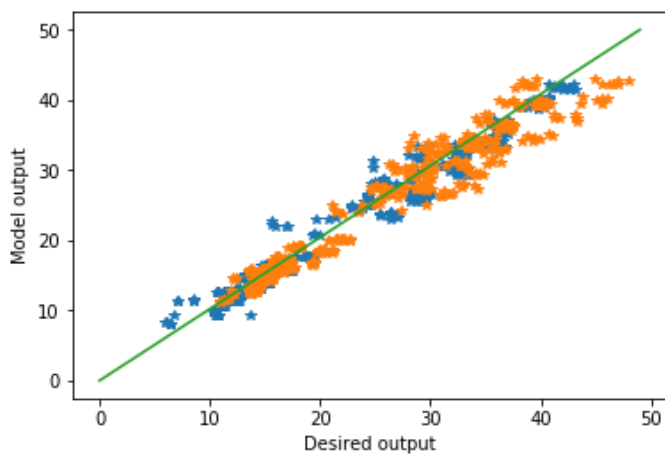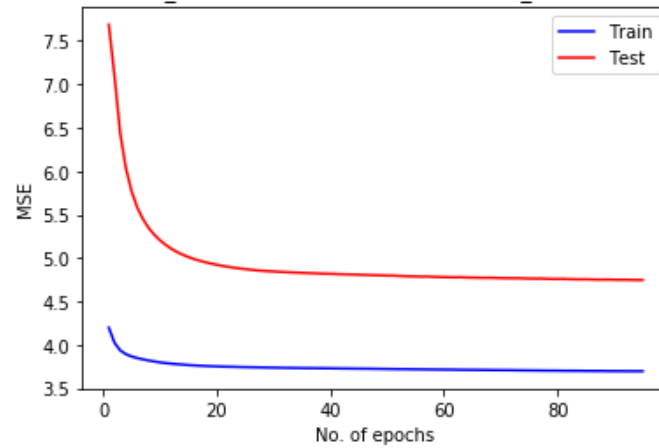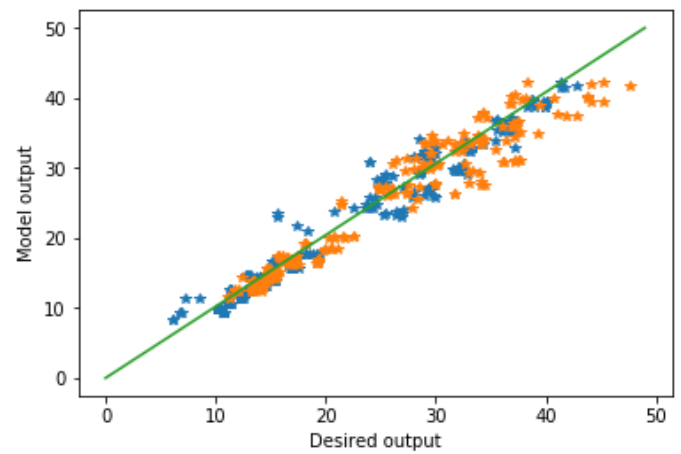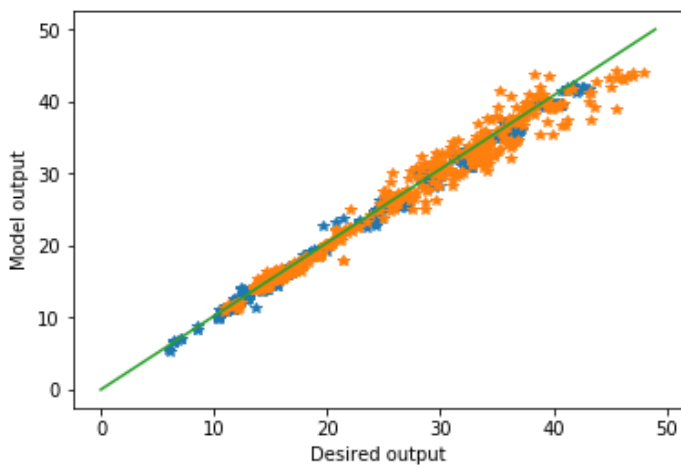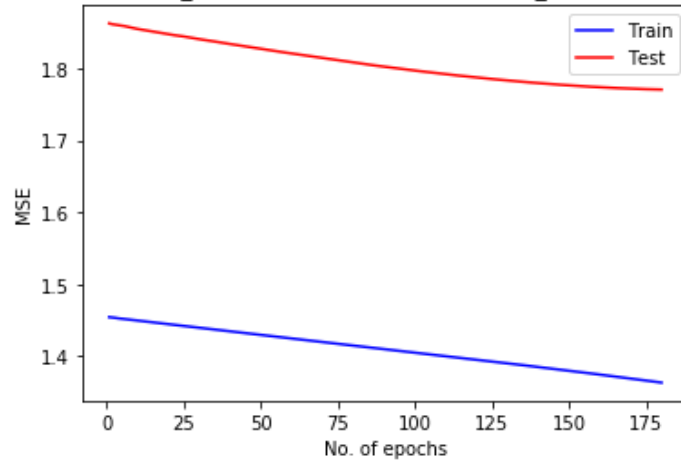

Scatterplot for Training Data
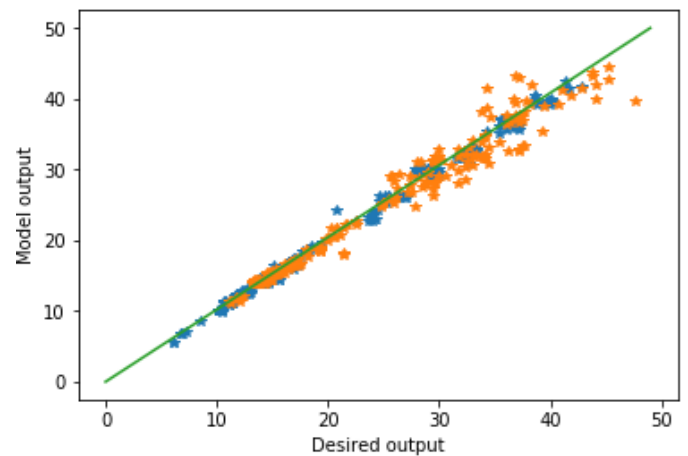
Scatterplot for Test Data

Features: Normalized
Mode of Learning: Batch
Activation Function: ReLU
Slope, β: 1
Number of Hidden Nodes: 15
Weight Update Rule: Adam
Learning Rate: 0.001

Optimizer: Adam Test_MSE: 9.877043351511027 Train_MSE: 6.476587780157697
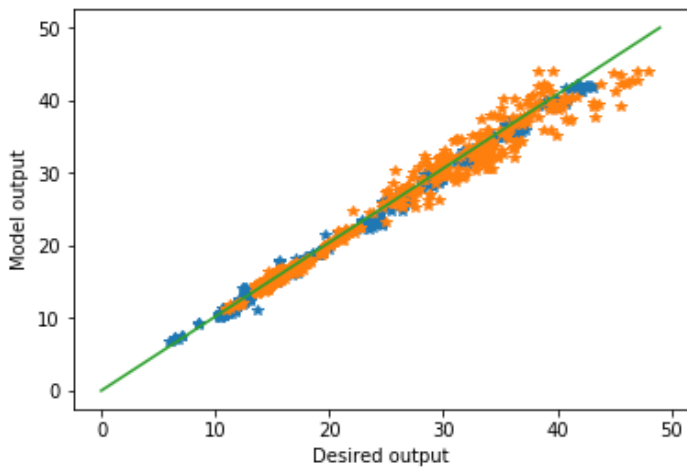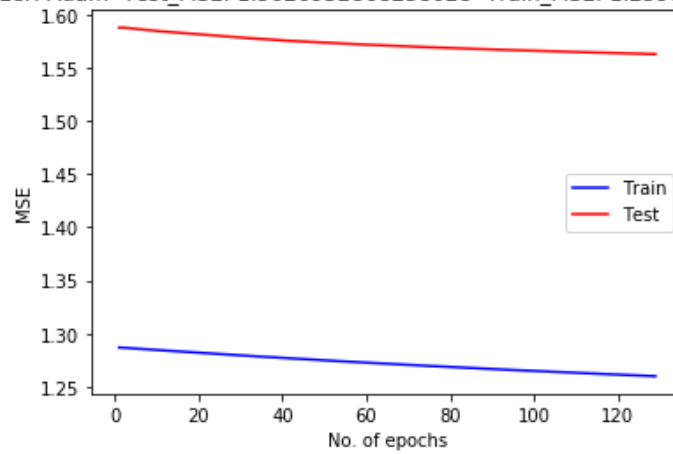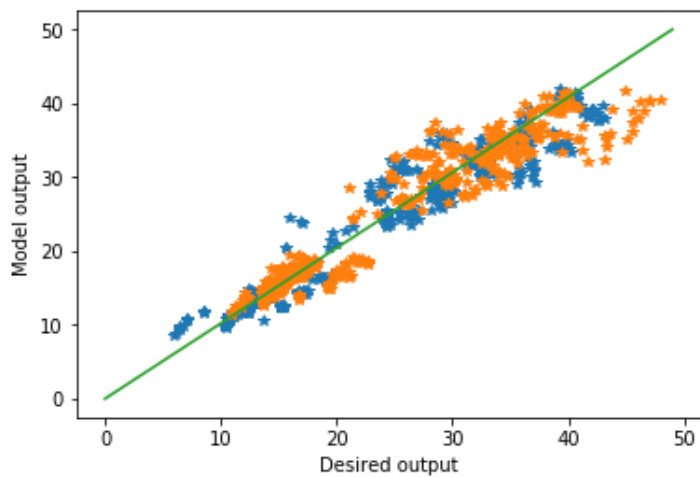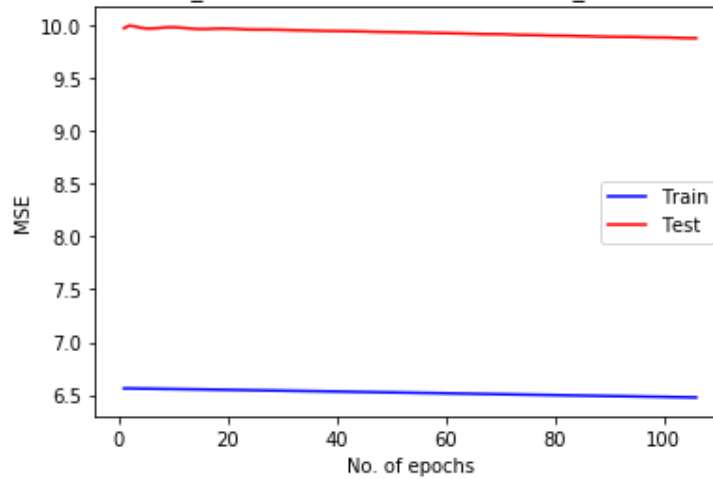

Scatterplot for Training Data
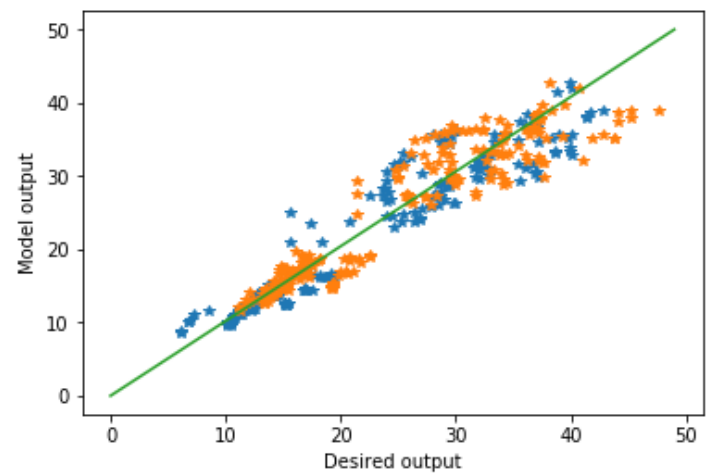
Scatterplot for Test Data

Task 2: Single Label Classification

Data:
The original data has 8 target variables encoded 0 to 7 for each class label. We have extracted only the classes and corresponding features assigned to our team.

Team 5 labels: ["coast", "street", "mountain", "forest", "inside city"]
The extracted image data has 60 features and 1662 examples. The target labels have been one-hot encoded.

Baseline System

Case 1: Multi-layer Feedforward Network with **1 hidden layer**

Features: Un-Normalized
Number of Hidden layers: 1
Mode of Learning: Pattern
Activation Function: Logistic
Activation for Output Layer: Softmax
Weight Update Rule: Delta
Learning Rate: 0.001

No. of hidden nodes	Slope, β	Train Error (cross-entropy)	Test Error (cross-entropy)	Epochs for convergence
25	0.2	0.7485	0.9894	278
	0.5	0.8282	1.0406	108
	1	0.8599	1.0867	68
	2	0.7236	1.1500	60
	3	0.8072	1.2000	55
	10	0.7500	1.3700	>2000
50	0.2	0.7258	0.9635	253
	0.5	0.7622	1.0096	109
	1	0.7333	1.0510	82
	2	0.0172	2.2840	>2000
	3	0.7500	1.2000	50
	10	0.7801	1.1060	301
100	0.2	0.7001	0.9370	226
	0.5	0.7282	0.9914	101
	1	0.7524	1.0539	54
	2	0.8016	1.1207	29
	3	0.0034	1.9755	>2000
	10	0.6853	1.1116	204

Case 2: Multi-layer Feedforward Network with **2 hidden layers**

Features: Un-Normalized
Number of Hidden layers: 2
Mode of Learning: Pattern
Activation Function: Logistic
Activation for Output Layer: Softmax
Weight Update Rule: Delta
Learning Rate: 0.001

No. of hidden nodes	Slope, β	Train Error (cross-entropy)	Test Error (cross-entropy)	Epochs for convergence
25	0.2	1.6059	1.6127	5
	0.5	0.7397	1.0664	389
	1	0.8334	1.1556	123
	2	0.8644	1.3391	64
	3	0.8500	1.2000	70
	10	0.5000	1.3000	400
50	0.2	1.6013	1.6129	296
	0.5	1.5957	1.6081	177
	1	1.4097	1.4443	>2000
	2	1.0249	1.1504	>2000
	3	0.8653	1.0994	>2000
	10	0.8000	1.2000	50
100	0.2	1.6102	1.6070	3
	0.5	0.7235	1.0218	295
	1	0.7928	1.0802	82
	2	0.0006	2.4921	>2000
	3	0.6286	1.2442	20
	10	0.5000	1.2000	100

- From the baseline system, we choose the best β, number of hidden nodes and number of hidden layers respectively which are to be used for further analysis.

$\beta = 0.5$, Number of hidden nodes = 100, Number of hidden layer = 1
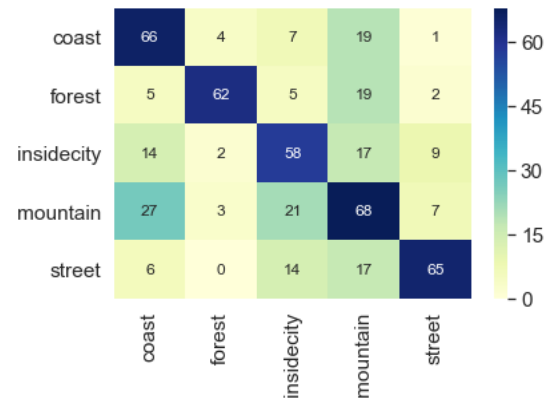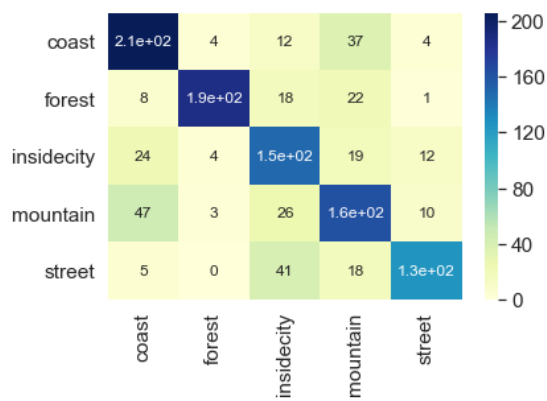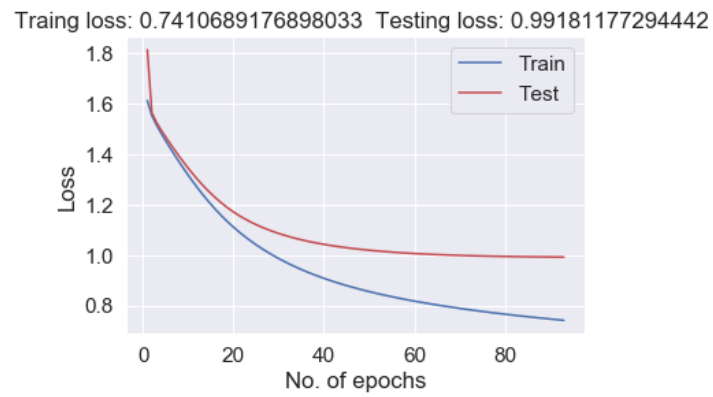
Traing loss: 0.7410689176898033 Testing loss: 0.99181177294442

Confusion Matrix for Train Data

Confusion Matrix for Test Data

Experiments

Features: Un-Normalized
Mode of Learning: Batch
Slope, β: 0.5
Number of Hidden Layers: 1
Number of Hidden Nodes: 100

Activation Function	Weight Update Rule	Learning Rate	Train Error (cross-entropy)	Test Error (cross-entropy)	Epochs for convergence
Logistic	Delta	0.0001	0.74145	0.97735	109
	Generalized Delta	0.0001	0.73966	0.97558	18
	AdaGrad	0.0001	0.72561	0.96453	35
	RMSprop	0.0001	0.71920	0.96114	8
	Adam	0.0001	0.71873	0.96111	3
Tanh	Delta	0.0001	1.13955	1.71156	>1000
	Generalized Delta	0.0001	0.99432	1.50000	>1000
	AdaGrad	0.0001	0.89069	1.3676	613
	RMSprop	0.0001	0.67462	1.09747	168
	Adam	0.0001	0.58219	1.04113	167
ReLU	Delta	0.001	0.85614	1.42818	>1000
	Generalized Delta	0.001	0.81227	1.38532	245
	AdaGrad	0.001	0.71321	1.32248	29
	RMSprop	0.0001	0.98089	1.80429	244
	Adam	0.0001	0.44432	1.54929	>1000

Features: Un-Normalized
Mode of Learning: Batch
Activation Function: Logistic
Slope, β: 0.5
Number of Hidden Nodes: 100
Weight Update Rule: Adam
Learning Rate: 0.0001

Adam Training loss: 0.7187291087741629 Testing loss: 0.9611167978294943


Confusion Matrix for Train Data

Confusion Matrix for Test Data

Features: Un-Normalized
Mode of Learning: Batch
Activation Function: Tanh
Slope, β: 0.5
Number of Hidden Nodes: 100
Weight Update Rule: Adam
Learning Rate: 0.0001

Adam Training loss: 0.5821934750921708 Testing loss: 1.0411372540739998

Confusion Matrix for Train Data

Confusion Matrix for Test Data

Features: Un-Normalized
Mode of Learning: Batch
Activation Function: ReLU
Slope, β: 0.5
Number of Hidden Nodes: 100
Weight Update Rule: Adagrad
Learning Rate: 0.001

Ada_grad Training loss: 0.7132100701237268 Testing loss: 1.322488168649481

Confusion Matrix for Train Data

Confusion Matrix for Test Data

Features: Normalized
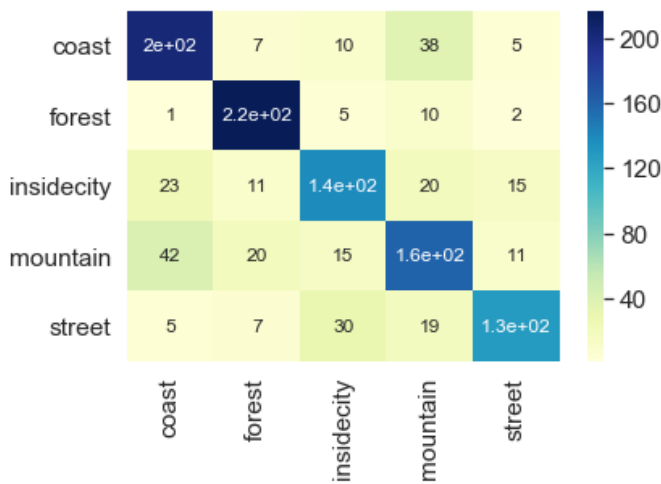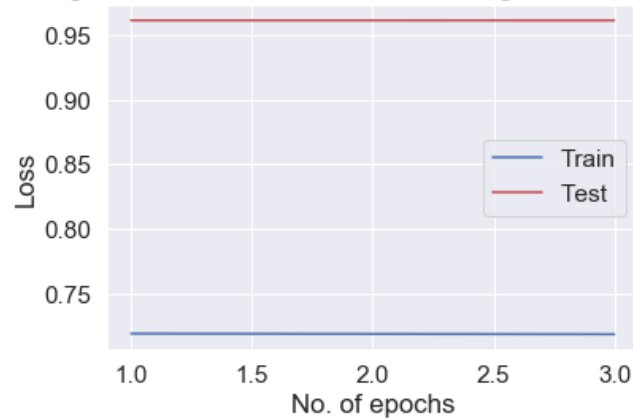Mode of Learning: Batch
Slope, β: 0.5
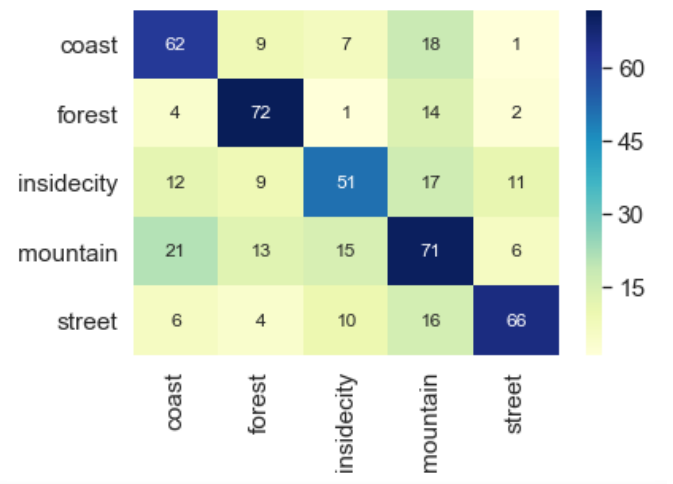Number of Hidden Layers: 1
Number of Hidden Nodes: 100

Activation Function	Weight Update Rule	Learning Rate	Train Error (cross-entropy)	Test Error (cross-entropy)	Epochs for convergence
Logistic	Delta	0.0001	0.69866	0.93539	3
	Generalized Delta	0.0001	0.69866	0.93538	3
	AdaGrad	0.0001	0.69705	0.93423	11
	RMSprop	0.0001	0.66535	0.93252	534
	Adam	0.0001	0.66436	0.93269	18
Tanh	Delta	0.0001	1.09645	1.93708	>1000
	Generalized Delta	0.0001	0.92444	1.70170	>1000
	AdaGrad	0.0001	0.87803	1.62062	>1000
	RMSprop	0.0001	0.87136	1.61260	11
	Adam	0.0001	0.59828	1.02060	955
ReLU	Delta	0.0001	0.85172	1.50753	>1000
	Generalized Delta	0.0001	0.72025	1.28554	>1000
	AdaGrad	0.0001	0.64589	1.11638	>1000
	RMSprop	0.0001	0.63072	1.09639	18
	Adam	0.0001	0.58619	1.06787	26

Features: Normalized
Mode of Learning: Batch
Activation Function: Logistic
Slope, β: 0.5
Number of Hidden Nodes: 100
Weight Update Rule: Adam
Learning Rate: 0.0001

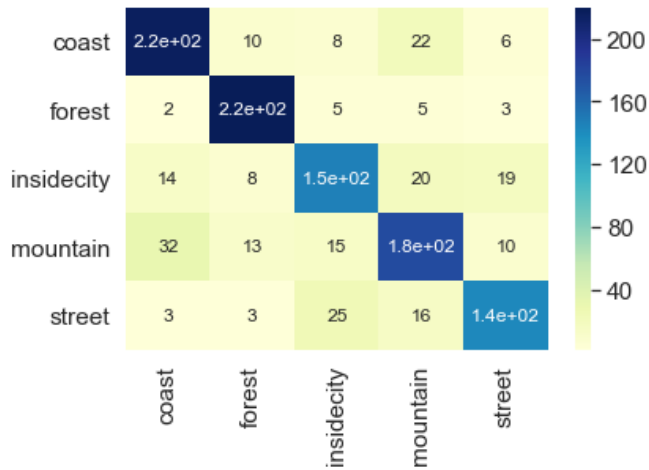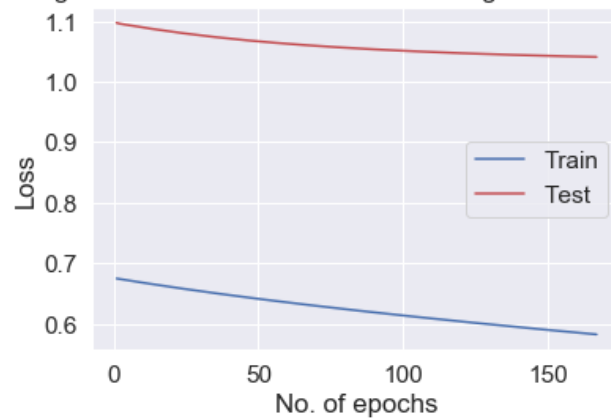Adam Training loss: 0.6643685131723785 Testing loss: 0.9326924543775059

Confusion Matrix for Train Data

Confusion Matrix for Test Data

Features: Normalized
Mode of Learning: Batch
Activation Function: Tanh
Slope, β: 0.5
Number of Hidden Nodes: 100
Weight Update Rule: Adam
Learning Rate: 0.0001

Adam Training loss: 0.5982865469110455 Testing loss: 1.0206069811593461
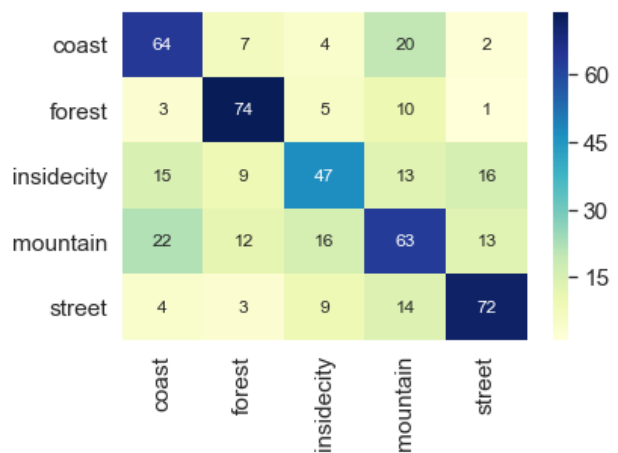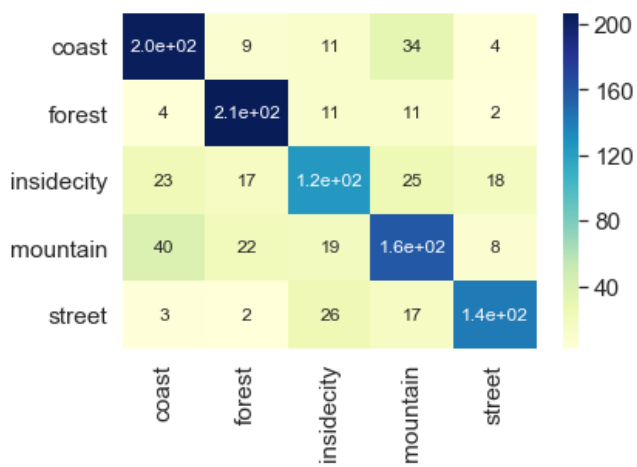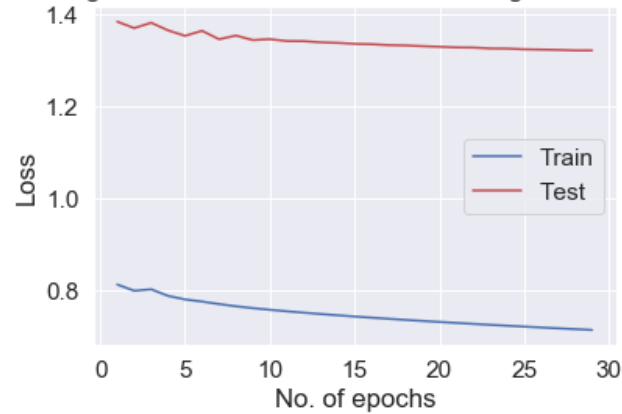

Confusion Matrix for Train Data
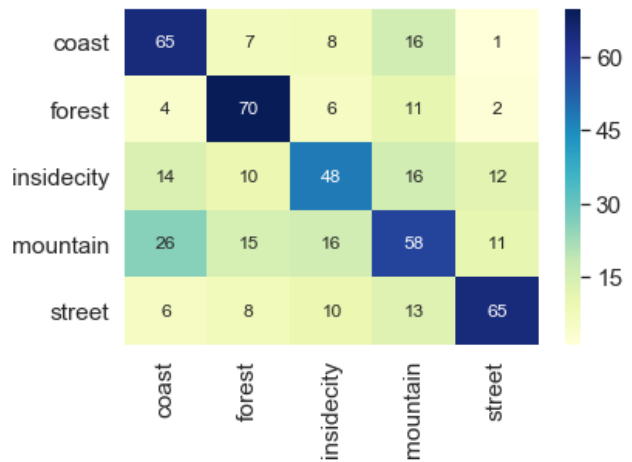
Confusion Matrix for Test Data

Features: Normalized
Mode of Learning: Batch
Activation Function: ReLU
Slope, β: 0.5
Number of Hidden Nodes: 100
Weight Update Rule: Adam
Learning Rate: 0.0001

Adam Training loss: 0.5861913880743191 Testing loss: 1.067875729321271
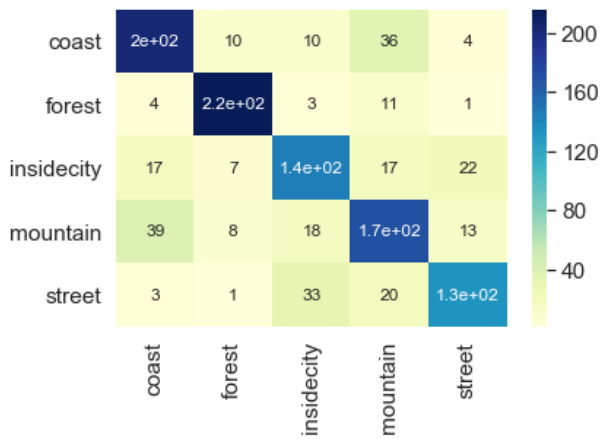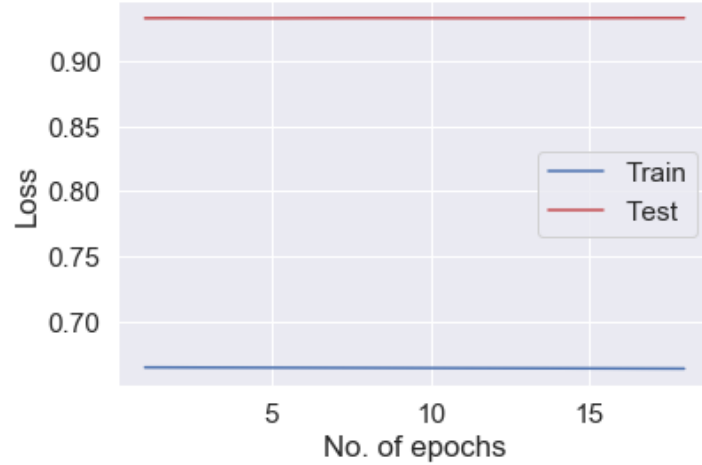

Confusion Matrix for Train Data

Confusion Matrix for Test Data

Features: Un-Normalized
Mode of Learning: Batch
Slope, β: 0.5
Number of Hidden Layers: 1
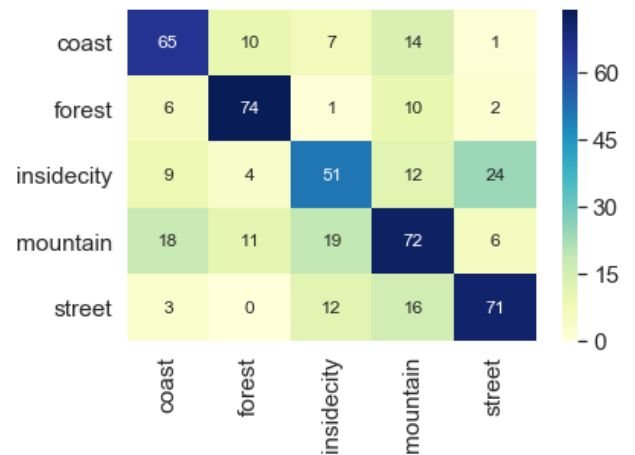Number of Hidden Nodes: 100

Activation Function	Weight Update Rule	Learning Rate	Train Error (SSE)	Test Error (SSE)	Epochs for convergence
Logistic	Delta	0.0001	0.19876	0.25379	3
	Generalized Delta	0.0001	0.19875	0.25378	3
	AdaGrad	0.0001	0.19403	0.25013	19
	RMSprop	0.0001	0.18878	0.24649	17
	Adam	0.0001	0.18893	0.24667	22
Tanh	Delta	0.0001	0.25150	0.33340	361
	Generalized Delta	0.0001	0.25049	0.33249	337
	AdaGrad	0.0001	0.60388	0.67571	221
	RMSprop	0.0001	0.23898	0.32308	276
	Adam	0.0001	0.24616	0.32624	339
ReLU	Delta	0.0001	0.27630	0.34713	70
	Generalized Delta	0.0001	0.27595	0.34702	65
	AdaGrad	0.001	0.31969	0.36941	158
	RMSprop	0.0001	0.26380	0.34053	63
	Adam	0.0001	0.26627	0.34175	70

Features: Un-Normalized
Mode of Learning: Batch
Activation Function: Logistic
Slope, β: 0.5
Number of Hidden Nodes: 100
Weight Update Rule: RMS_prop
Learning Rate: 0.0001

Training loss: 0.18878467926414147 Testing loss: 0.24649885903410995
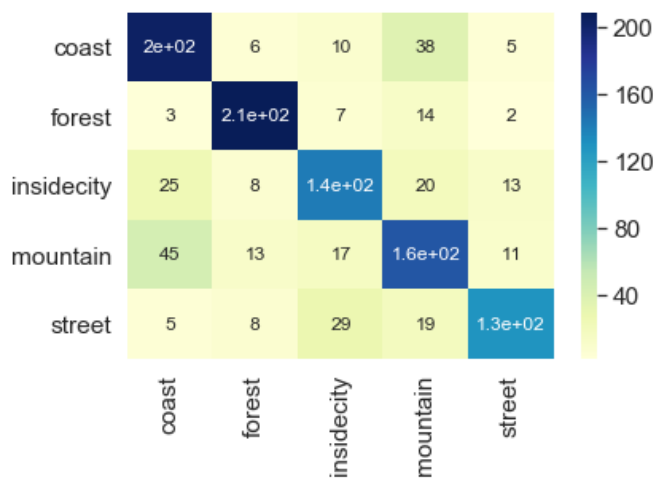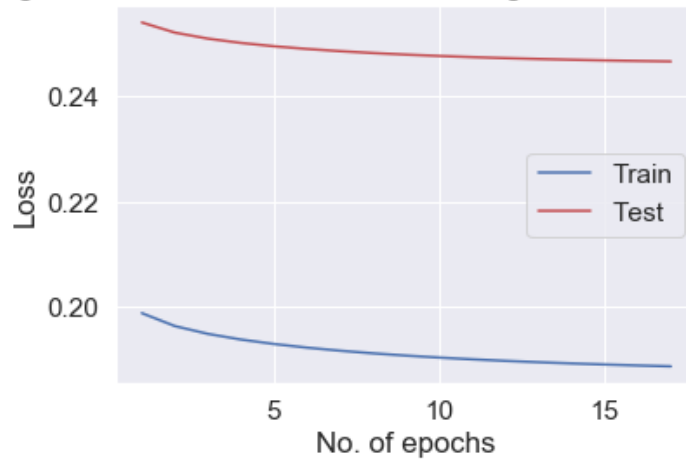
Confusion Matrix for Train Data
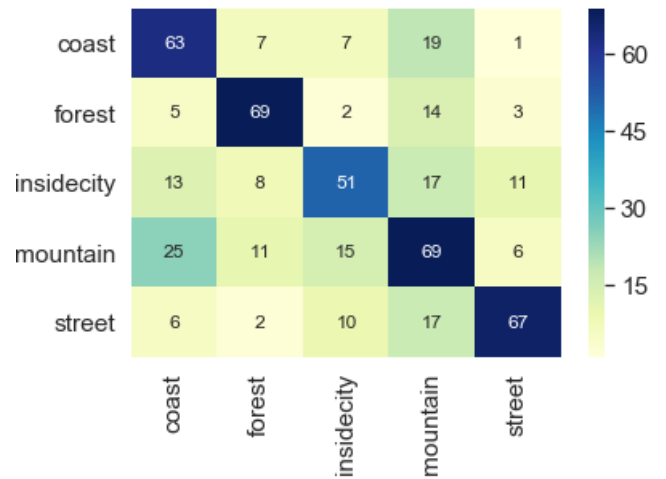
Confusion Matrix for Test Data

Features: Un-Normalized
Mode of Learning: Batch
Activation Function: Tanh
Slope, β: 0.5
Number of Hidden Nodes: 100
Weight Update Rule: RMS_prop
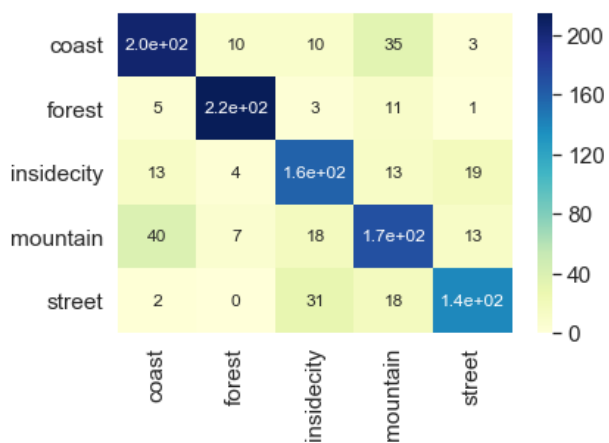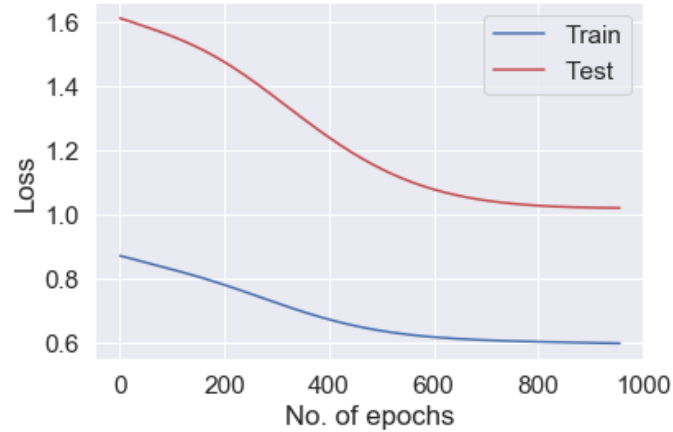Learning Rate: 0.0001
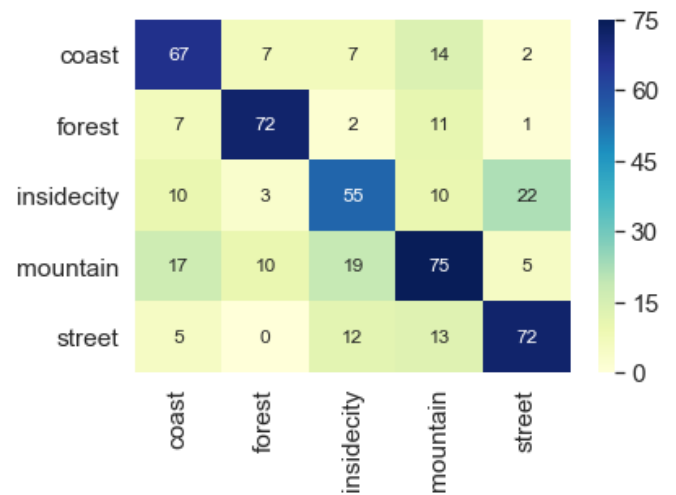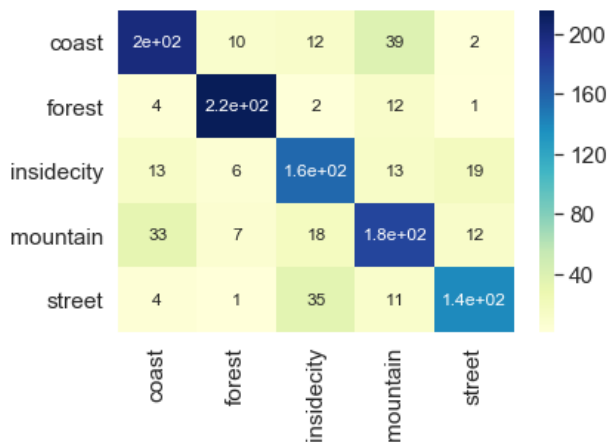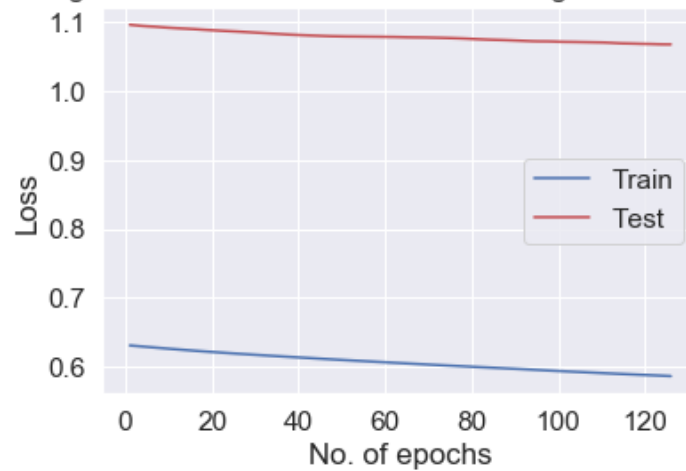

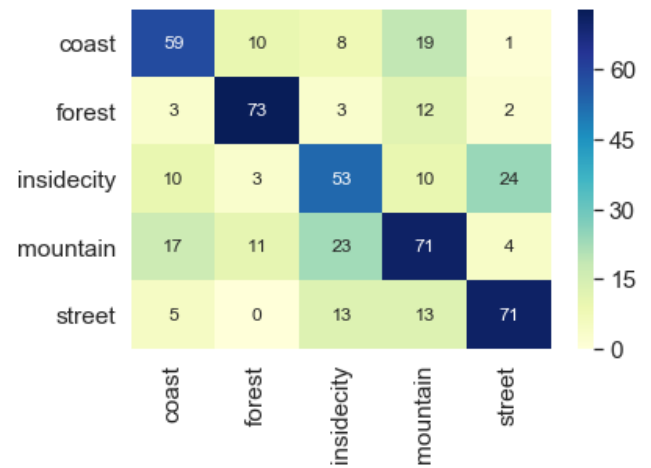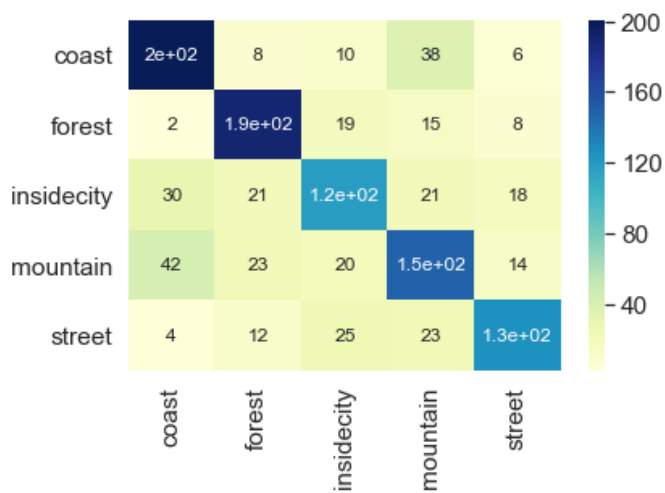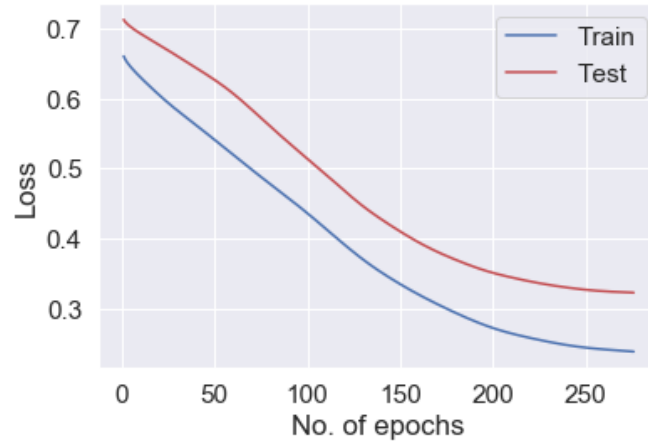Confusion Matrix for Train Data
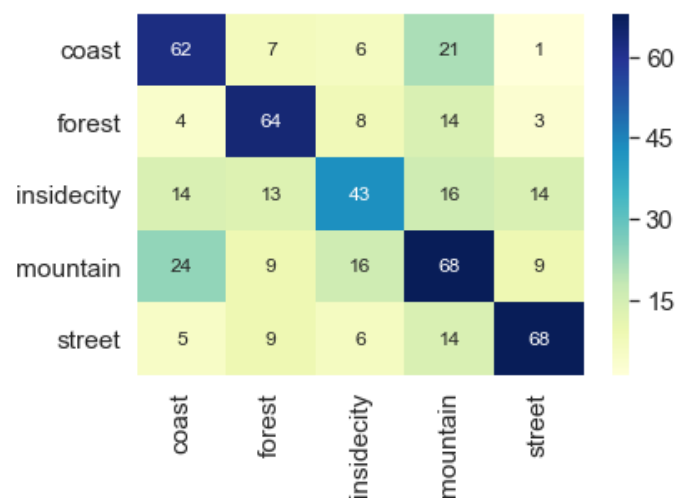
Confusion Matrix for Test Data

Features: Un-Normalized
Mode of Learning: Batch
Activation Function: ReLU
Slope, β: 0.5
Number of Hidden Nodes: 100
Weight Update Rule: RMS_prop
Learning Rate: 0.0001

Training loss: 0.2638048892196105 Testing loss: 0.3405338006569516
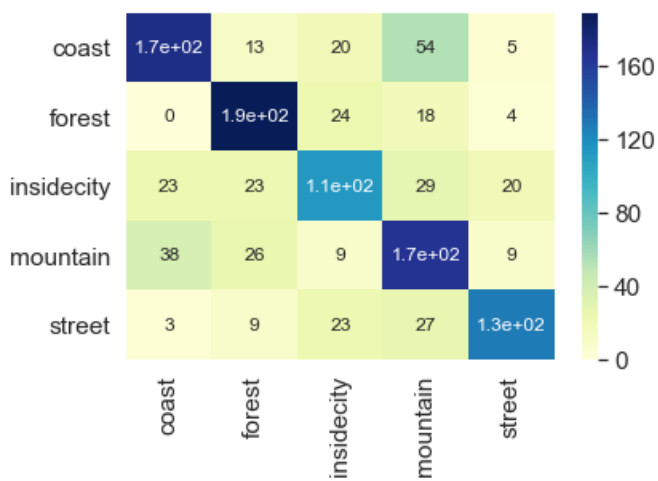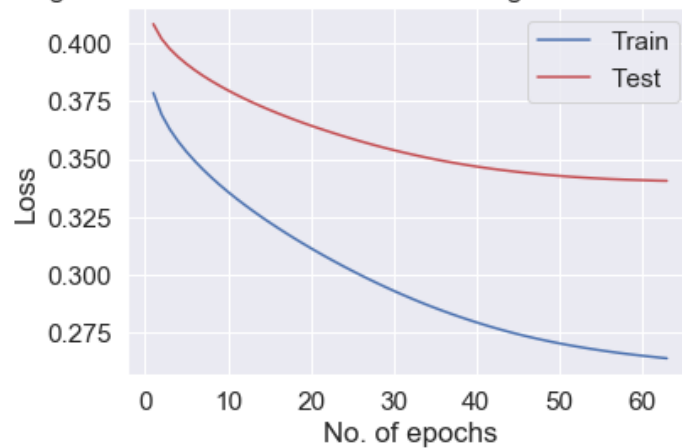
Confusion Matrix for Train Data
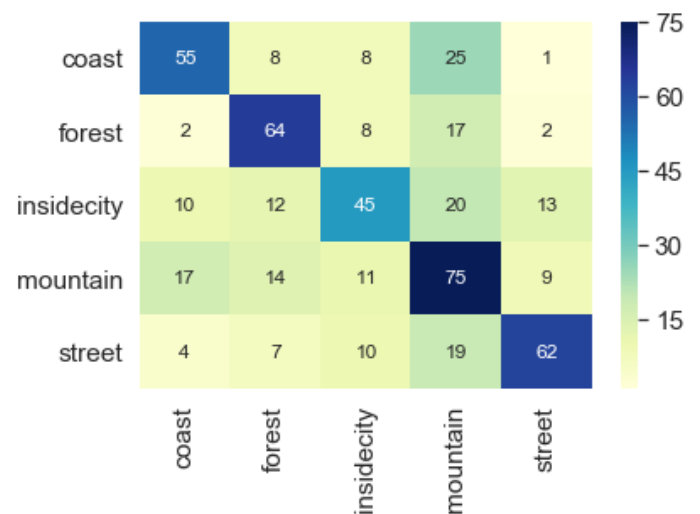
Confusion Matrix for Test Data

Task 3: Multi - Label Classification

Data:

The image data has 32 features and 9999 examples.It has 6 target variables "flowers","sunset","window","buildings","sky","person" encoded o's and 1's.

BASELINE

Number of hidden layers : **2**
Number of neurons in each hidden layer: **[30,50,100]**
Mode of learning**: Pattern**
Features: **Un-Normalized**
Activation Function: **Logistic**
Activation for Output Layer: **Logistic**
Weight Update Rule: **Delta**
Learning Rate: **0.001**

Small(30):

Slope, β	Train Error (SSE)	Test Error(SSE)	Epochs for convergence
0.01	0.4991	0.4998	14
0.2	0.4991	0.4998	14
1	0.3989	0.3996	930
5	0.3704	0.3807	398
10	0.3680	0.3852	162
20	0.3770	0.3936	71

Medium(50):

Slope, β	Train Error (SSE)	Test Error(SSE)	Epochs for convergence
0.01	0.4991	0.4998	11
0.2	0.4991	0.4998	11
1	0.3992	0.4001	1074
5	0.3714	0.3815	360
10	0.3655	0.3828	150
20	0.3658	0.3907	71

Large(100):

Slope, β	Train Error (SSE)	Test Error(SSE)	Epochs for convergence
0.01	0.4993	0.4999	3
0.2	0.4993	0.4999	3
1	0.3997	0.4001	964
5	0.3690	0.3792	396
10	0.3591	0.3802	150
20	0.3659	0.3888	48

From the baseline system, we choose the best β to be 20


```
Number of epochs for convergence = 48

Traing loss: 0.3659790065086011  Testing loss0.3888719164682771
```
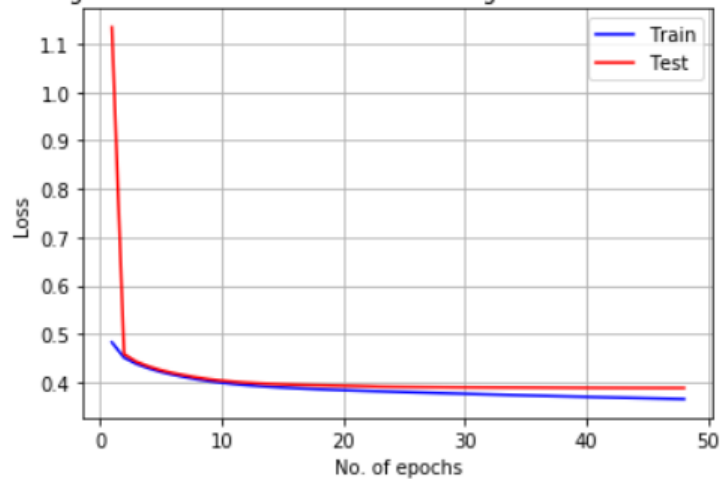
Training:

```
precision =  0.7666284950854089
recall =  0.6414853516556922
F_measure = 0.698496007119455
```

Testing:

```
precision =  0.7414537006306007
recall =  0.6140059741121805
F_measure = 0.6717381714821798
```

Experiments

Features: Un-Normalized
Mode of Learning: Batch

Activation Function	Weight Update Rule	Learning Rate	Train Error (SSE)	Test Error (SSE)	Epochs for convergence
Logistic	Delta	0.0001	0.3623	0.3862	3
	Generalized Delta	0.0001	0.3623	0.3862	3
	AdaGrad	0.0001	0.3606	0.3856	5
	RMSprop	0.0001	0.3601	0.3855	5
	Adam	0.0001	0.3716	0.3880	11
Tanh	Delta	0.0001	0.6244	0.6245	>1000
	Generalized Delta	0.0001	0.5917	0.5917	>1000
	AdaGrad	0.0001	0.4814	0.4817	>1000
	RMSprop	0.0001	0.3957	0.3965	313
	Adam	0.0001	0.3905	0.3927	210

<u>Logistic & Adam</u>

Training :

```
precision =  0.7723303750357858
recall =  0.627469224162611
F_measure = 0.6924041719949621
```

Testing:

```
precision =  0.7471512335435335
recall =  0.6057362540103994
F_measure = 0.669052812668381
```

<u>Tanh & Adam</u>

Training:

```
precision =  0.7562983109075293
recall =  0.5710229983777078
F_measure = 0.650729745904512
```

Testing:

```
precision =  0.755393295718553
recall =  0.5708319504369953
F_measure = 0.6502705774784766
```

Features: Normalized
Mode of Learning: Batch

Activation Function	Weight Update Rule	Learning Rate	Train Error (SSE)	Test Error (SSE)	Epochs for convergence
Logistic	Delta	0.0001	0.4937	0.4947	3
	Generalized Delta	0.0001	0.4936	0.4947	3
	AdaGrad	0.0001	0.4572	0.4599	>1000
	RMSprop	0.0001	0.4117	0.4169	188
	Adam	0.0001	0.3585	0.3875	514
Tanh	Delta	0.0001	0.6510	0.6539	>1000
	Generalized Delta	0.0001	0.5852	0.5875	>1000
	AdaGrad	0.0001	0.4981	0.4994	736
	RMSprop	0.0001	0.4051	0.4059	521
	Adam	0.0001	0.3919	0.3943	393

Logistic & RMSprop:

Training :

```
precision = 0.7672417856974497
recall = 0.6302503182011031
F_measure = 0.6920316447213394
```

Testing:

```
precision = 0.7362818761384334
recall = 0.6125056921675773
F_measure = 0.6687144080679971
```

Tanh & Adam:

Training:

```
precision = 0.7599113751001745
recall = 0.5945882242021402
F_measure = 0.6671605592272649
```

Testing:

```
precision = 0.7423155737704918
recall = 0.5894808743169399
F_measure = 0.6571286987192557
```

Features: Un-Normalized
Mode of Learning: Pattern

Activation Function	Weight Update Rule	Learning Rate	Train Error (SSE)	Test Error (SSE)	Epochs for convergence
Logistic	Delta	0.0001	0.3608	0.3860	21
	Generalized Delta	0.0001	0.3606	0.3860	3
	AdaGrad	0.0001	0.3596	0.3855	6
	RMSprop	0.0001	0.3207	0.4536	>100
	Adam	0.0001	0.3038	0.4543	4
Tanh	Delta	0.0001	0.3901	0.3927	3
	Generalized Delta	0.0001	0.3901	0.3927	3
	AdaGrad	0.0001	0.3900	0.3926	4
	RMSprop	0.0001	0.3935	0.3938	10
	Adam	0.0001	0.3686	0.3819	31

Logistic & AdaGrad:

Training:

```
precision =  0.7640951303446
recall =  0.6432494225239239
F_measure = 0.698483893437037
```

Testing:

```
precision =  0.7300204918032787
recall =  0.6256545992714025
F_measure = 0.6738202704558384
```

Tanh & Adam

Training:

```
precision =  0.755267996040164
recall =  0.5932329232074671
F_measure = 0.6645154403690471
```

Testing:

```
precision =  0.7395264116575591
recall =  0.5858948087431693
F_measure = 0.653806773046275
```

Features: Normalized
Mode of Learning: Pattern

Activation Function	Weight Update Rule	Learning Rate	Train Error (SSE)	Test Error (SSE)	Epochs for convergence
Logistic	Delta	0.0001	0.3671	0.3915	>100
	Generalized Delta	0.0001	0.3623	0.3906	60
	AdaGrad	0.0001	0.3606	0.3905	3
	RMSprop	0.0001	0.3833	0.4255	70
	Adam	0.0001	0.2664	0.4613	>100
Tanh	Delta	0.0001	0.4005	0.4029	80
	Generalized Delta	0.0001	0.4001	0.4025	4
	AdaGrad	0.0001	0.3996	0.4023	3
	RMSprop	0.0001	0.3983	0.3974	16
	Adam	0.0001	0.3982	0.3962	8

Logistic & AdaGrad:

Training:

```
precision = 0.7677603356432376
recall = 0.628635742233536
F_measure = 0.6912674650138149
```

Testing:

```
precision = 0.7381602914389799
recall = 0.6132172131147541
F_measure = 0.6699128781156759
```

Tanh & RMSprop

Training:

```
precision = 0.752557394050818
recall = 0.6006811860651486
F_measure = 0.6680966308273658
```

Testing:

```
precision = 0.7353142076502732
recall = 0.5947176684881602
F_measure = 0.657584767742023
```