**CHAPTER 4**

**EXPERIMENTAL RESULTS AND COMPARATIVE ANALYSIS**

This chapter list out all experiments, results and comparative analysis related to the proposed framework. Each component will be defined as a milestone in the proposed model. In fact, each of these components contains comparison with another approach to highlight the improvements of the proposed approach.

Section 42 will present about datasets used in this experiment. Section 5.3 will explain the experiment as well as comparative analysis of the controlled environment.

comparative analysis are presented in Section 5.4. Next, Section 5.5 will describe the implementation results, accuracy measurement and comparative analysis. The evaluation and analysis result on malware classification are presented in Section 5.6. Section 5.7 discuss about the proposed framework by highlighting overall milestone of the components. Lastly, Section 5.8 summarizes the chapter.

**4.1 IMPLEMENTATION PHASE**

SIMULATION

MATLAB

REASON TO CHOOSE MATLAB

COMPONENT OF MATLAB FOR CLOUD COMPUTING

SIMULATION DETAILS

Compare the algorithm privileges with other respected scholar works

Simulation Parameters and Performance Metrics

**4.1.1 REASON TO CHOOSE MATLAB**

MATLAB has several advantages over other methods or languages:

* Its basic data element is the matrix. A simple integer is considered an matrix of one row and one column. Several mathematical operations that work on arrays or matrices are built-in to the Matlab environment. For example, cross-products, dot-products, determinants, inverse matrices.
* Vectorized operations. Adding two arrays together needs only one command, instead of a for or while loop.
* The graphical output is optimized for interaction. You can plot your data very easily, and then change colors, sizes, scales, etc, by using the graphical interactive tools.
* Matlab’s functionality can be greatly expanded by the addition of toolboxes. These are sets of specific functions that provided more specialized functionality. Ex: Excel link allows data to be written in a format recognized by Excel, Statistics Toolbox allows more specialized statistical manipulation of data (Anova, Basic Fits, etc)

There are also disadvantages:

* It uses a large amount of memory and on slow computers it is very hard to use.
* It sits “on top” of Windows, getting as much CPU time as Windows allows it to have. This makes real-time applications very complicated.

**security tool performance parameter**

**Cloud Computing**

Cloud computing is the way to use computing resources as an utility. Both hardware and software can be delivered as a service over a network (typically the Internet).According to Ref. [43], there are two types of cloud: public cloud and private cloud.

building cloud computing

A public cloud is designed to apply a pay-as-you-go manner to the general public to provide services. And a private cloud is usually used to deal with the inside data of an organization which are not open to the public. For example, Ref. shows that the IBM smart cloud can provide private cloud service by giving threat protection for every layer of virtual infrastructure, limiting access to critical data, tracking user access and getting virtual infrastructure reports.

## 4.3.3 CLOUD WEB SERVER

The Java platform has turned into a pillar of big business IT improvement since the presentation of the Enterprise Edition in 1998, in two unique routes: Through the coupling of Java to the cloud web server, the Java platform has turned into a main platform for coordinating the Web with big business backend frameworks. This has permitted organizations to move part or the majority of their business to the Internet environment by method for very intuitive online situations.

The Java platform has developed into an Enterprise Integration part in which legacy frameworks are opened to the outside world through extensions based on the Java stage. This pattern has been bolstered for Java stage support for EAI norms like informing and Web benefits and has powered the incorporation of the Java platform as an advancement premise in such measures as SCA, XAM and others.These platforms shown in table 4.3

Table 4.3: Software tools and platforms

|  |  |  |  |
| --- | --- | --- | --- |
| **Component** | **JBoss EWS 2.0.0** | **JBoss EWS 2.0.1** | **JBoss EWS 2.1.0** |
| Apache Web Server (httpd) | 2.2.22 | 2.2.22 | 2.2.26 |
| Apache Tomcat 7 | 7.0.30 | 7.0.40 | 7.0.54 |
| Apache Tomcat 6 | 6.0.35 | 6.0.37 | 6.0.41 |
| Apache Tomcat Native | 1.1.24 | 1.1.27 | 1.1.30 |
| Mod.jk | 1.2.36 | 1.2.37 | 1.2.40 |
| Mod cluster | 1.2.3.Final | 1.2.4.Final | 1.2.9.Final |

As a major aspect of the JBoss Enterprise Web Server (JBoss EWS) discharge handle, a few segments may be included or evacuated for reliance or similarity purposes. Also, a few segments may be incorporated as innovation sneak peaks. The following are the individual segment subtle elements for each JBoss EWS discharge.

## 4.3.4 JBOSS ENTERPRISE WEBSERVER 2

The following components have been integrated into JBoss EWS 2.0 and are fully supported for use in both development and production (according to your corresponding subscription agreement). The following components have been tested with JBoss EWS 2.0 and are fully supported for use in both development and production with a subscription to EWS Plus (according to your corresponding subscription agreement).

# 4.3.5 TOMCAT

Apache Tomcat, regularly alluded to as Tomcat, is an open-source web server created by the Apache Software Foundation (ASF). Tomcat executes a few Java EE determinations including Java Servlet, JavaServer Pages (JSP), Java EL, and WebSocket, and gives an "immaculate Java" HTTP web server environment for Java code to keep running in. Tomcat is created and kept up by an open group of designers under the support of the Apache Software Foundation, discharged under the Apache License 2.0 permit, and is open-source programming. Tomcat began off as a servlet reference usage by James Duncan Davidson, a product designer at Sun Microsystems. He later made the undertaking open source and assumed a key part in its gift by Sun Microsystems to the Apache Software Foundation. The Apache Ant programming fabricate robotization device was produced as a side-

Impact of the formation of Tomcat as an open source project.Davidson had at first trusted that the undertaking would get to be publicly released and, since numerous open source ventures had O'Reilly books connected with them highlighting a creature on the spread, he needed to name the task after a creature. Tomcat was discharged with four parts: Catalina (servlet holder), cote (HTTP connector) Jaspera (JSP motor). Catalina is tomcat servlet holder, it actualizes Sun Microsystem determinations for servlet and Java server pages. In tomcat a domain component speaks to a Database of username, secret word and moves like UNIX gatherings relegated to those clients. Distinctive executions a domain permit Catalina to be incorporated into situations where such validation data is as of now been made and kept up and after that utilization that data to actualize contain an administration security as portrayed in the servlet determination

Coyote is Tomcats HTP connector component. It support HTTP1.1 protocol for the web cloud server application container. Onyote listens for incoming connections on a specific TCP port the server and forwards the resource to the Tomcat Engine. The Tomcat in process the request and sends back a response to requesting client. Coytote can execute the JSPs and servlets

Jesper is Tomcats JSP engine and JSP files to compile them into Java codes as servlets. The compile Java code can be handled by Catlina and runtime Jasper detects changes to JSP files and recomple slim. Jesper recompiles JSP files. As version 5 tomecat uses Jesper 2. Who is an implemntaion of Sun Microsystem is Jay sp2 specification.

**4.5 DEPLOYMENT OPTIONS**

Converting your Matlab code to a web service makes it possible that this code is used anywhere on the world. Your software can be called from a HTTP server, directltyfrom a browser, or by another computer program. Apart from this interoperability advantage, making software available as a webservice may help you comply withcorporate IT requirements, allows you to protect sensitive data, or simply increase computing capacity by running many instances of single webservice at the sametime.The toolbox has many deployment options (see Figure 1), and new ones can beadded upon request. The table below shows options that have been applied in the past.

# **4.5.1 REQUIREMENTS FOR DEVELOPMENT ENVIRONMENT**

To develop a webservice software package ready for deployment the following isneeded:

1. A Windows or Linux computer with Matlab 2006b or higher installed on it, (optional) Matlab compiler
2. Matlab Web service toolbox

Matlab compiler is not a requirement, but is recommended. Otherwise each web service instance will occupy a Matlab license. In compiled mode, multiple web service instances can run at no additional costs.

To deploy this web service software package you need the following:

1. A Windows or Linux machine with Matlab or Matlab Runtime installed
2. Apache Tomcat
   1. **Dataset**

Two DDoS datasets are used for evaluating the proposed attack detection technique. One of these datasets represents a network layer flooding attack while the other two represent different types of application level attacks. Each of these attack datasets are simulated for a period of 5 minutes: first by two minutes of normal activity, followed by three minutes of flooding. Each of these datasets is now described.

* 1. The ICMP dataset is based on the CAIDA DDoS Attack Dataset, which represents a real-world DDoS attack and contains approximately an hour(20:50:08 UTC to 21:56:16) of pseudonymised traces from a DDoS attack that occurred on August 4, 2007 . The traffic profile of the CAIDA dataset represents a flooding attack where the attacking machines send a large number of ICMP Echo Requests (ping) to the victim aimed at consuming the victim’s networking bandwidth. The simulated attack uses an identical packet rate and the same number of source IP addresses, as in the CAIDA dataset, to generate a 5-minute attack dataset containing one-way traffic i.e., directed against the target host
  2. Jkjk
  3. Jkj
  4. Ll

Table 4.1: Macro-level Statistics of the 1998 FIFA World Cup and the CAIDA  
dataset.

|  |  |  |
| --- | --- | --- |
| **Characteristic Features** | **CAIDA Dataset** | UCSL Daataset |
| Activity type | DDoS Attack |  |
| Duration | 66 min |  |
| Packets typ e | ICMP ECHO |  |
| Numb er of target(s) | 1 |  |
| Numb er of packets sen | 359,655,826 |  |
| File format | p cap |  |
| Size (uncompressed) | 5.3 GB |  |

Table 4.2: Micro-level statistics of the 1998 FIFA World Cup and the CAIDA  
dataset during the observation period.

|  |  |  |
| --- | --- | --- |
| **Characteristic Features** | **CAIDA Dataset** | UCSL Daataset |
| Observation time | 10 mins |  |
| **Numb er of Requests/Packets** | 74,478,486 |  |
| **Numb er of Source IPs** | 8,585 |  |
| Requests p er source IP | 8675.42 |  |
|  |  |  |
|  |  |  |
|  |  |  |

* 1. : **Dataset Pro cessing**

A subset of CAIDA and as used for the experiments described in this section. The dataset is a six hour network trace around the1st semi-final match of the CAIDA . The original dataset was first filtered to remove all traffic except HTTP GET requests. Secondly, only packets containing 200, 206 and 400 HTTP status codes were extracted and used as the input trace (processed dataset) for the experiments. Table 4. shows the highlevel characteristics of the original dataset and the processed dataset used in thesimulations

Table 5.1: Characteristic features of the dataset used for simulations

|  |  |  |
| --- | --- | --- |
| **Parameters** | **Original Dataset** | **Pro cessed Dataset** |
| Numb er of requests | 29,662,465 | 19,615,109 |
| Numb er of unique source IPs | 79,033 | 78,646 |
| Numb er of unique source IPs | 11,885 | 11,839 |
| Numb er of unique source IPs | 4,089 | 2,376 |
| Numb er of unique source IPs | GET, POST, HEAD | GET |
| HTTP status co des | all | 200, 206 and 400 |
|  |  |  |

## In order to ensure that the same web-resources are downloaded during the simulation as were present in the original trace, a fake directory structure was created and installed on the target machine’s cloud aptachie web server . All the files within this directory structure used the same file name and contained the same amount of data as in the original traffic trace, although the contents were junk data.

**Incoming Traffic**

The incoming traffic for the simulated and the original

The total traffic volume of the simulated

**Source IP Addresses**

**The source IP address parameter of the original traces compared with the simulated traffic in two forms: the total number of unique source IPs, and the variations in the number of unique source IPs over time.**

## Hardware Platform

One of the major ECC requirement is to implement in high speciation of hardware; the complexity of district logarithms.

Previous studies employed relevant implementation of ECC software. Dedicated PC resources were used in implementing ECC in the study by Diffie Hellman ([Sankar, Subashri, & Vaidehi, 2011](#_ENREF_35)). Other studies show several types of hardware used ([Gupta, Sikka, & Katiyar](#_ENREF_15); [Hardie & Connell, 2004](#_ENREF_17); [Katiyar, Dutta, & Gupta, 2010](#_ENREF_22); [Shohdy, Elsisi, & Ismail, 2010](#_ENREF_41)).

Table 4.1 shows the PC specifications used in implementing our proposed ECC compiler.

Table 4.1: PC specifications used in implementation

|  |  |  |
| --- | --- | --- |
| Processor | Main Circuit Board | Memory Modules |
| 2.93 gigahertz Intel Core2 Duo.64 kilobyte primary memory cache 3072 kilobyte secondary memory cache 64-bit ready Multi-core (2 total) Not hyper-threaded | Board: ASUSTeK Computer INC. P5G41C-M LX Rev X.0x Bus Clock: 266 megahertz | 1024 Megabytes Usable Installed Memory  Slot 'DIMM A1' has 1024 MB Slot 'DIMM B1' has 2048 MB |

Table 5.1: Summary of the Datasets

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | DDoS | Total | Purpose |
|  |  |  |  |
|  |  |  |  |

Features for All attacks

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **UDP Features** | **ICMP features** | **TCP Sync** | **TCP PSH** |
| 1 | No of packets | No of packets | No of packets | No of packets |
| 2 | Average packet size | Average packet size | Average packet size | Average packet size |
| 3 | No of bytes | No of bytes | No of bytes | No of bytes |
| 4 | Packet rate | Packet rate | Packet rate | Packet rate |
| 5 | Bit rate | Bit rate | Bit rate | Bit rate |
| 6 | Ratio udp | Ratio udp | Ratio udp | Ratio udp |
| 7 | Ratio tcp | Ratio tcp | Ratio tcp | Ratio tcp |
| 8 | Ratio icmp | Ratio icmp | Ratio icmp | Ratio icmp |
| 9 | prototype | prototype | prototype | prototype |
| 10 | Source count | Source count | Sync Count | Sync Count |
| 11 | Destination count | Destination count | Push count | Push count |
| 12 | - | - | Fin count | Fin count |
| 13 | - | - | Ack count | Ack count |

**4.6 IDPS Metrics and parameters**

The validation of feature selection depended on the performances obtained by classifiers when working with the subsets under test.

**4.6.1 Percentage Correct Classification**

**4.6.2 Percentage Incorrect Classification**

**4.6.3 Accuracy**

Accuracy, i.e., percentage of correctly classified vectors.

**Percentage In Correct Classification**

**Percentage Correct Classification**

**Precision**

(or positive predictive value for attacks), i.e., the fraction of predicted

**Sensitivity**

**Recall**

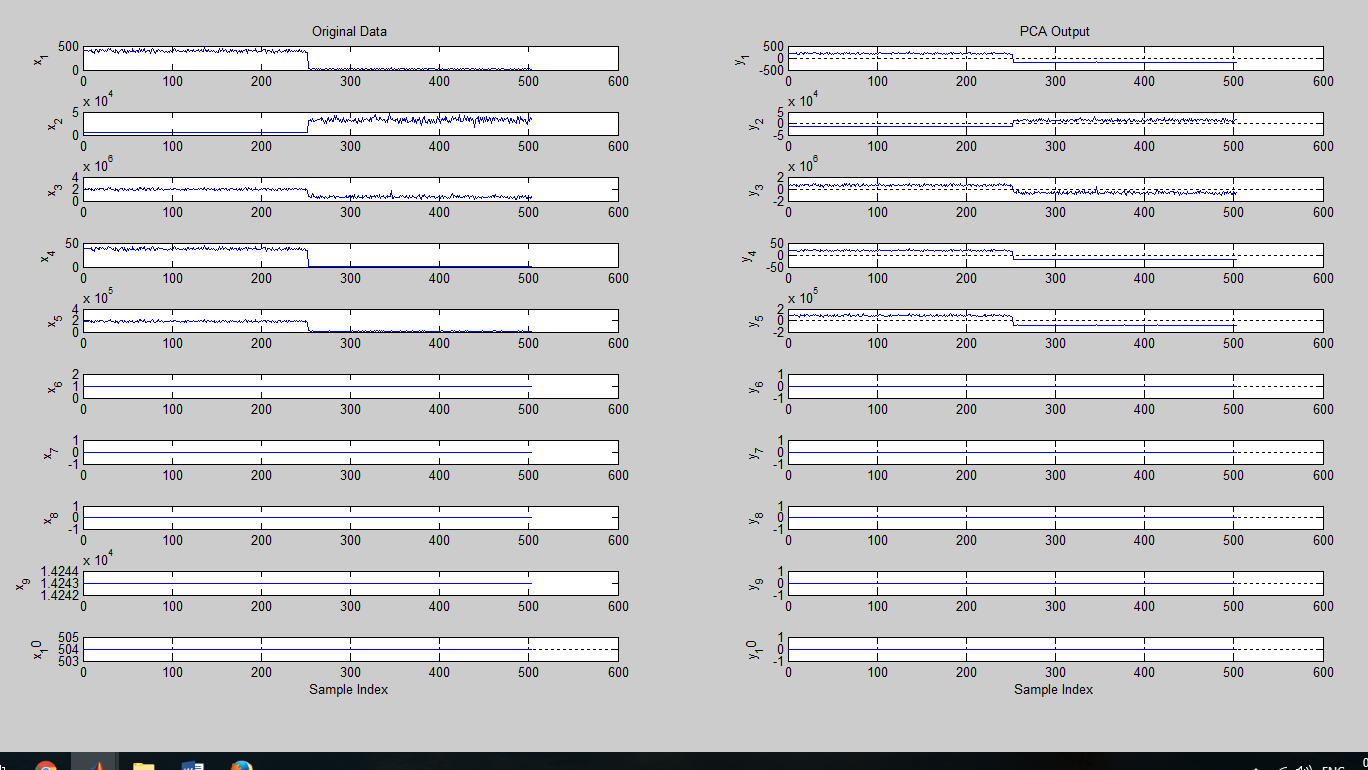
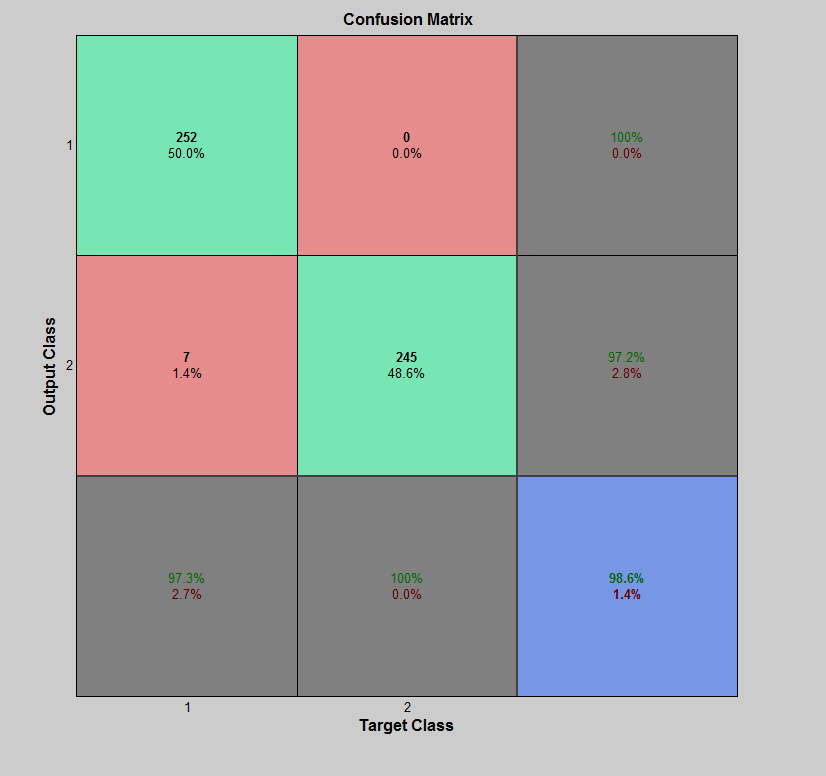
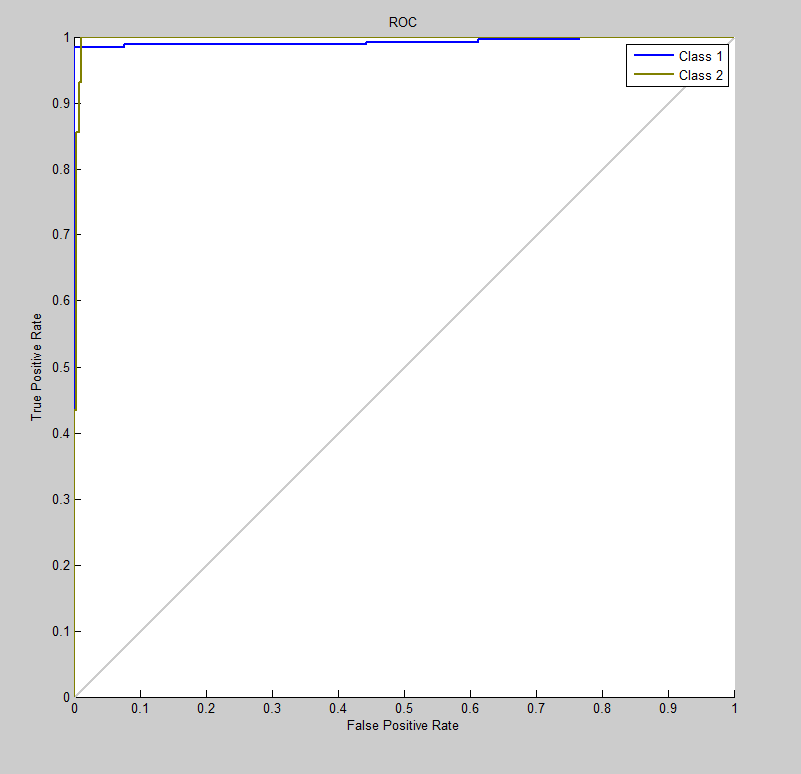
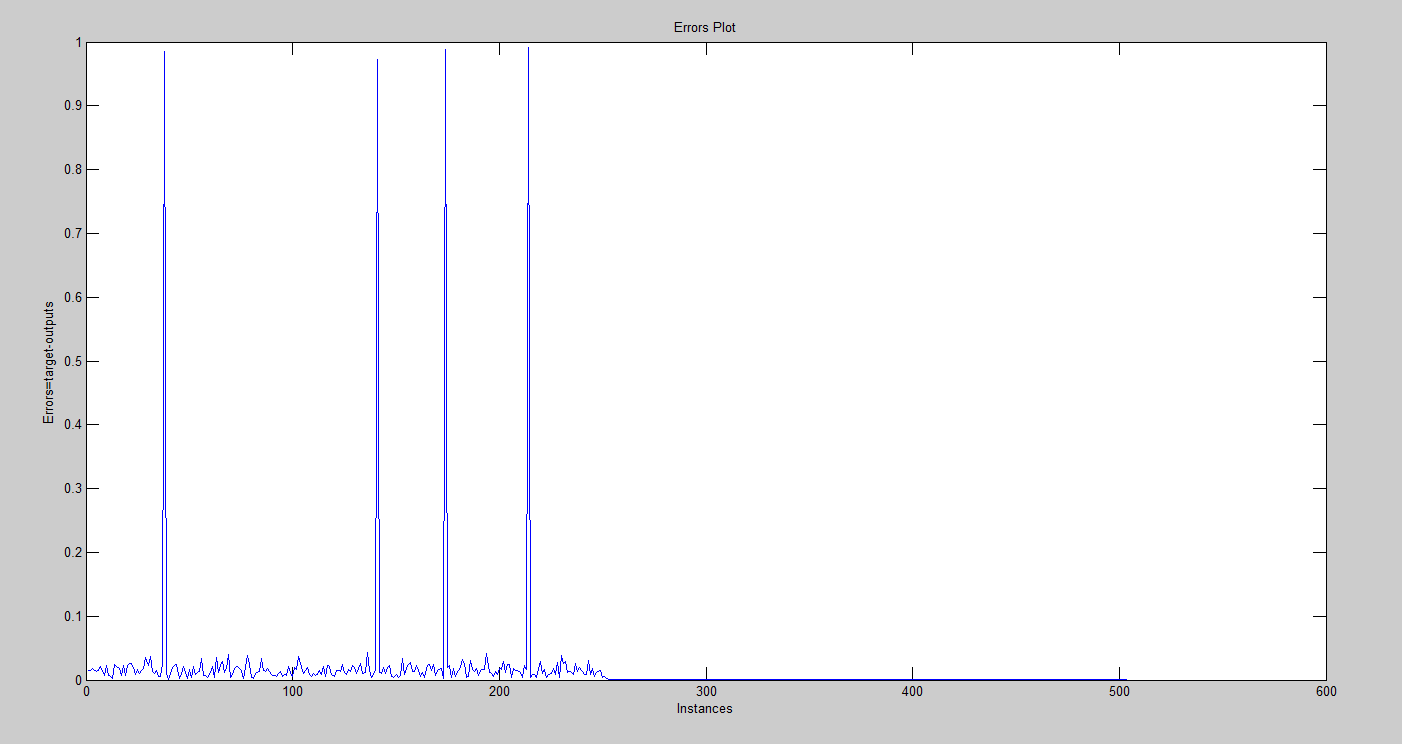
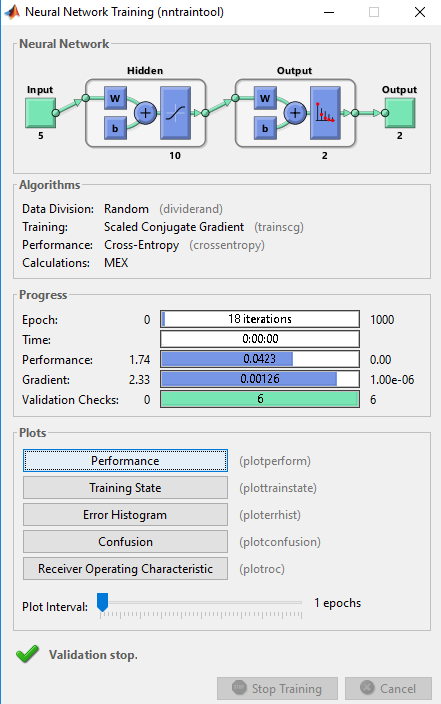
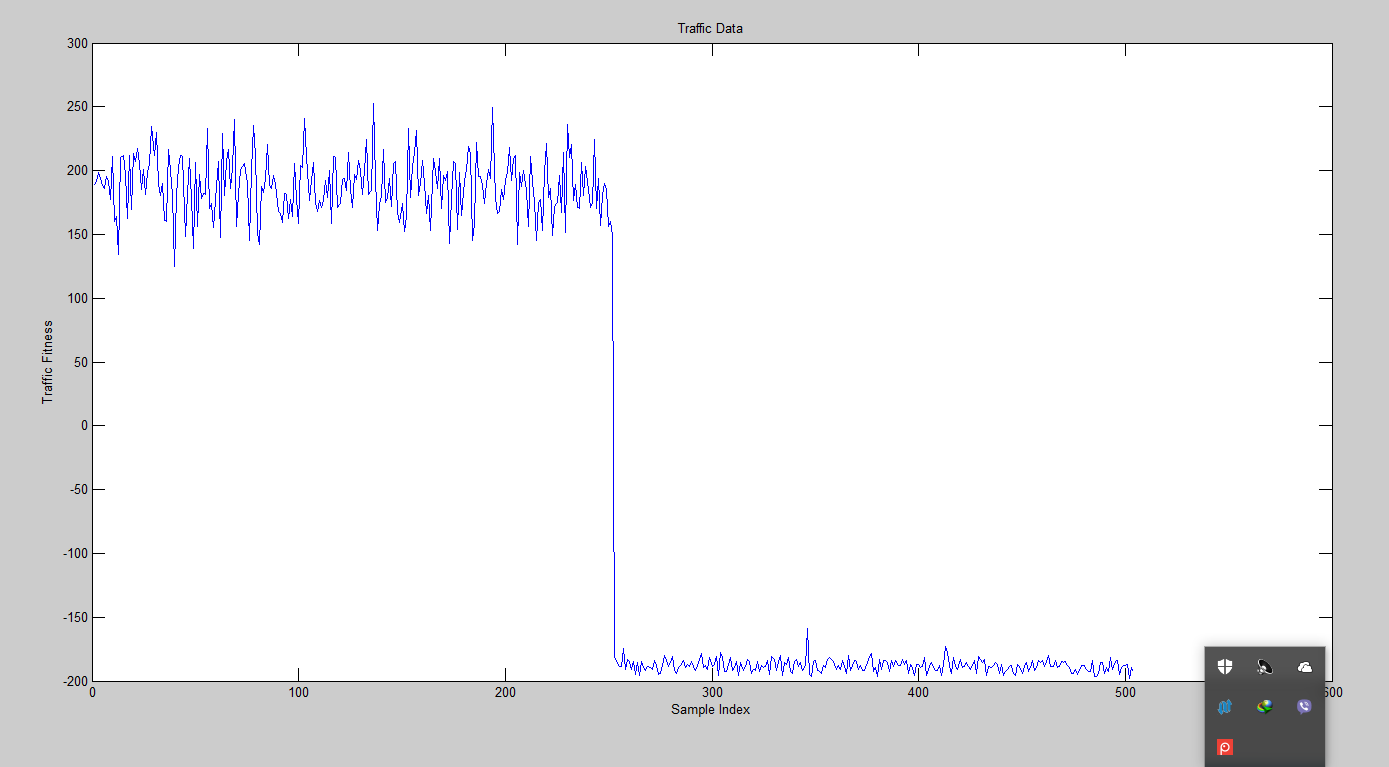
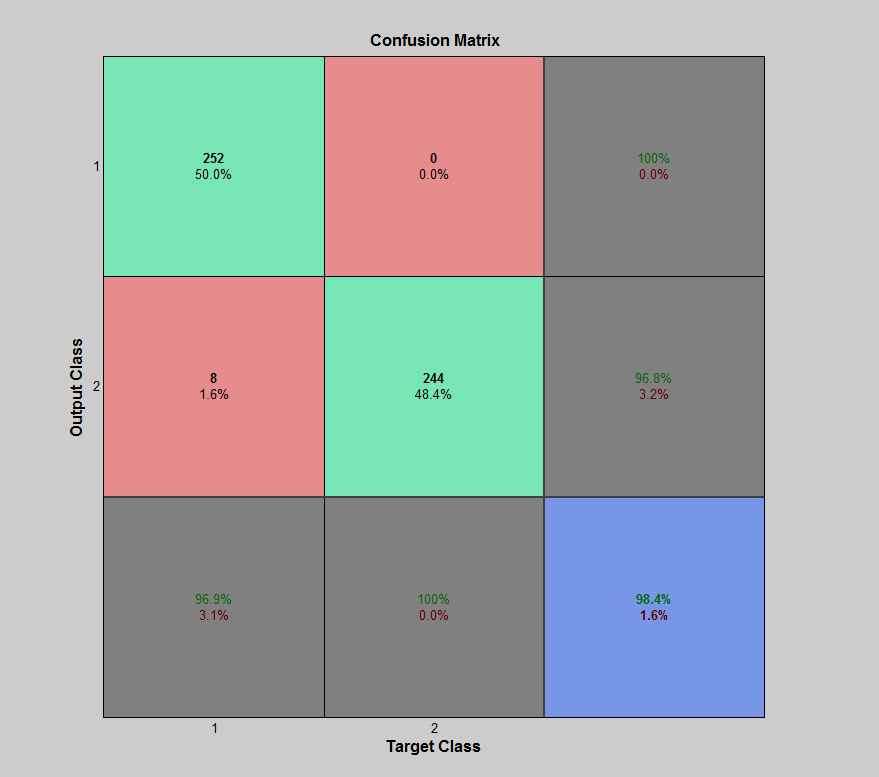
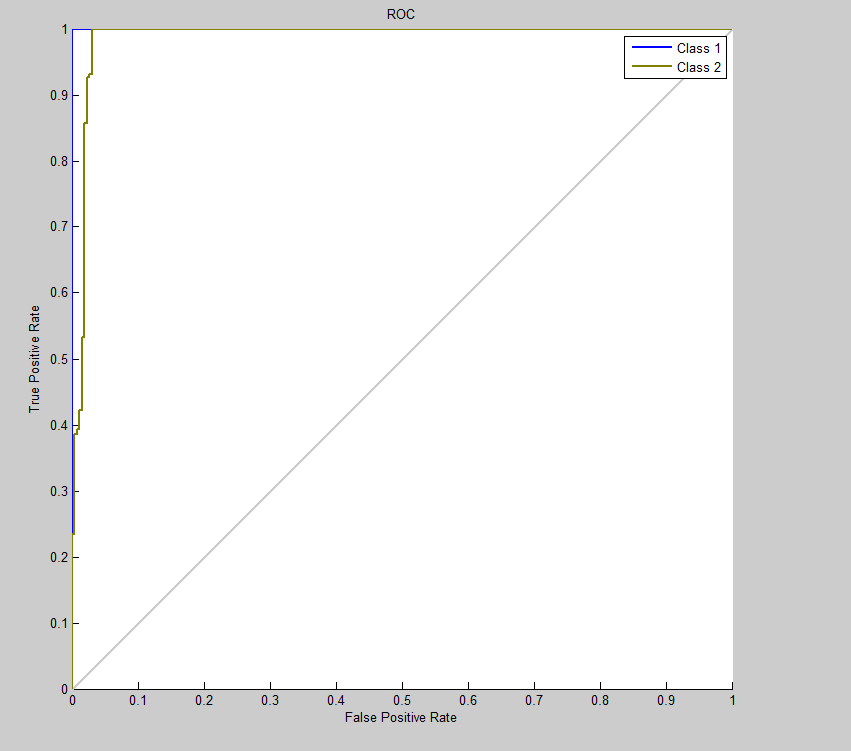
Attack recall (or sensitivity for attacks), i.e., the fraction of real attacks correctly predicted

**Fscore**

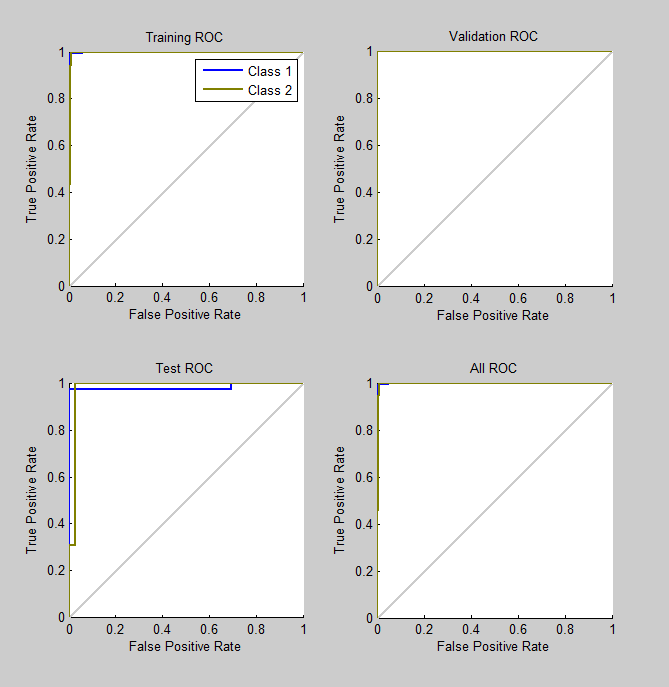
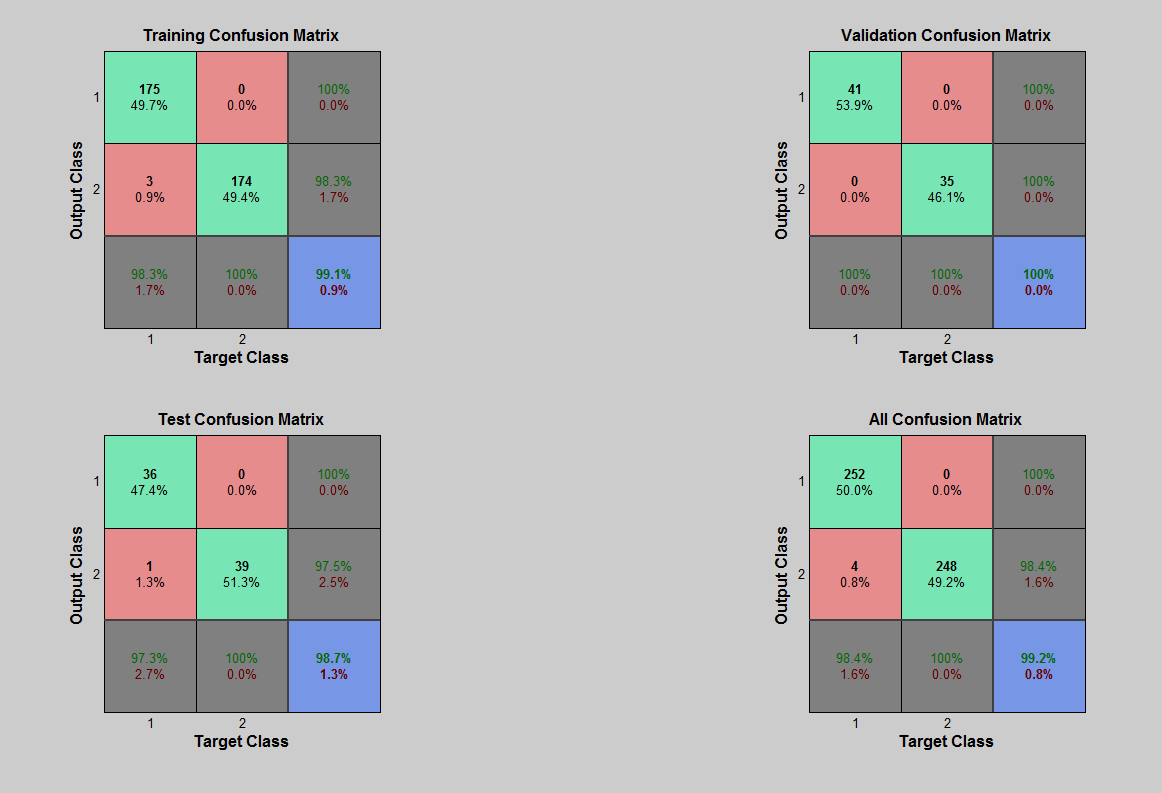
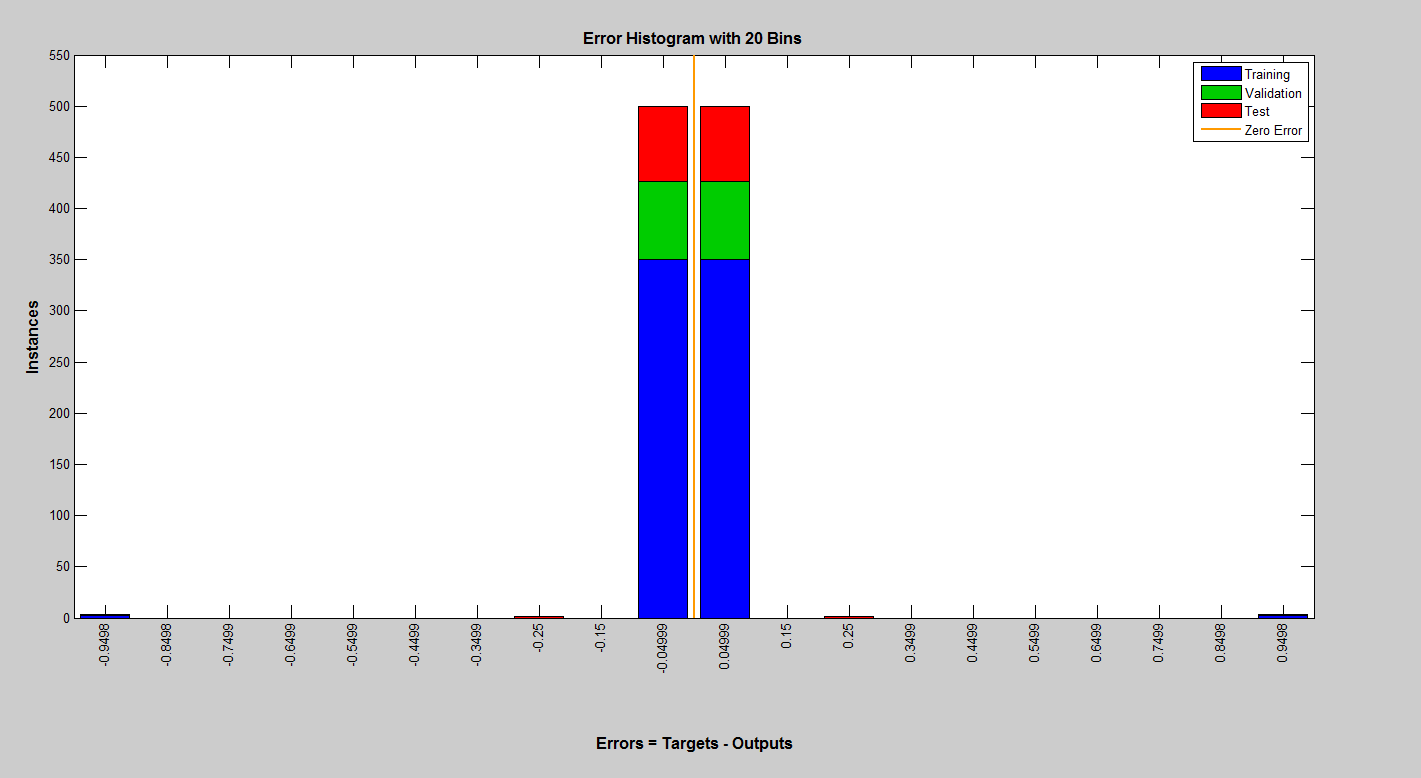
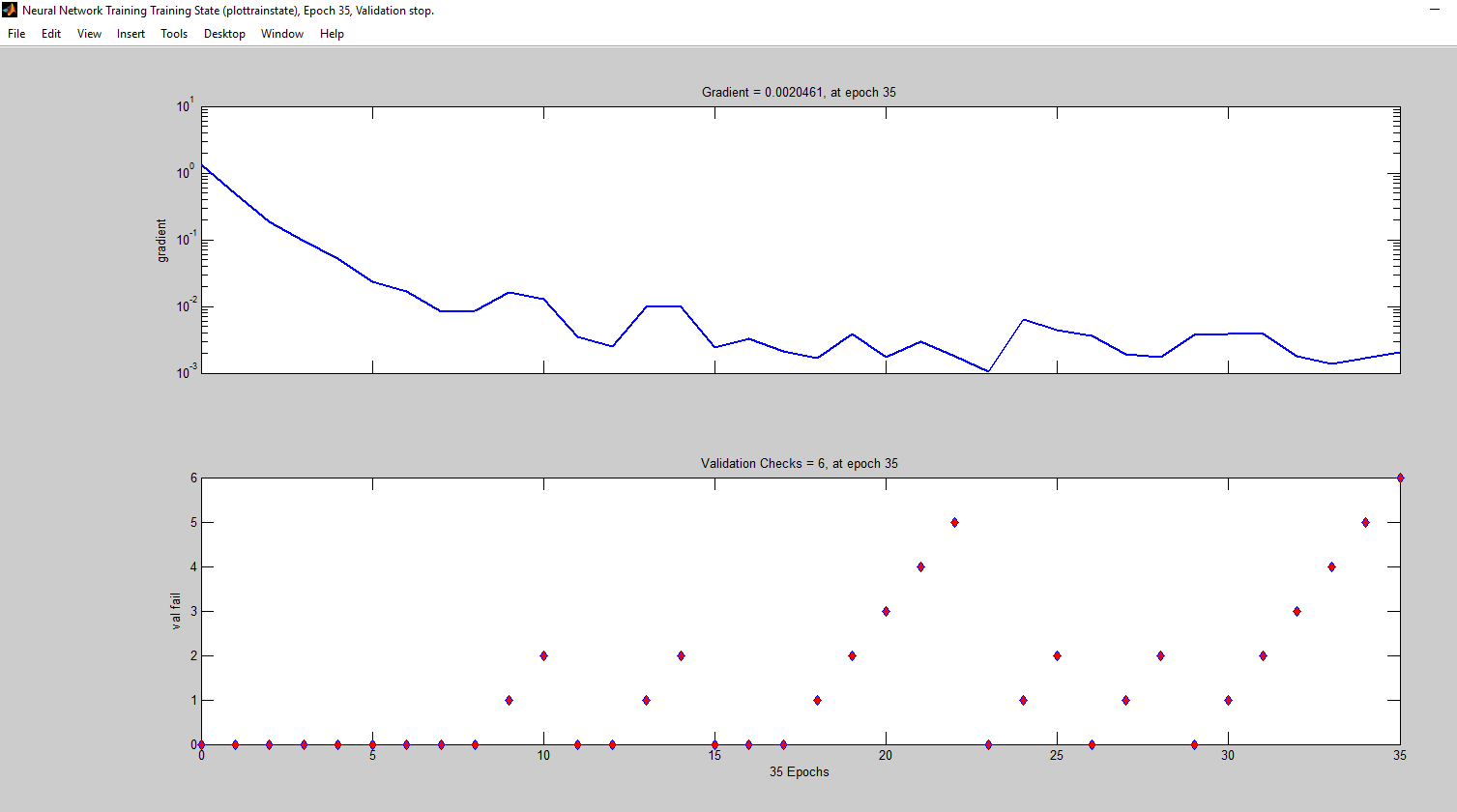
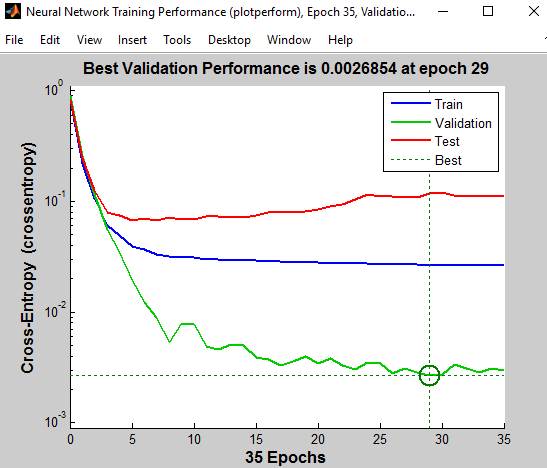
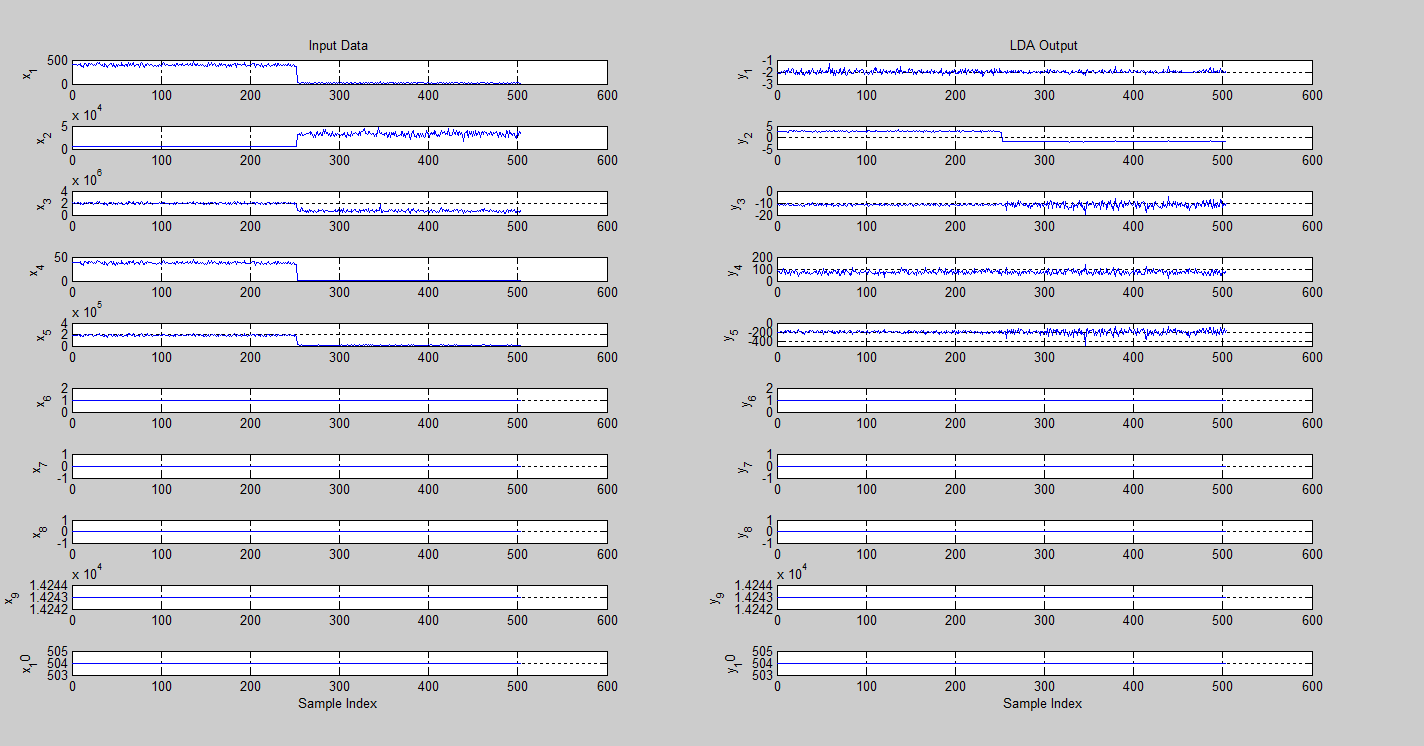
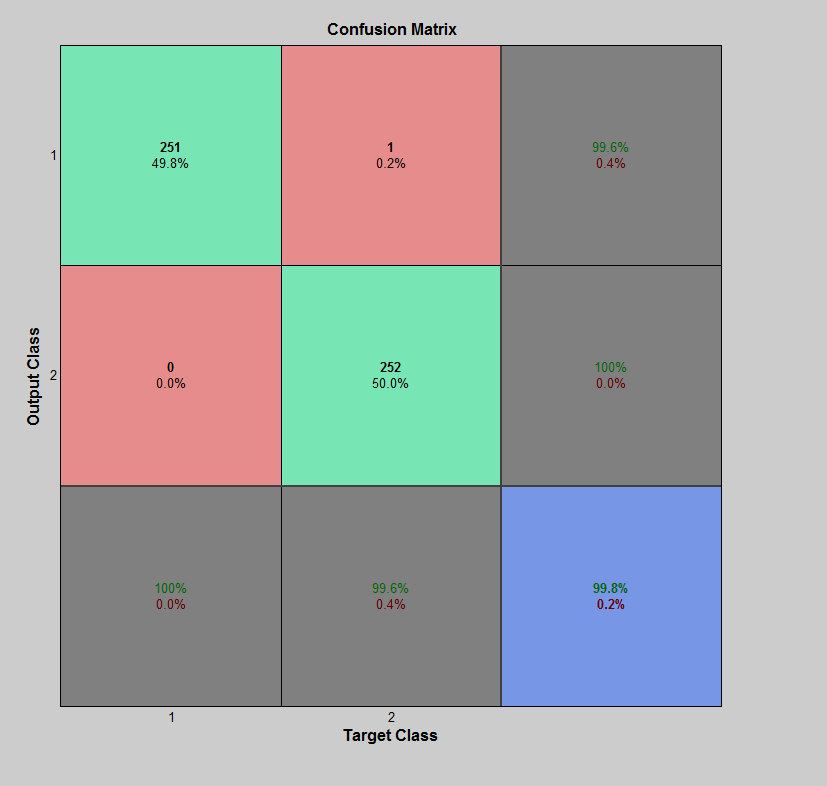
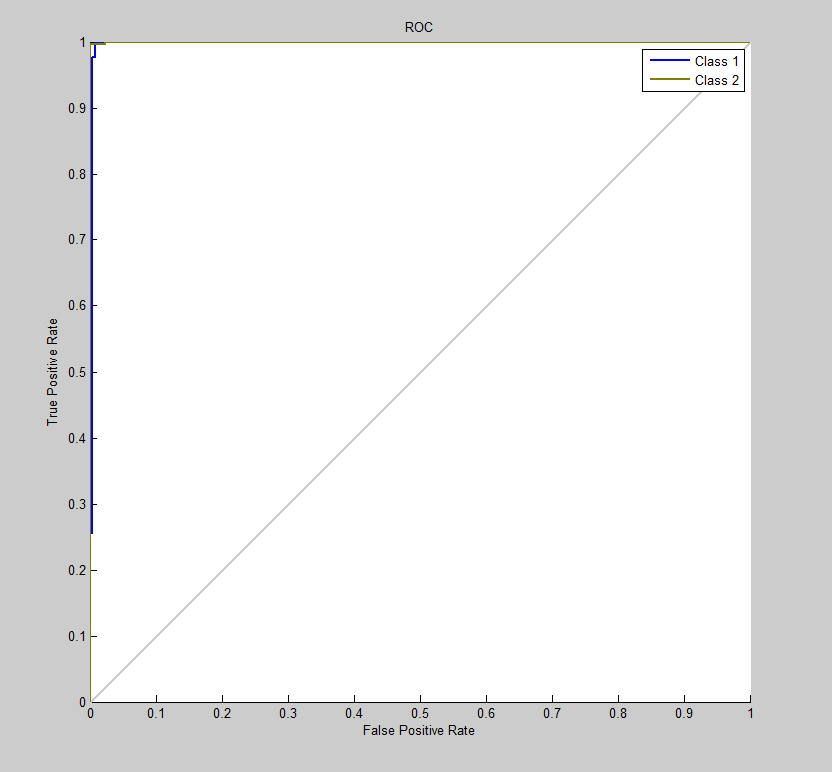
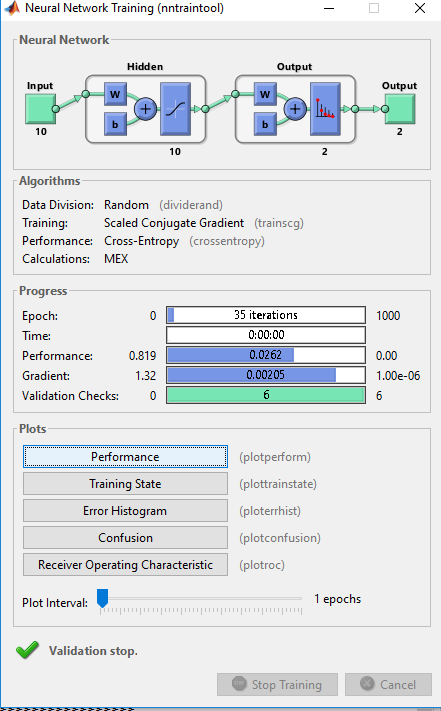
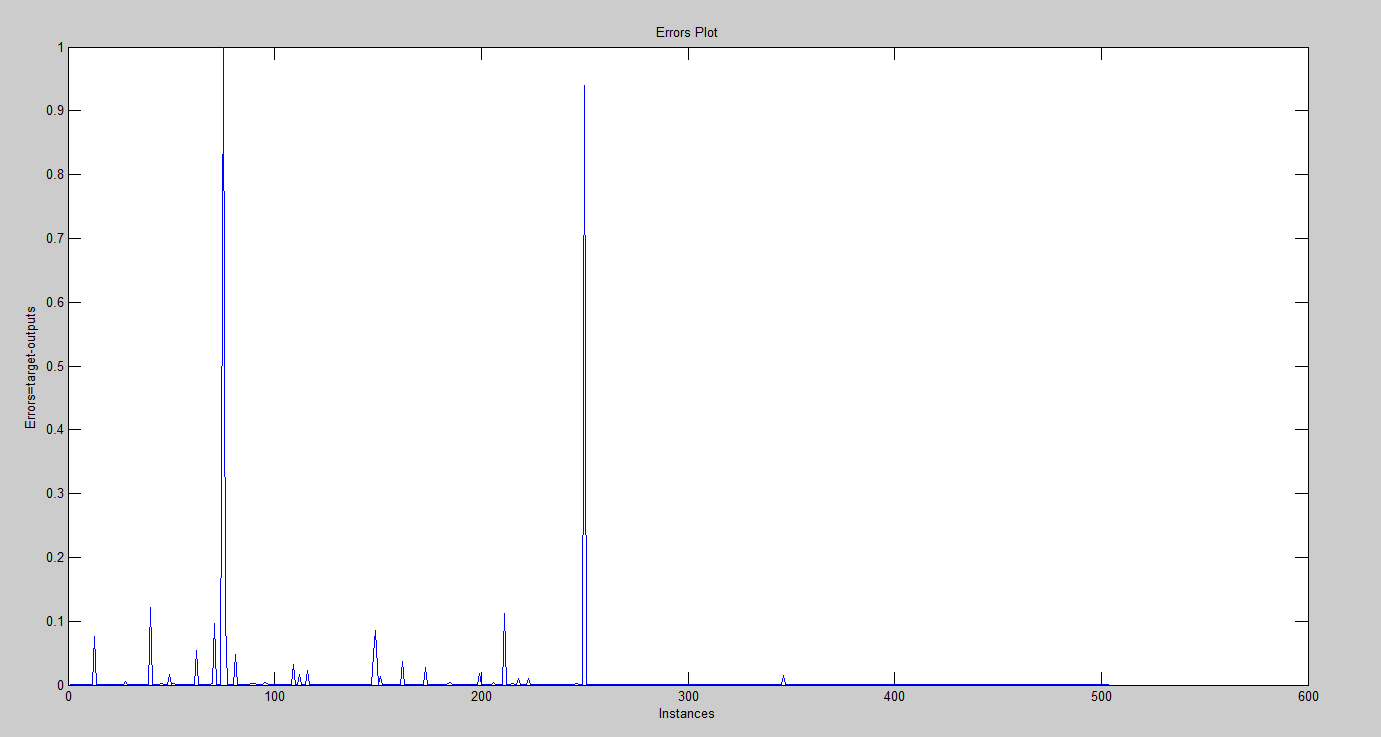
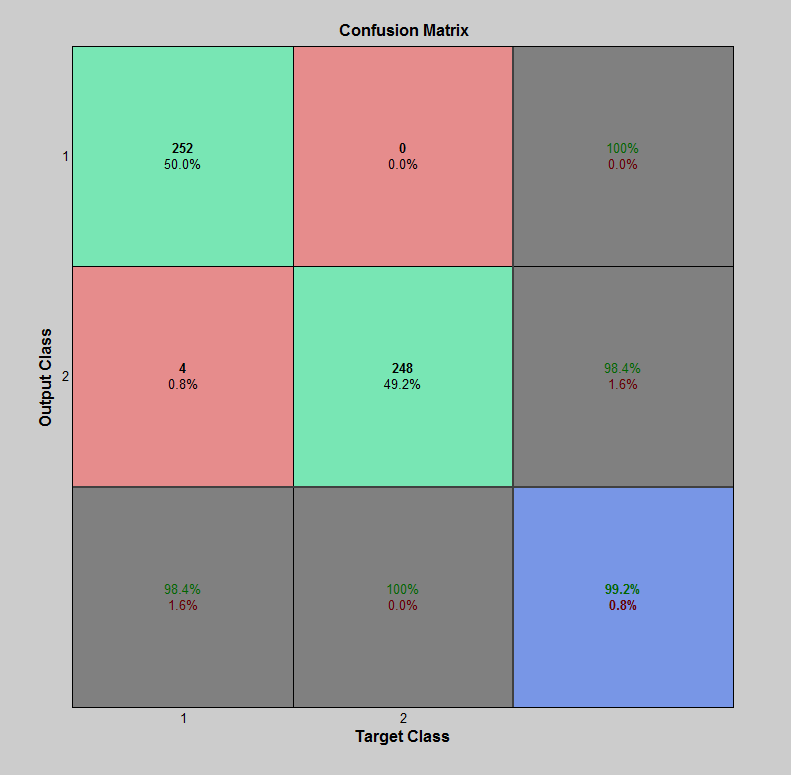
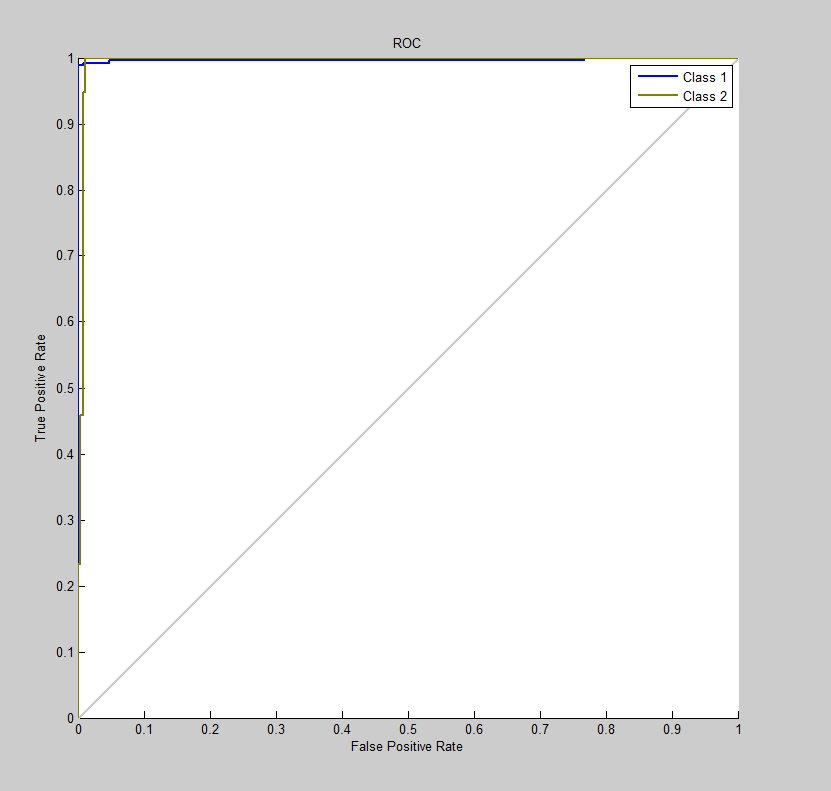
Area Under ROC Curve (AUC).

The AUC is a metric to compare classifiers. In general terms, AUC is suitable for problems with biased or unbalanced class distribution. Moreover, the AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive sample higher than a randomly chosen negative sample(Fawcett 2006). It is important to remark that, even in situations where distinct classification approaches obtain even performances, binary classifiers (DTC, kNN) get worse AUC than probabilistic classifiers (e.g., Bayes-based). Moreover, as smartly analyzed inHand (2009), AUC values can give potentially misleading results, hence they should be interpreted together with other indices

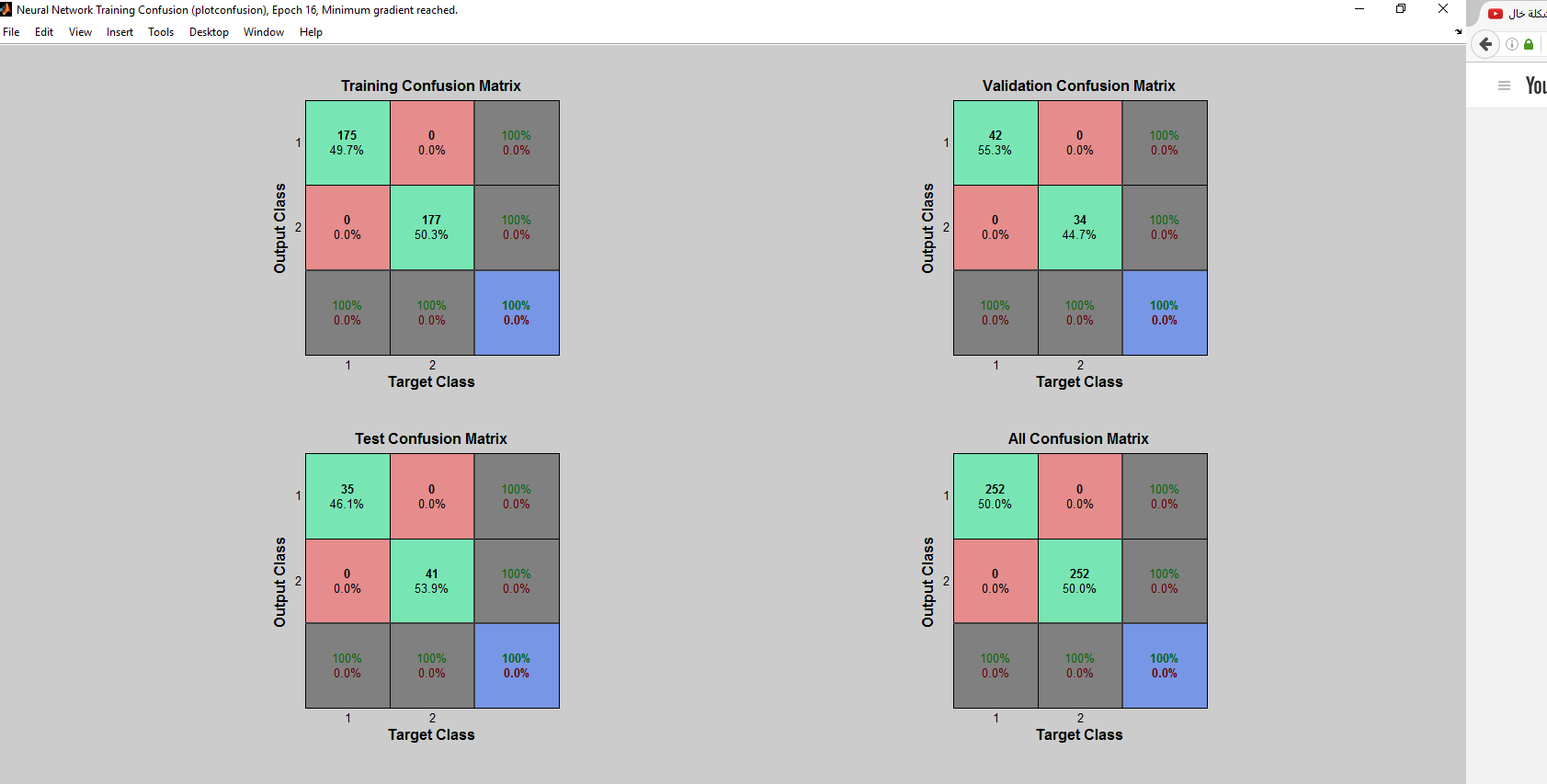
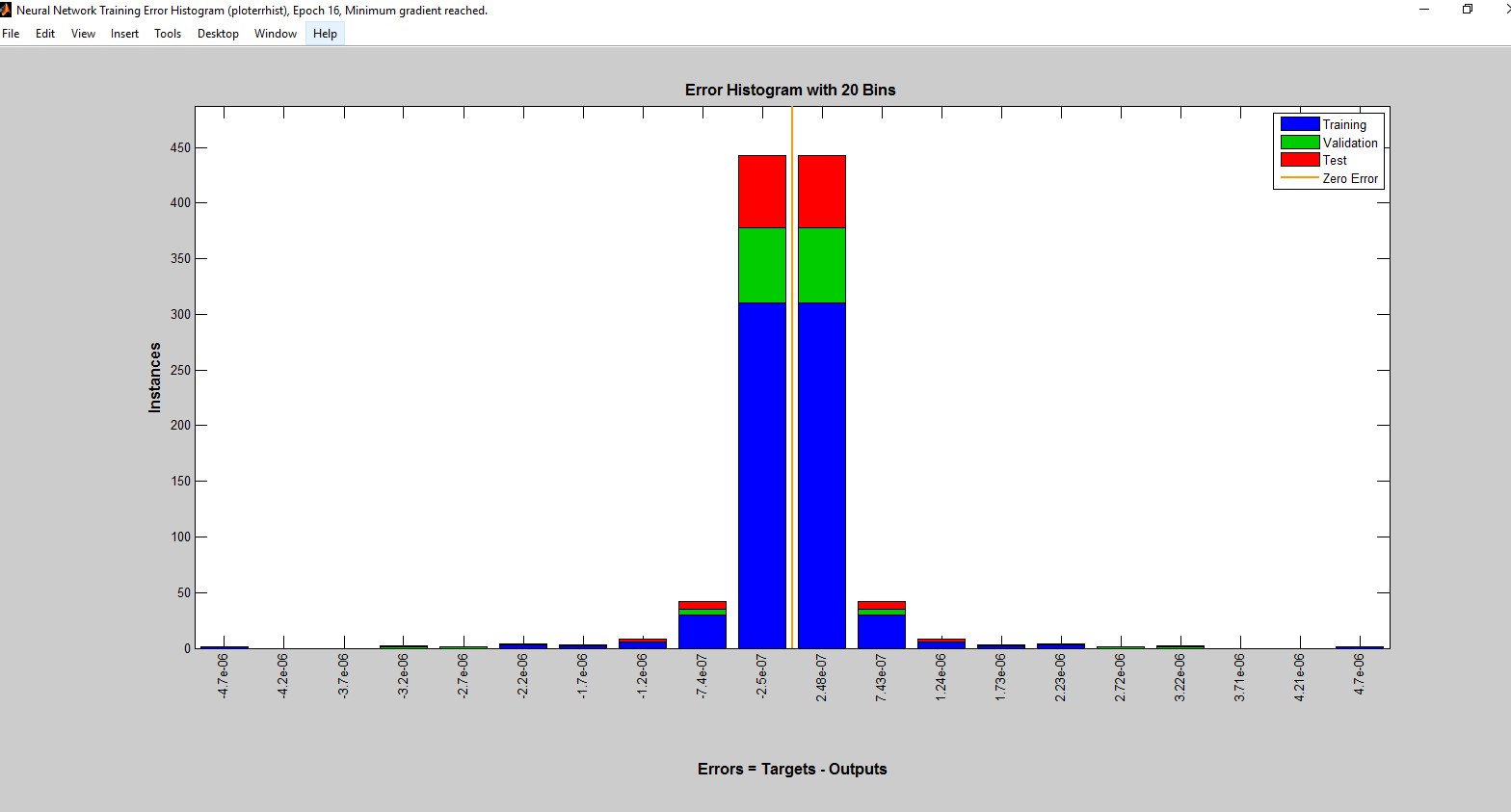
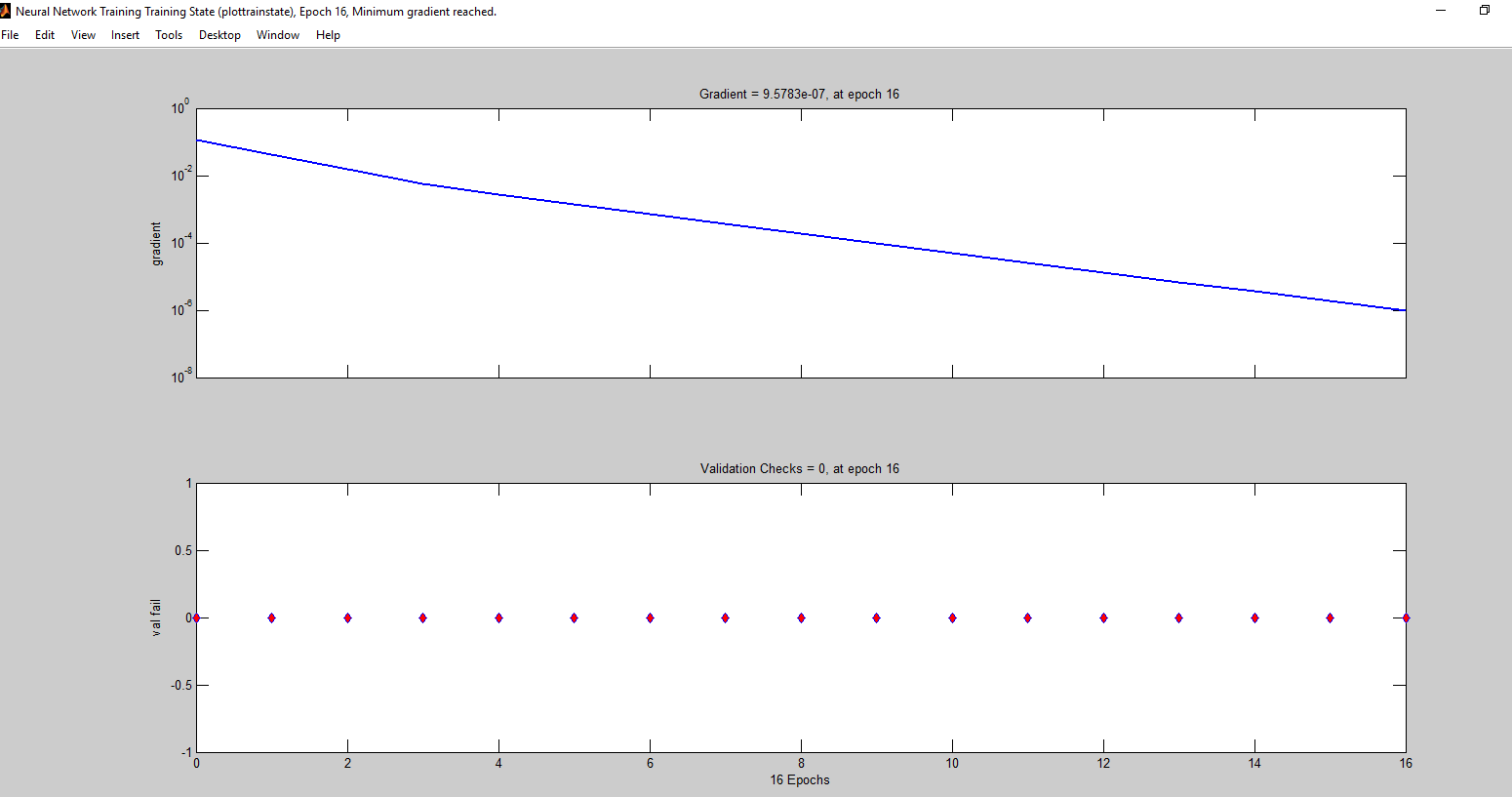
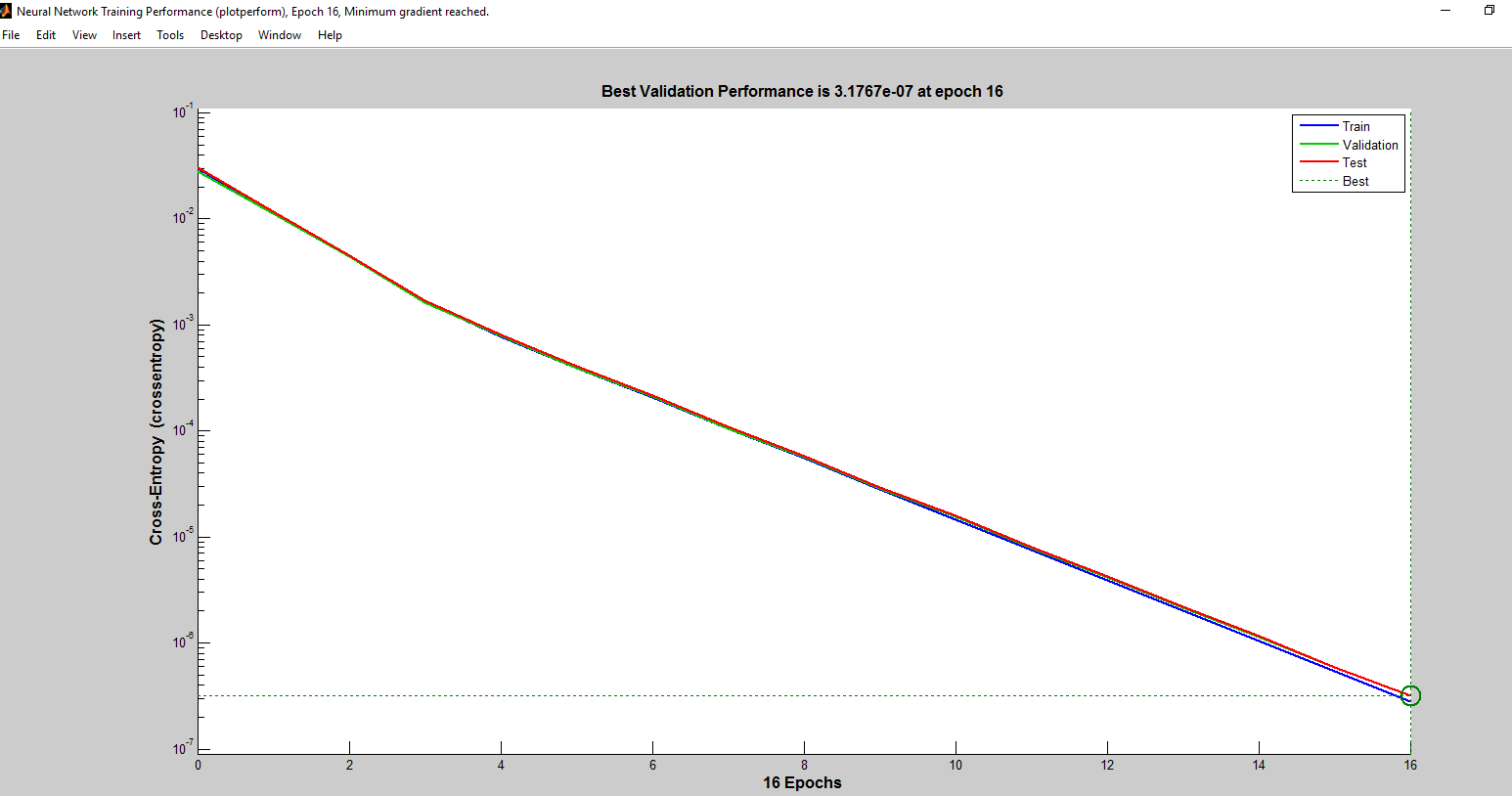
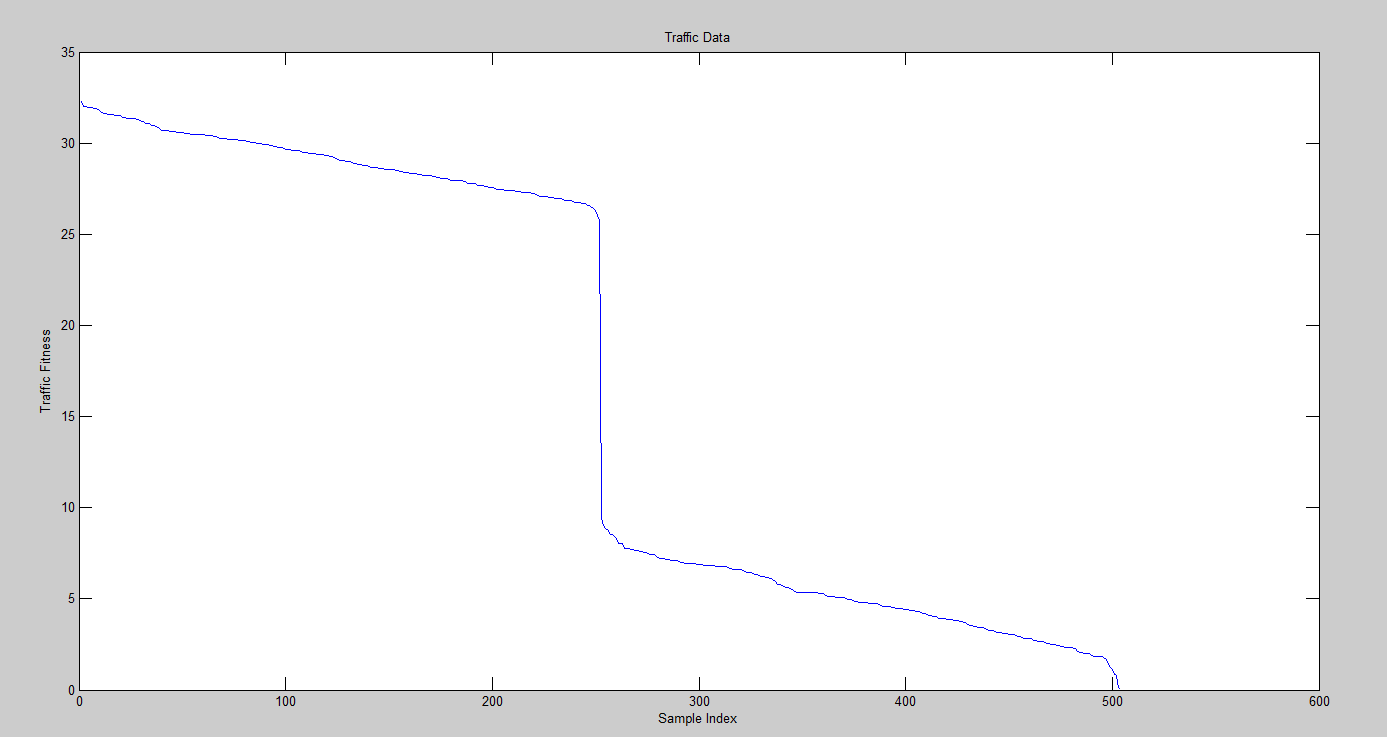
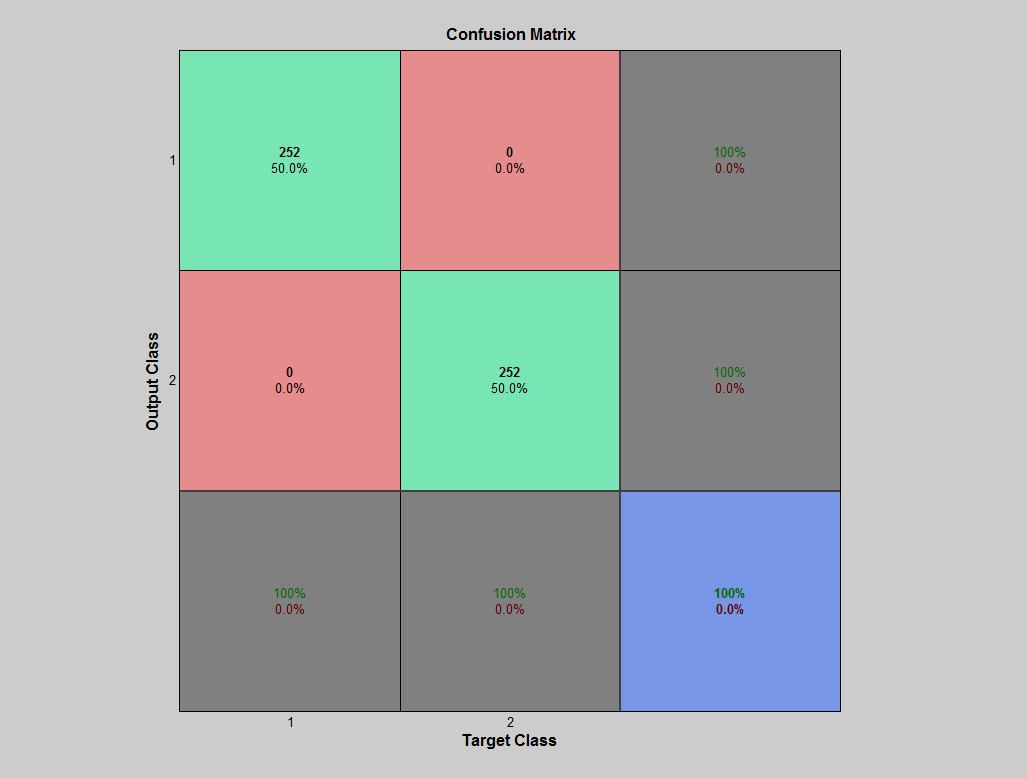
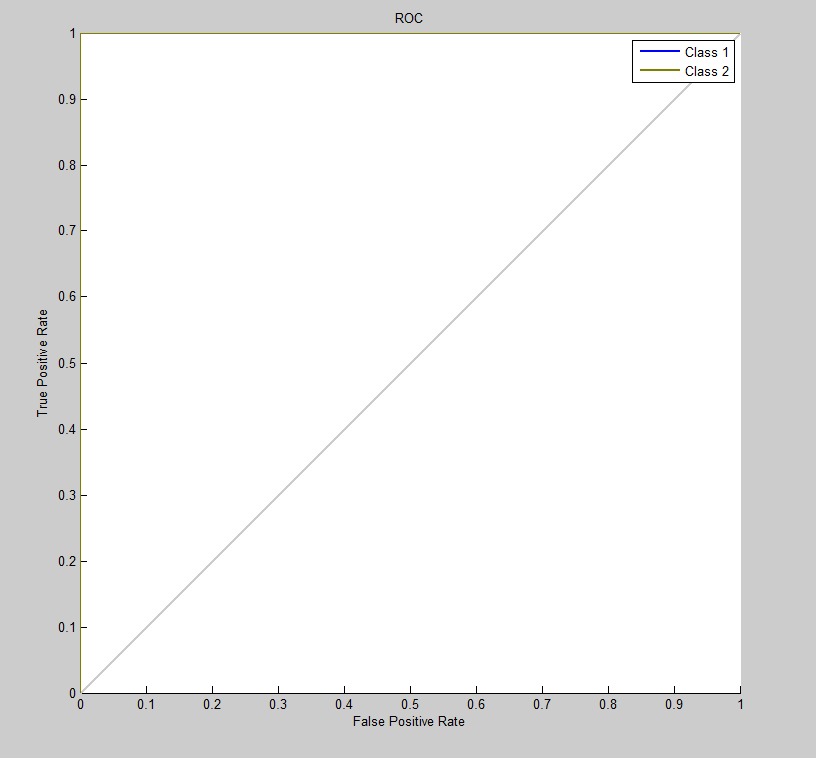
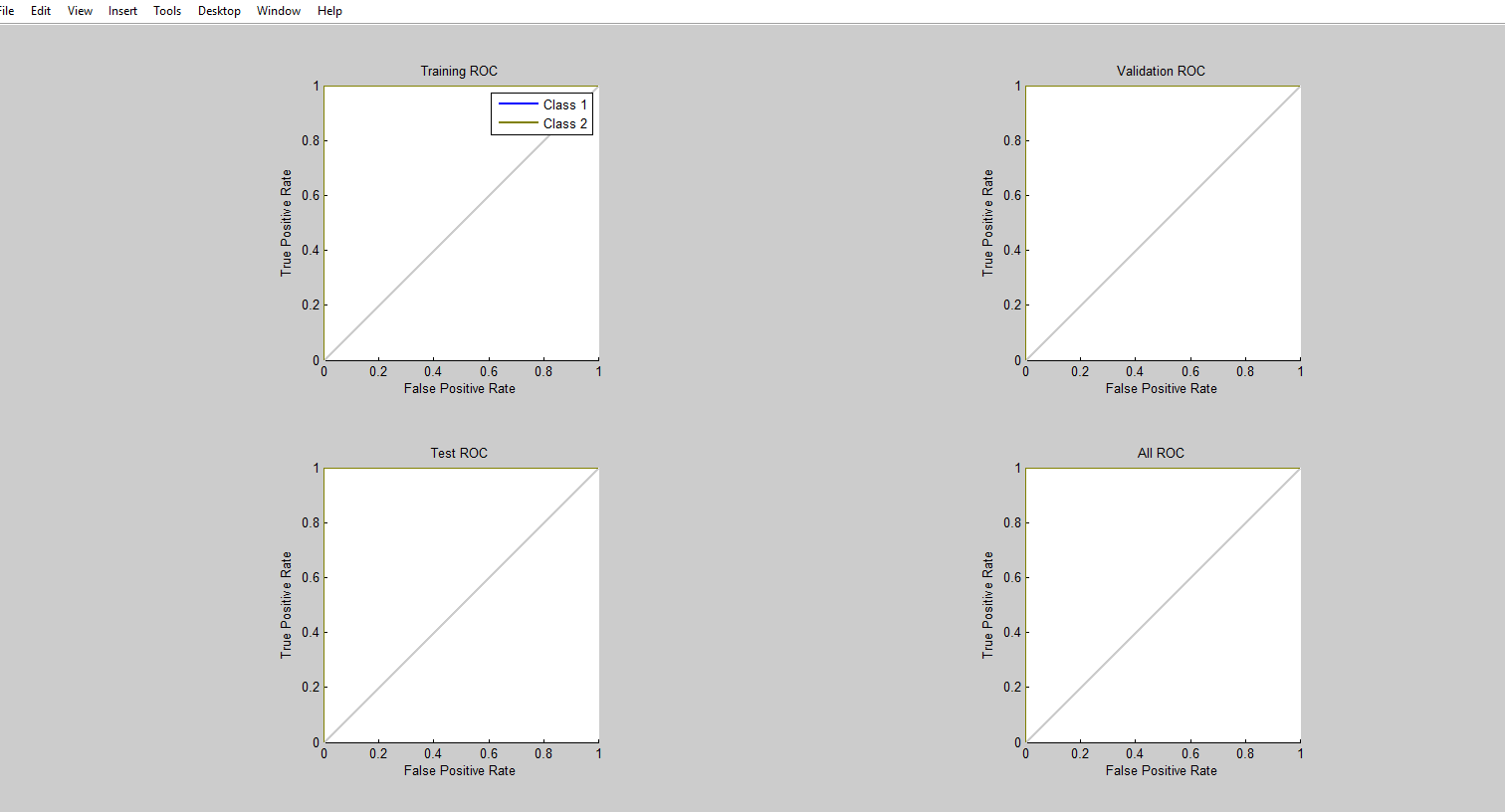
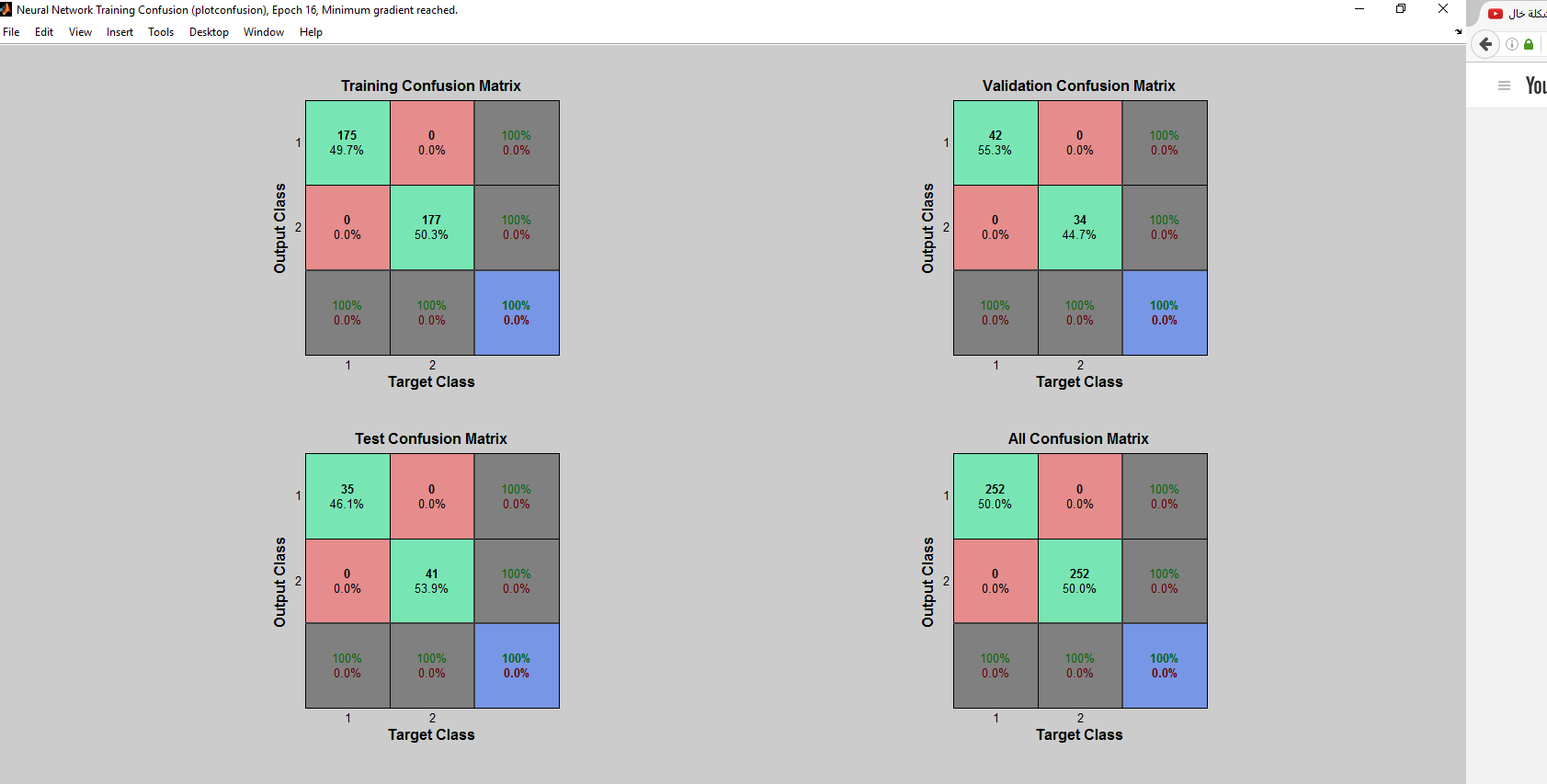
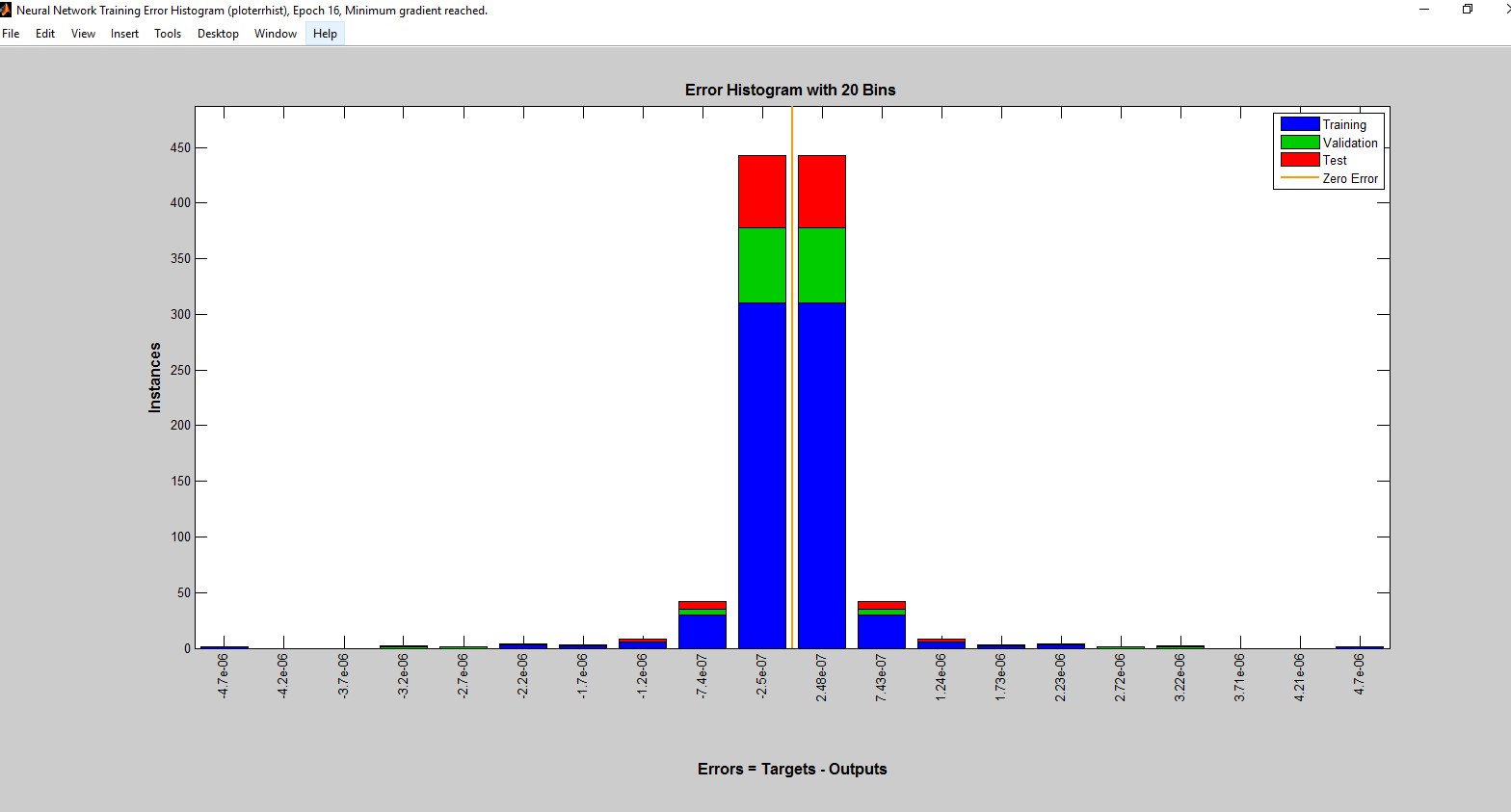
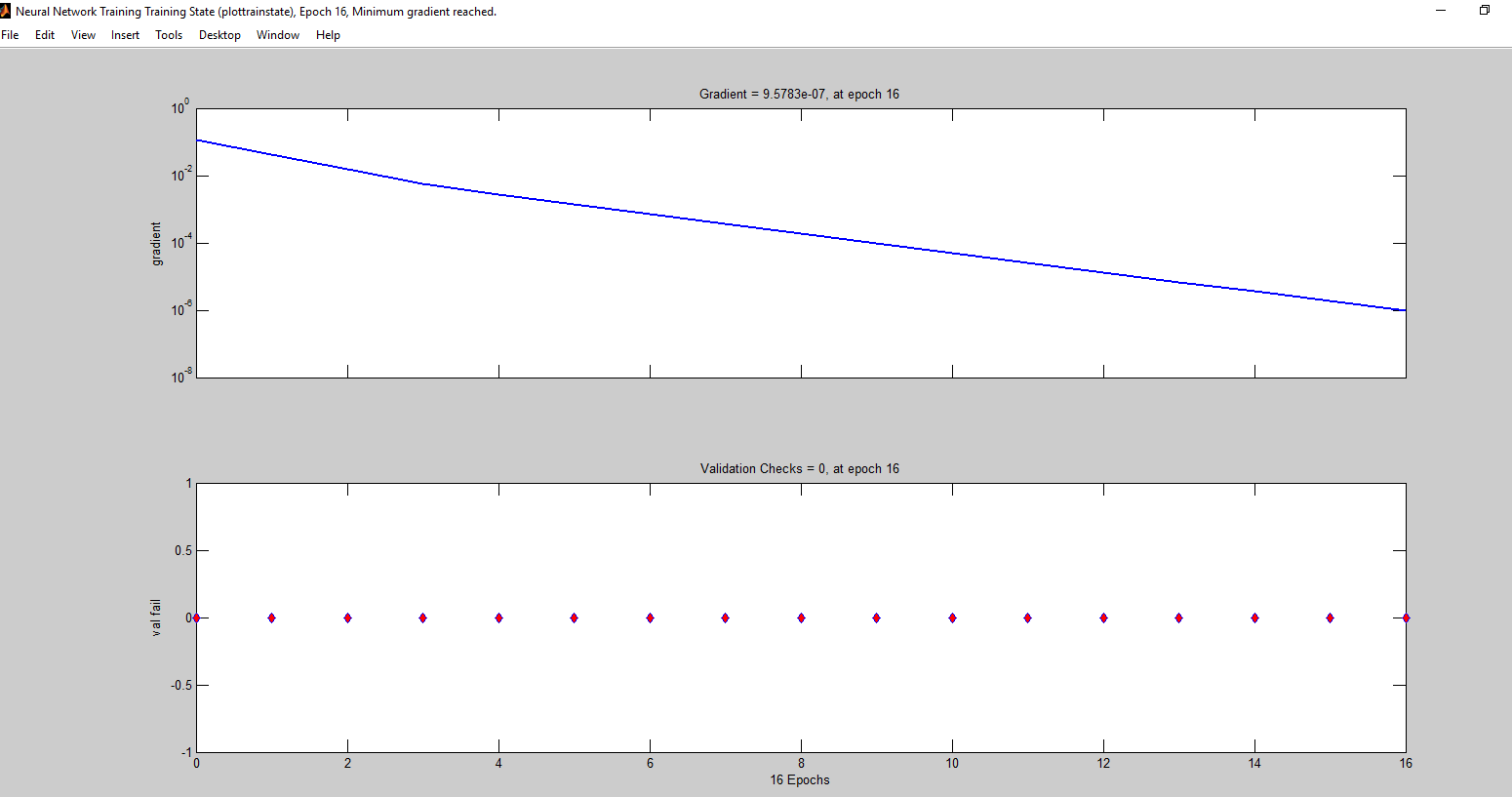
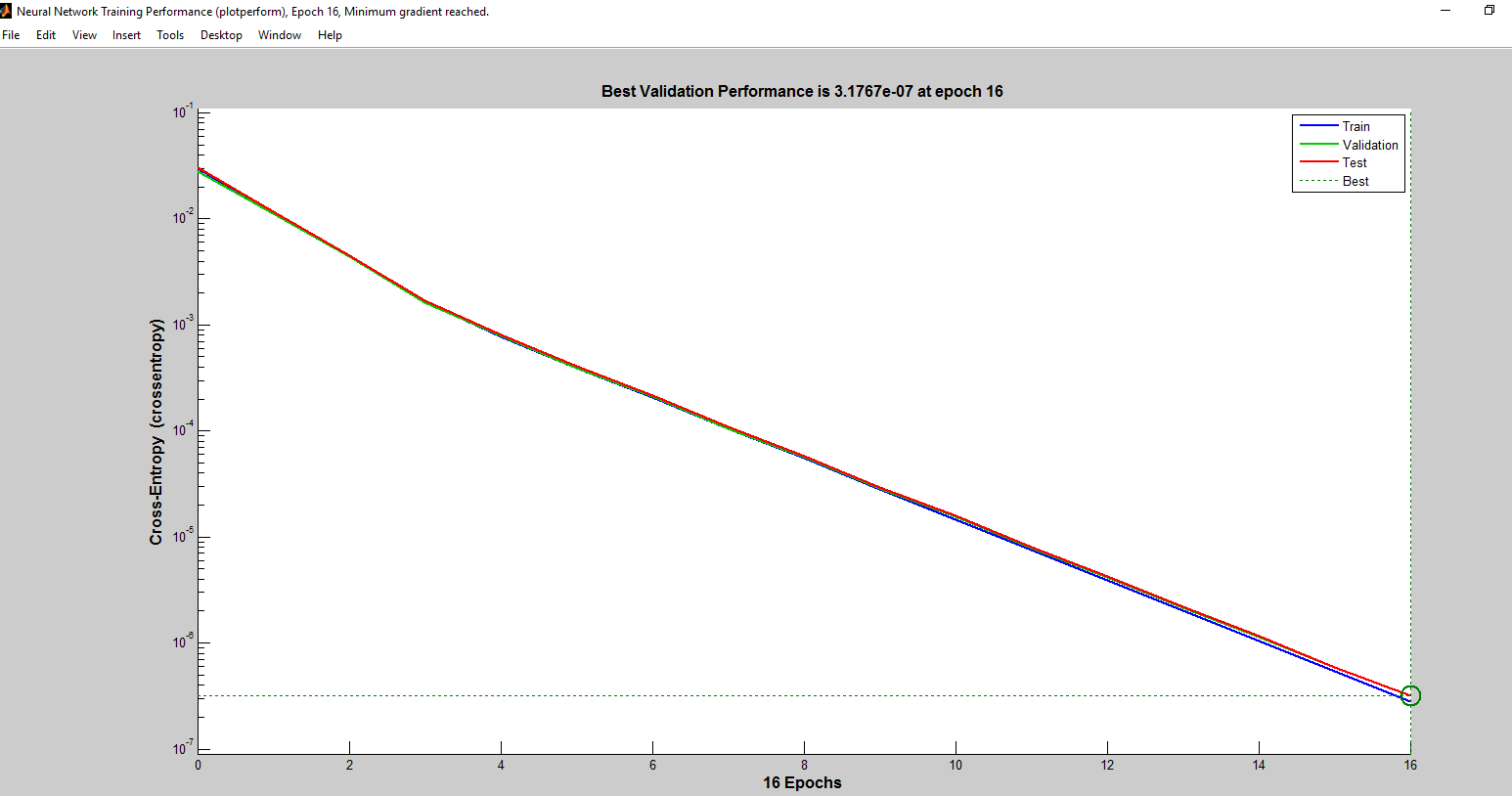
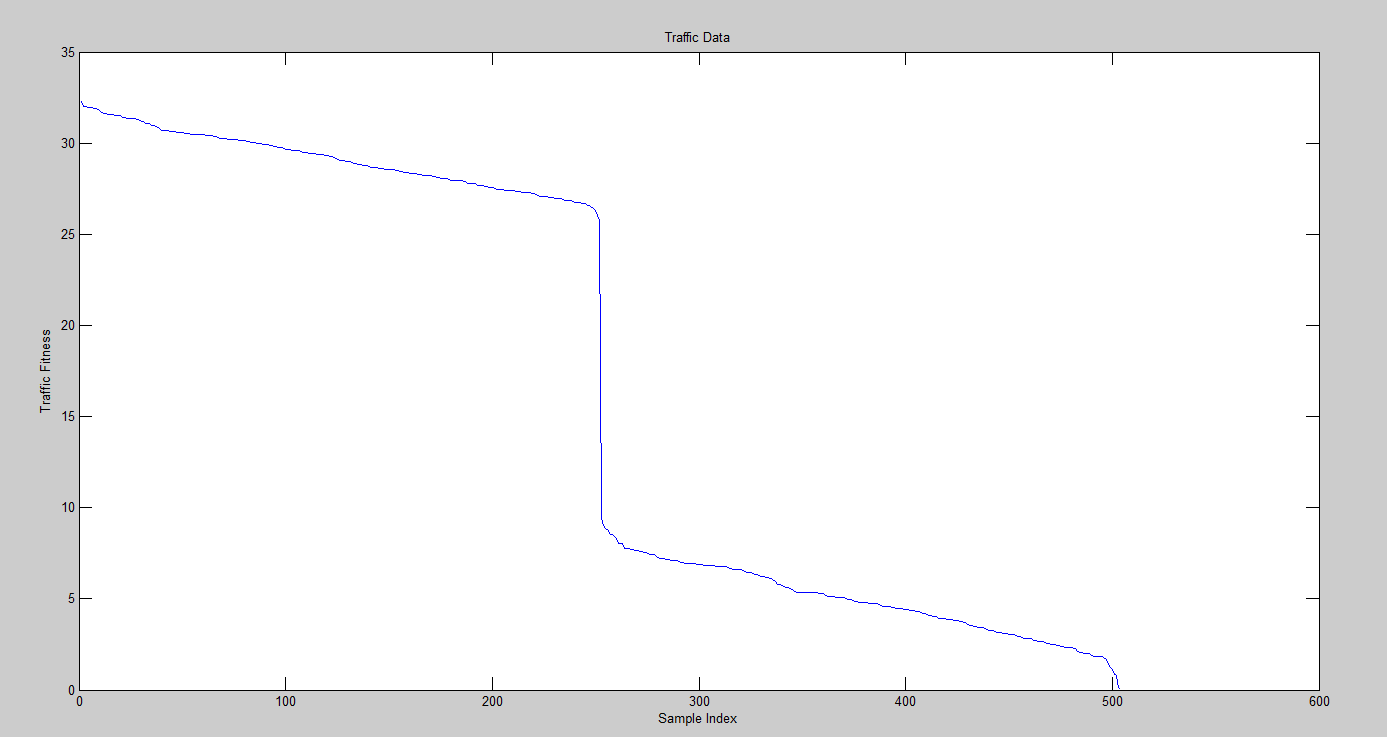
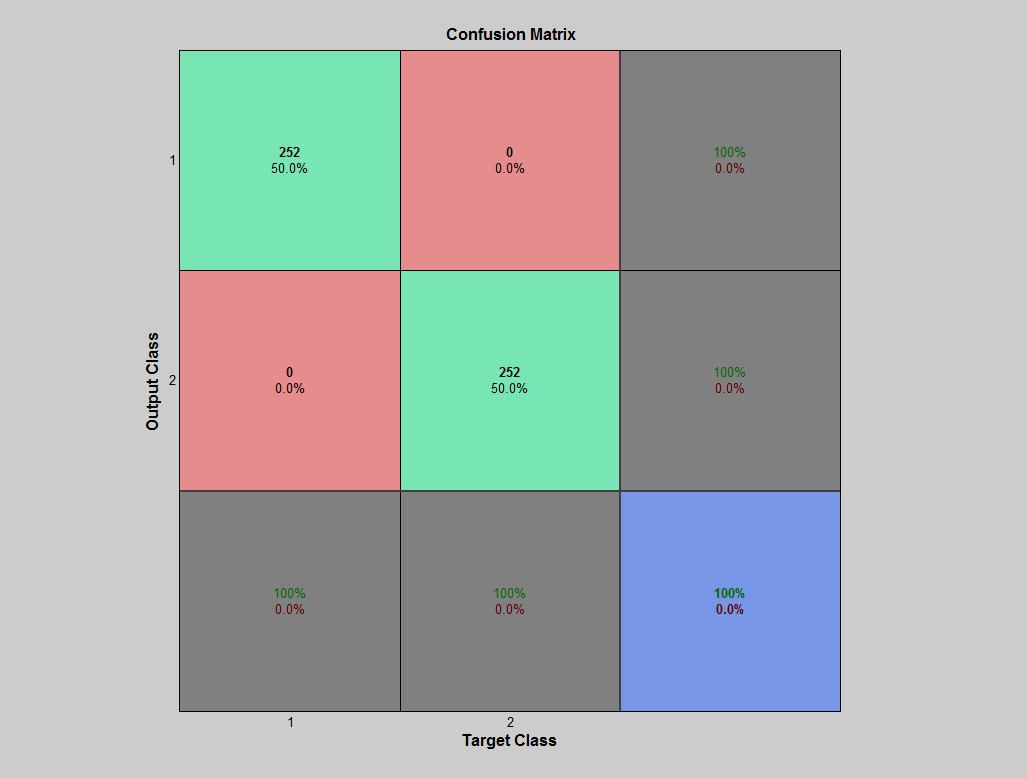
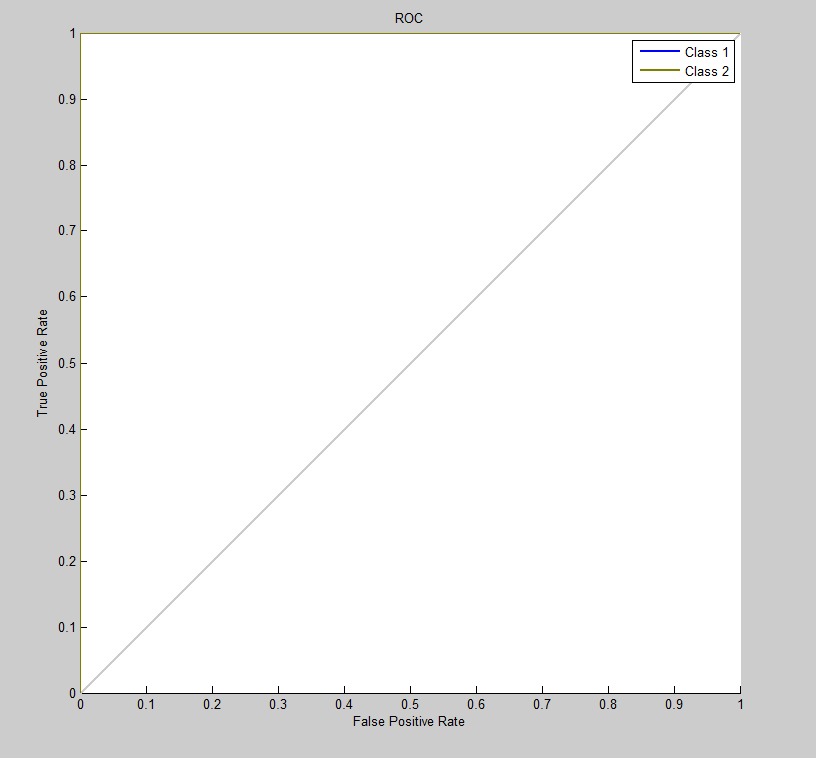
* 1. **Evaluating Controlled Environment**
  2. **Attacks results**
  3. **UDP Attack on cloud computing**
     1. **PCA**



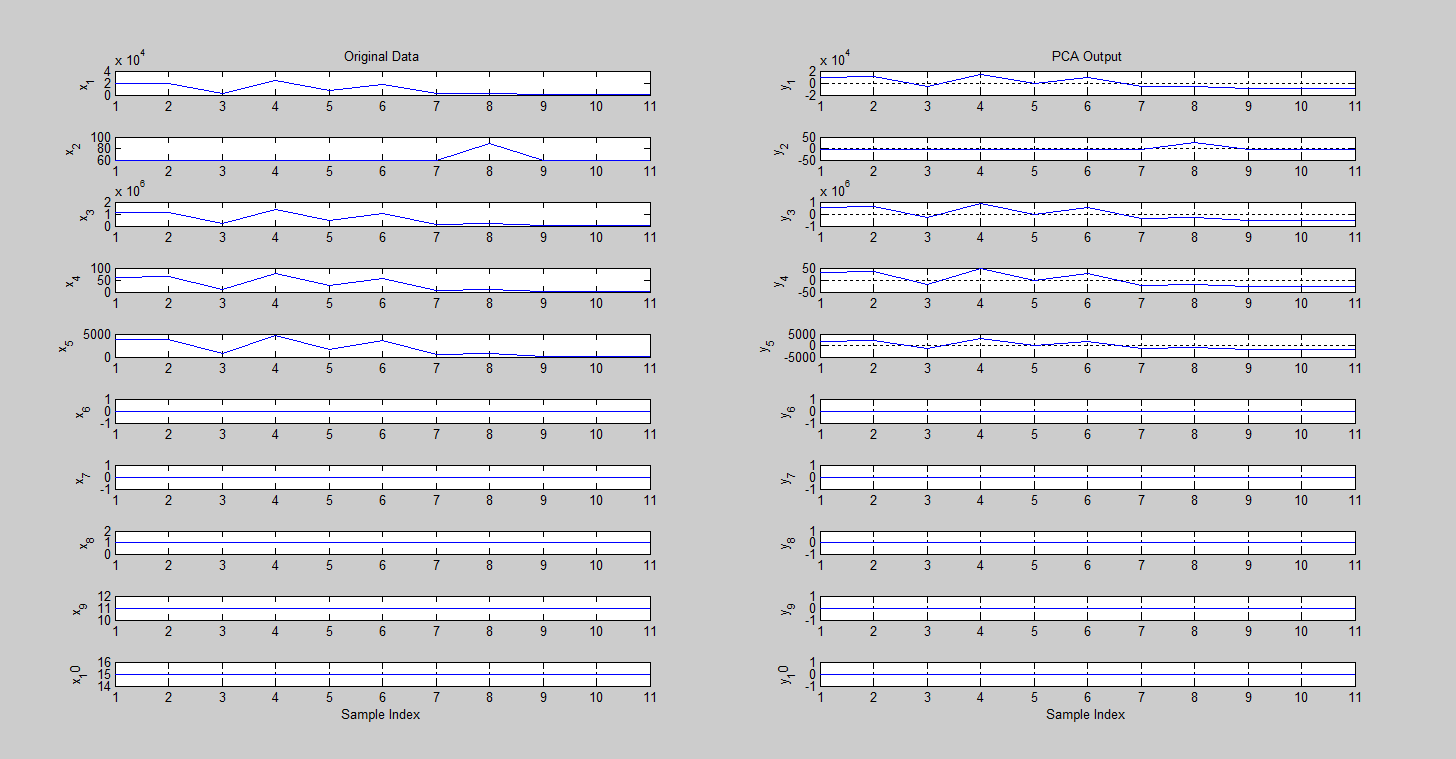
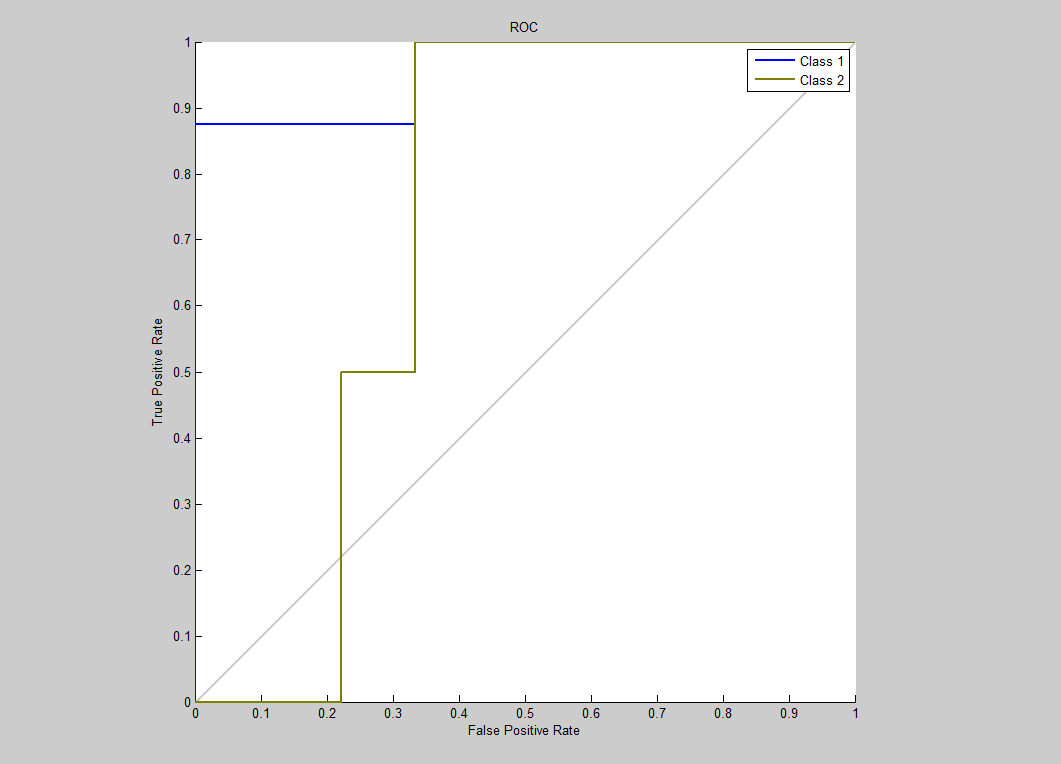
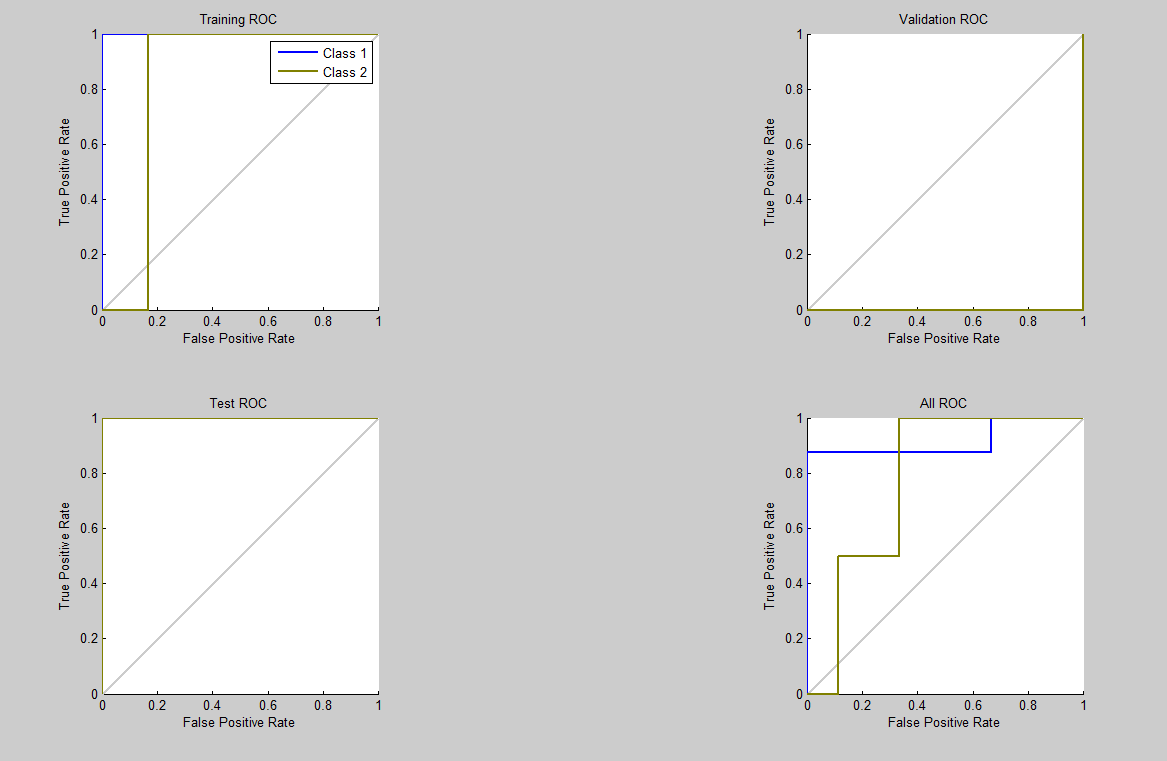
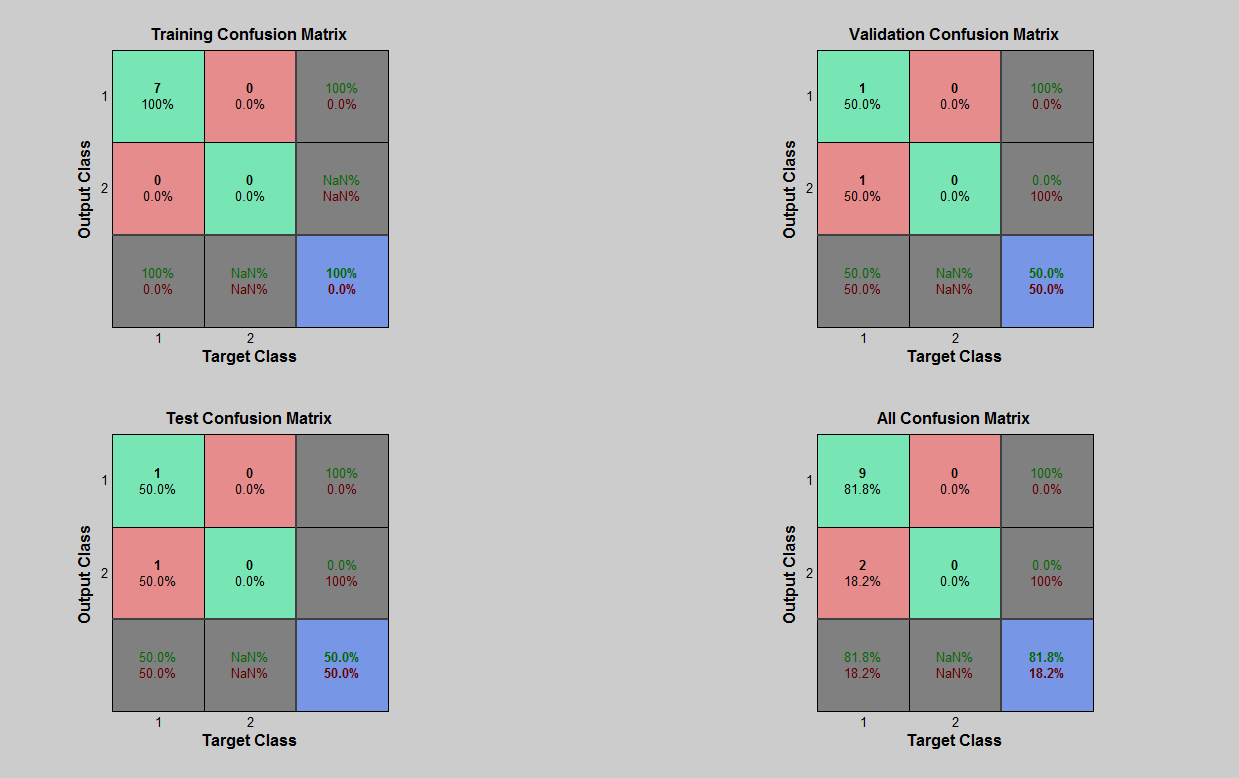
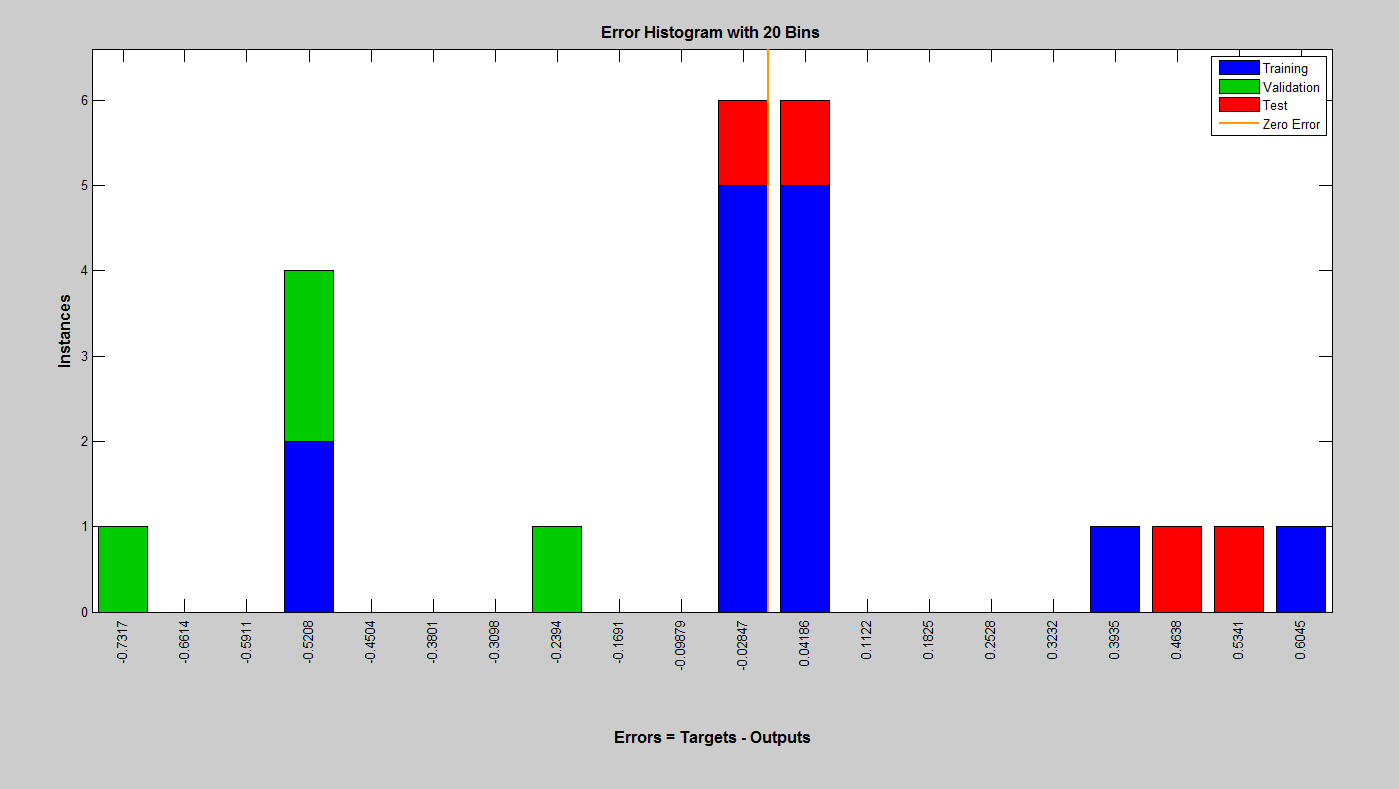
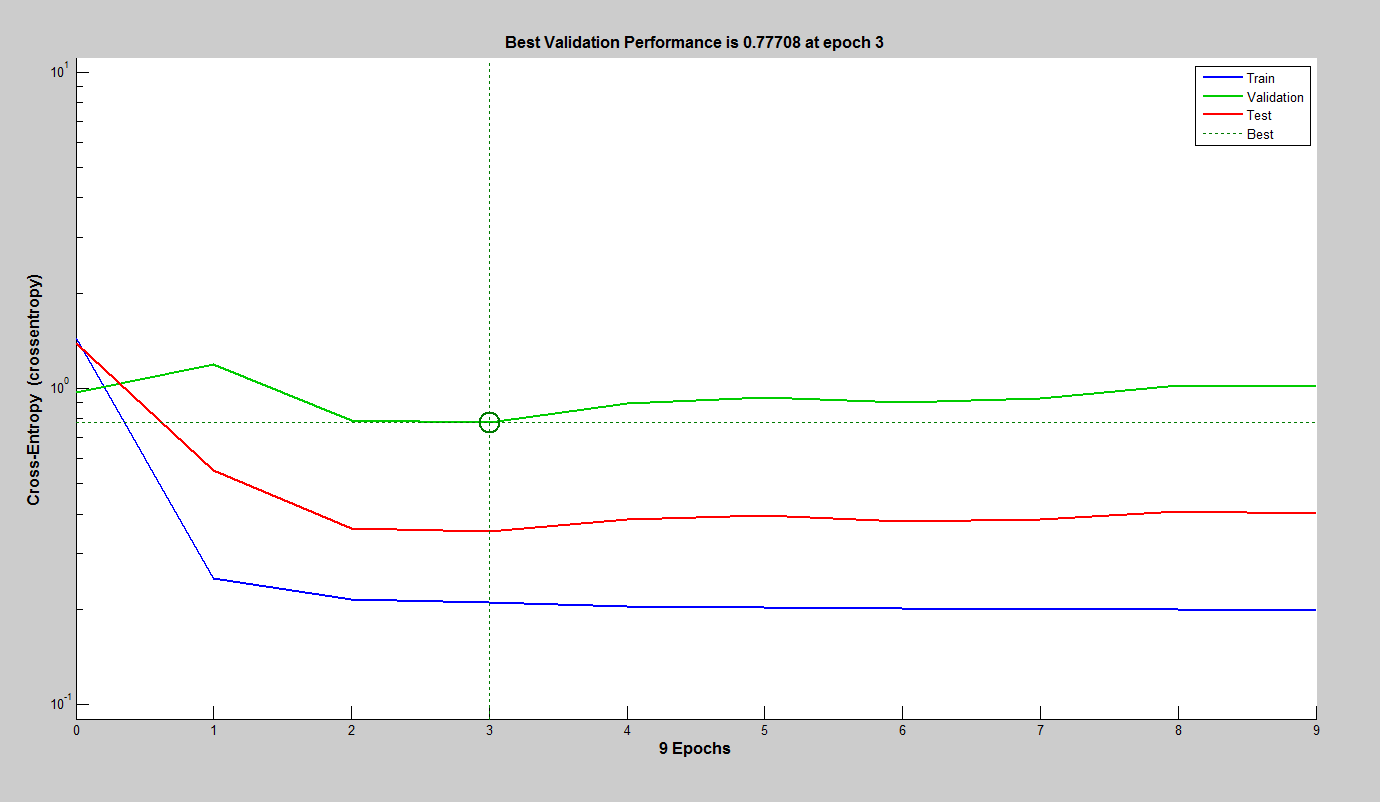
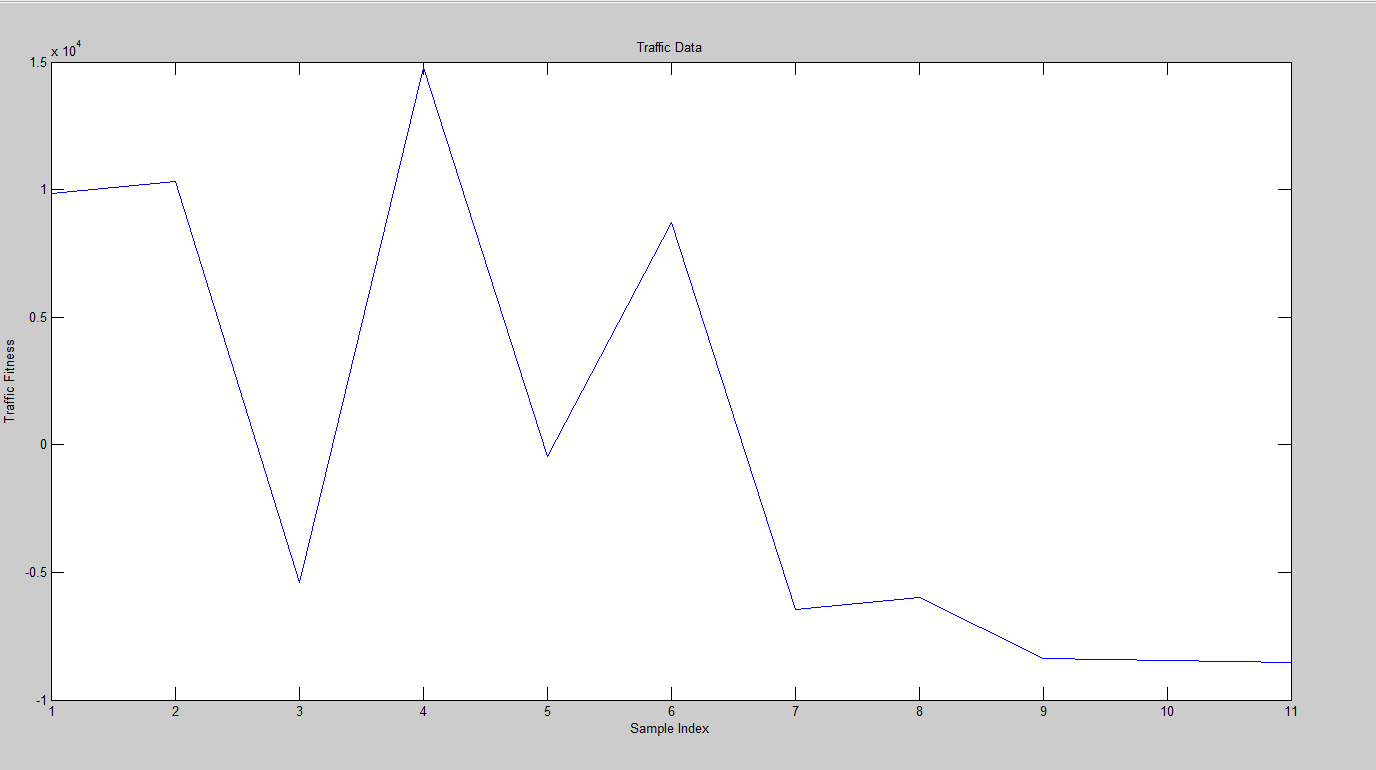
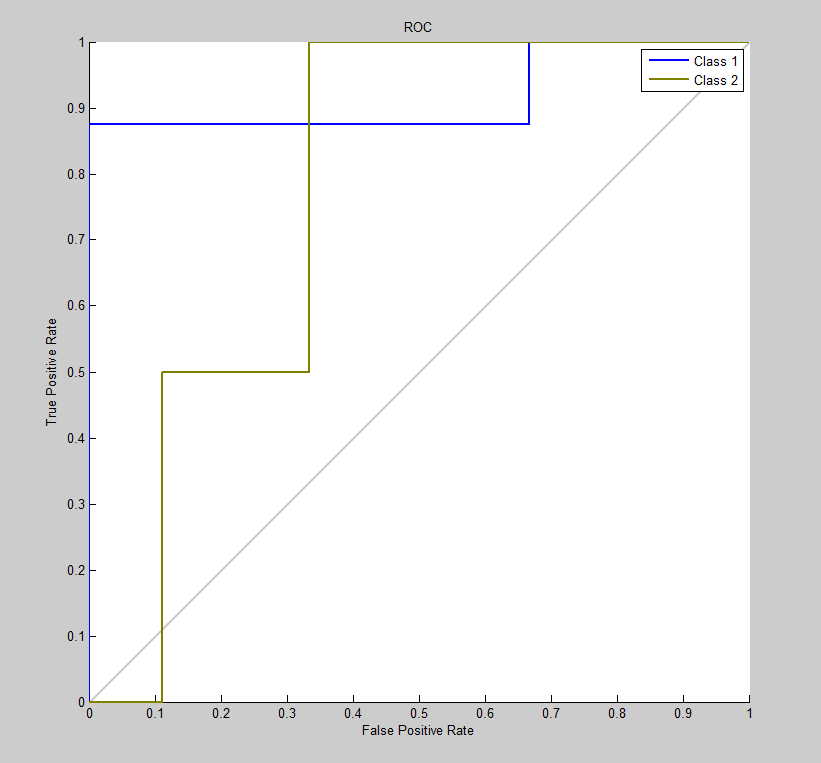
* + 1. **LDA**



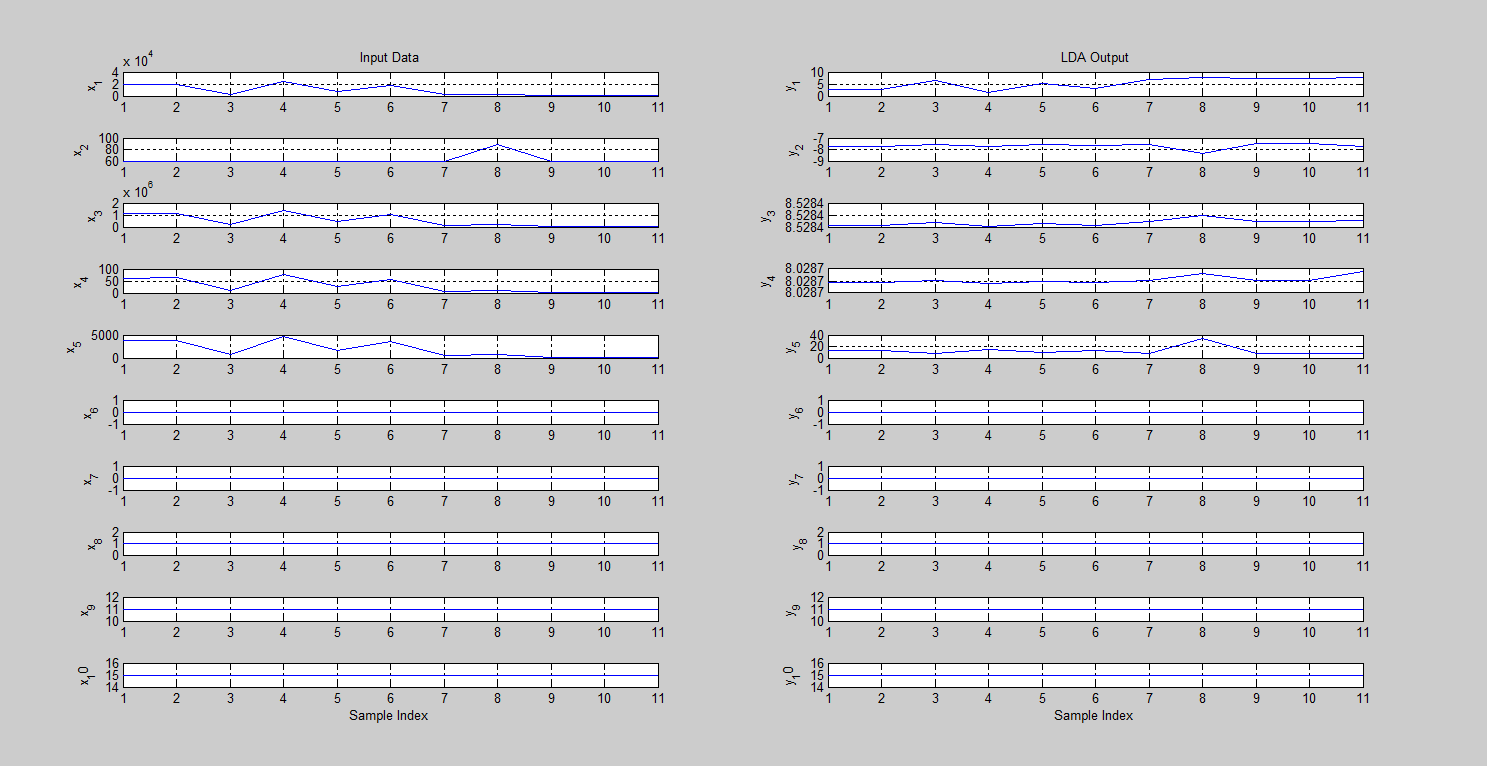
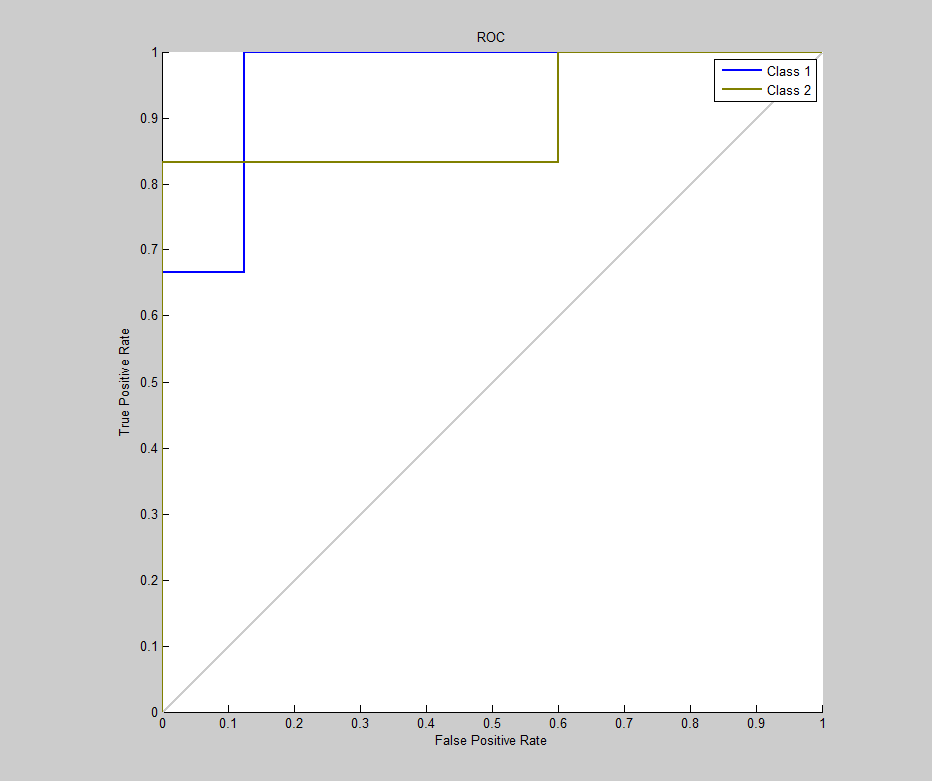
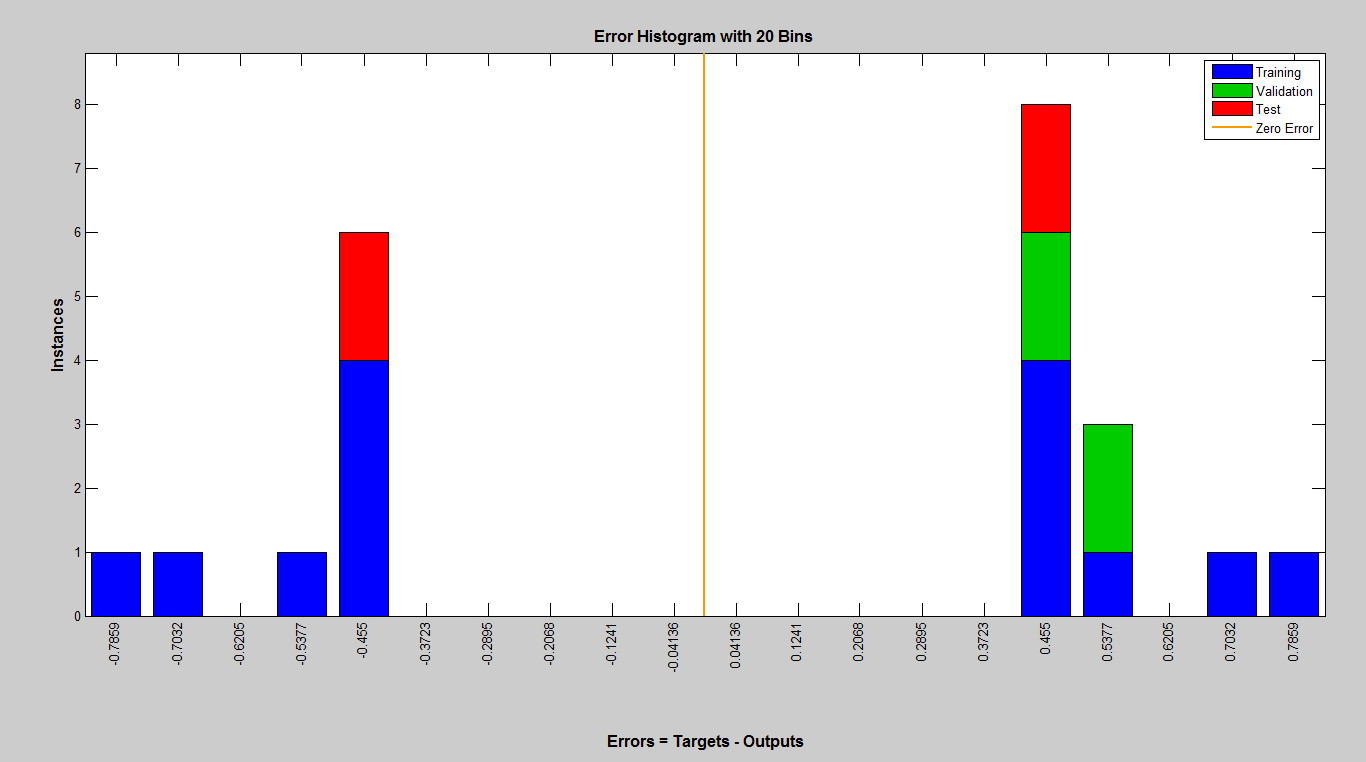
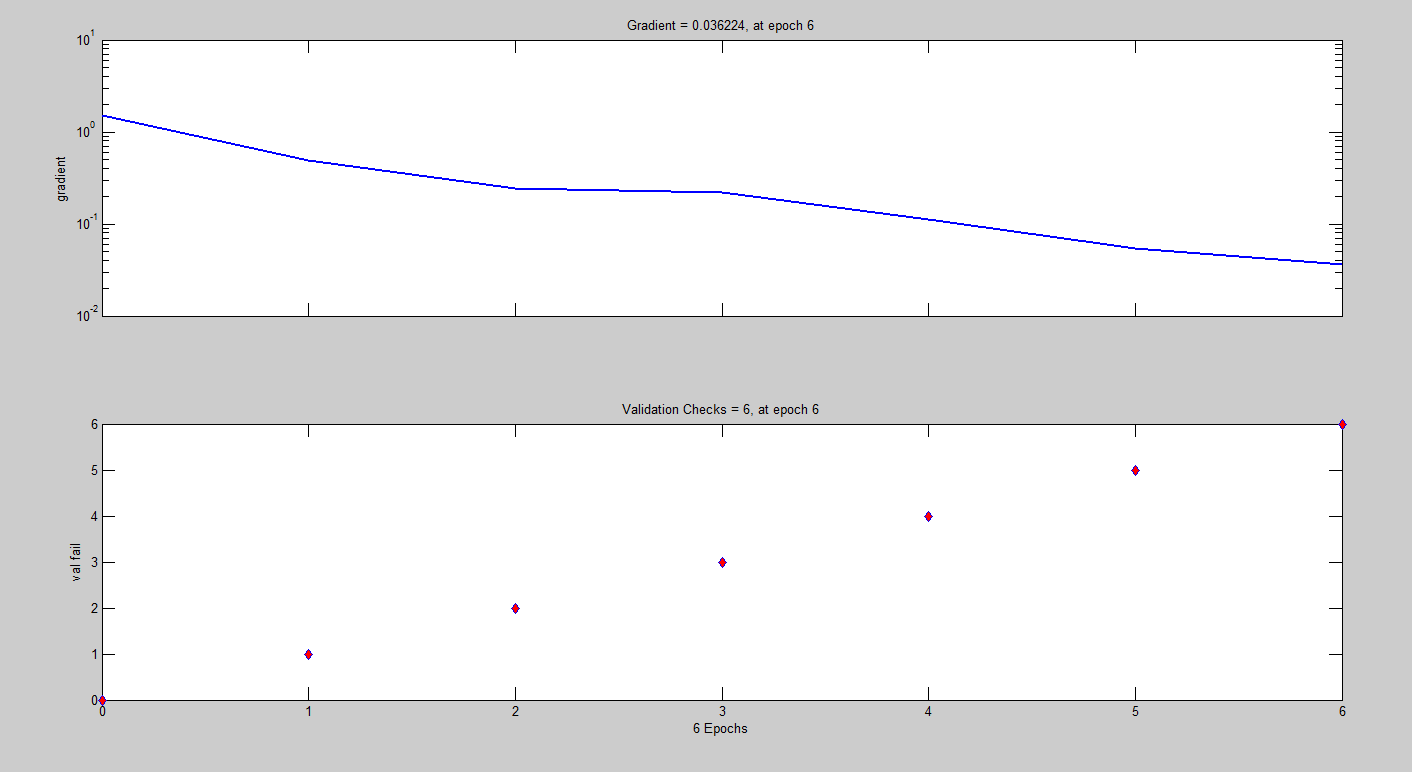
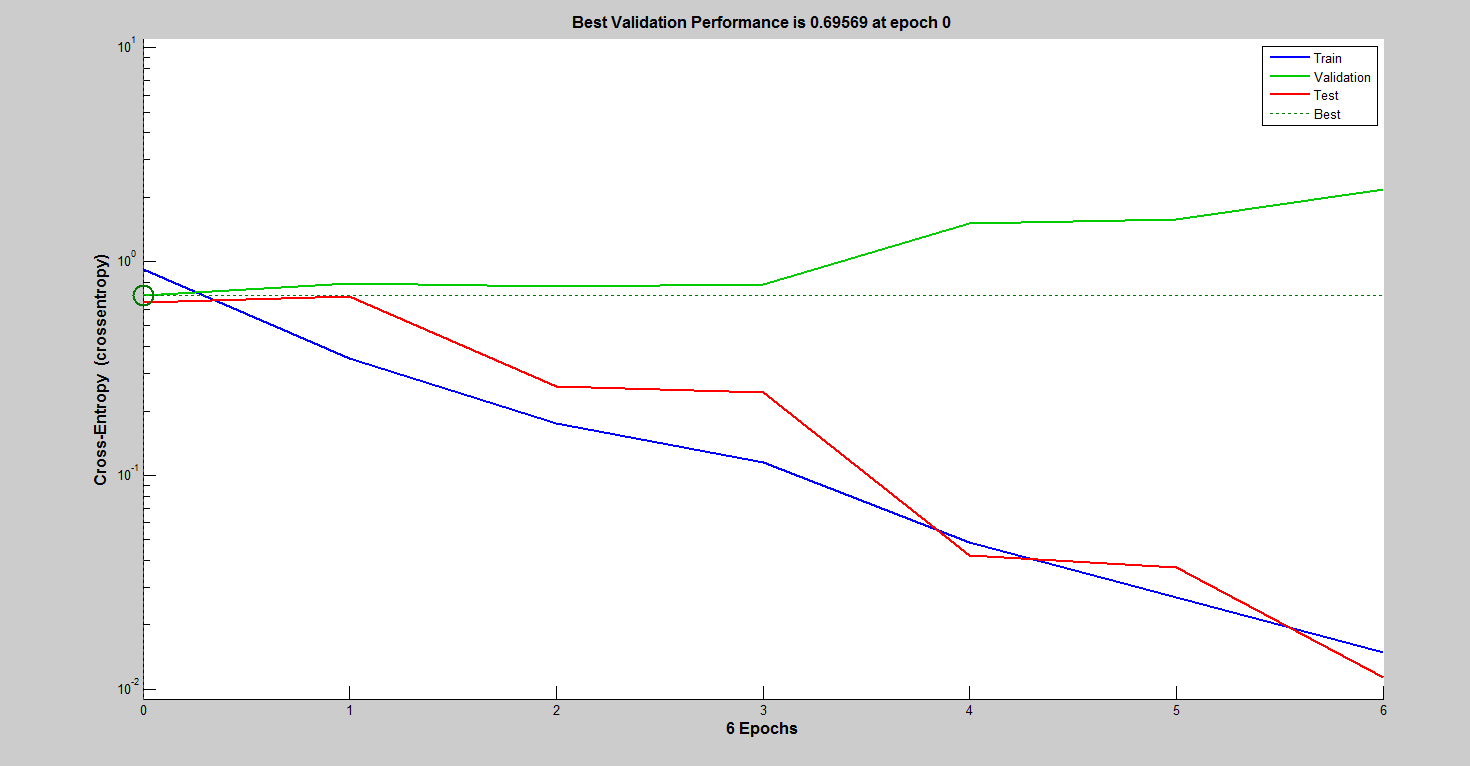
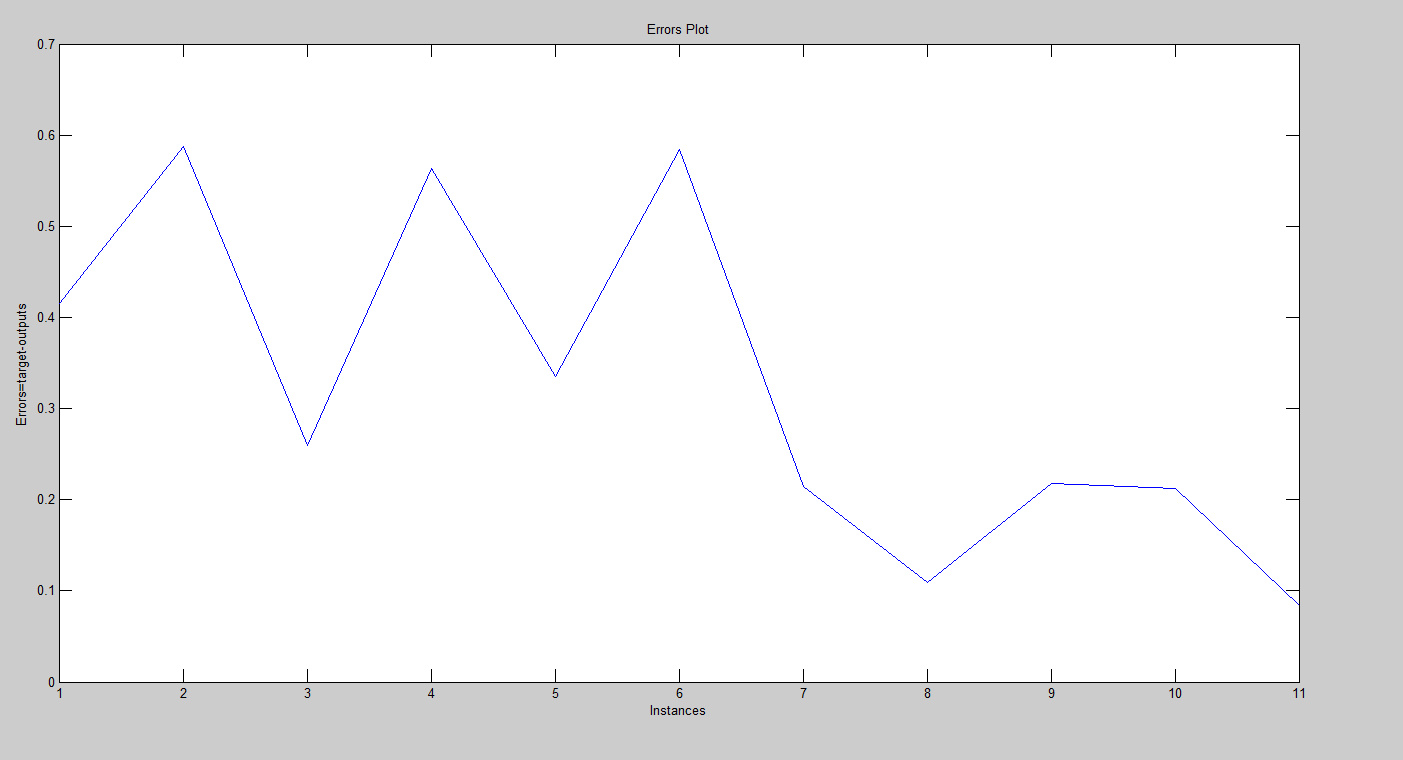
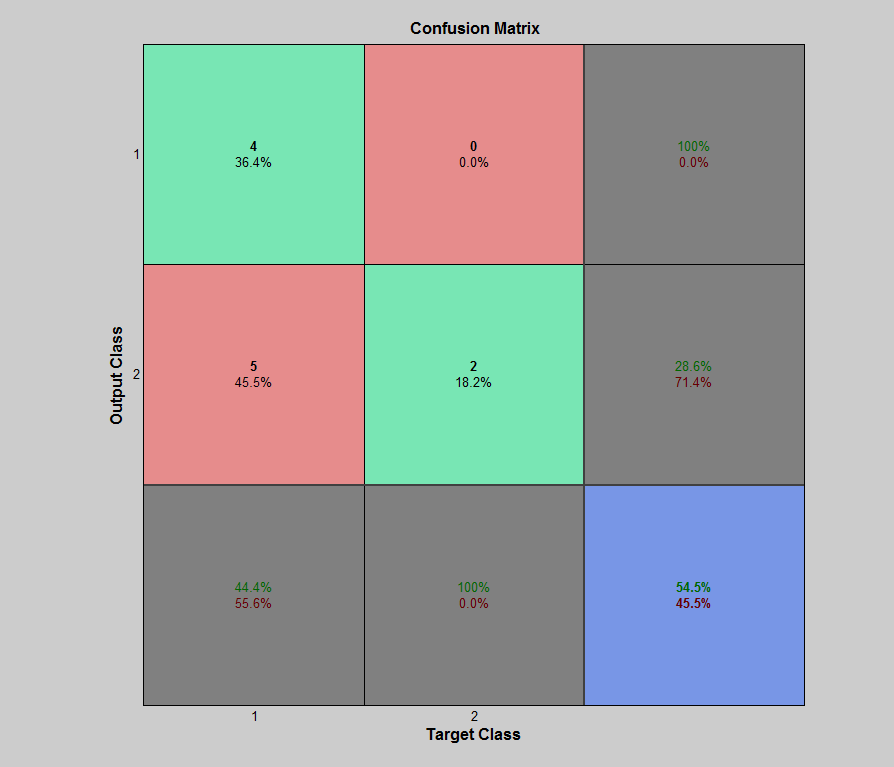
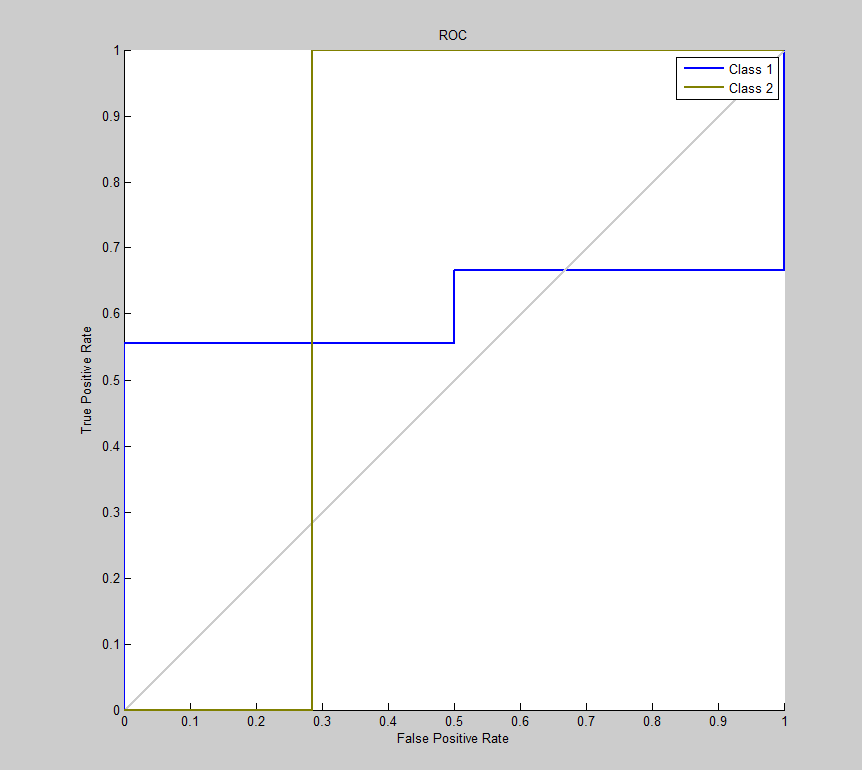
* + 1. **Proposed model (PCA+LDA+ALO)**



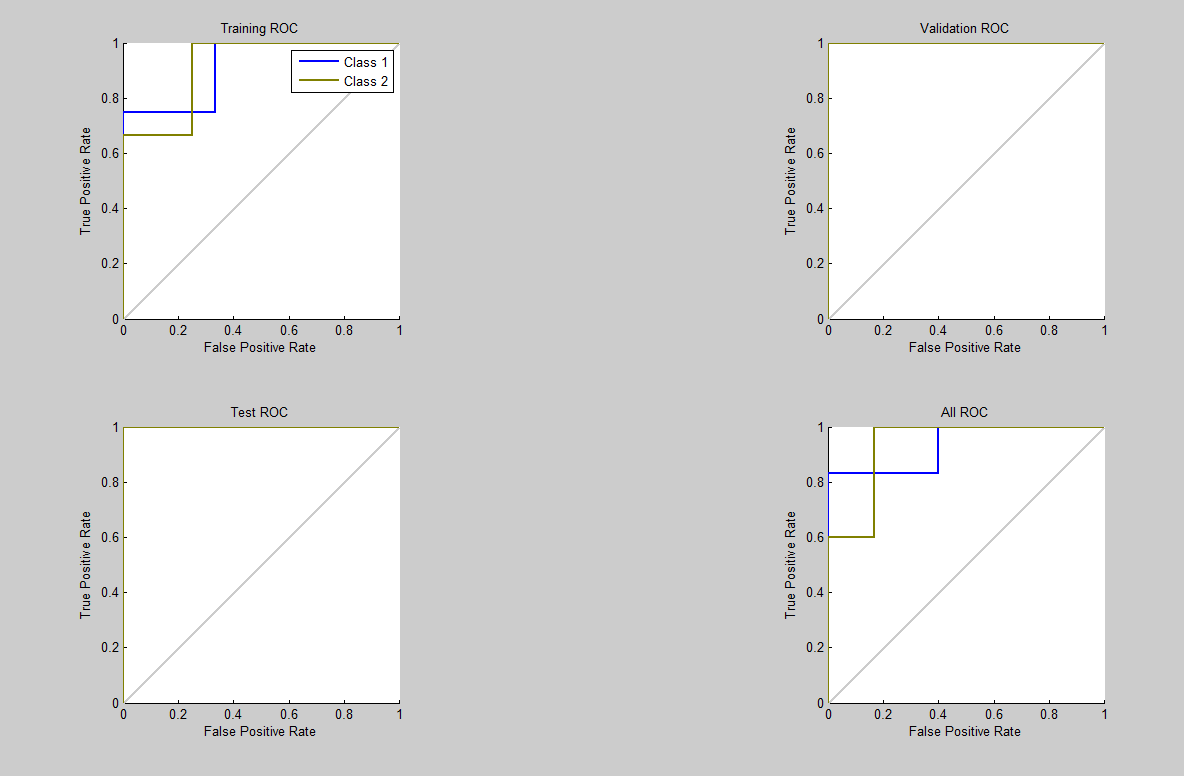
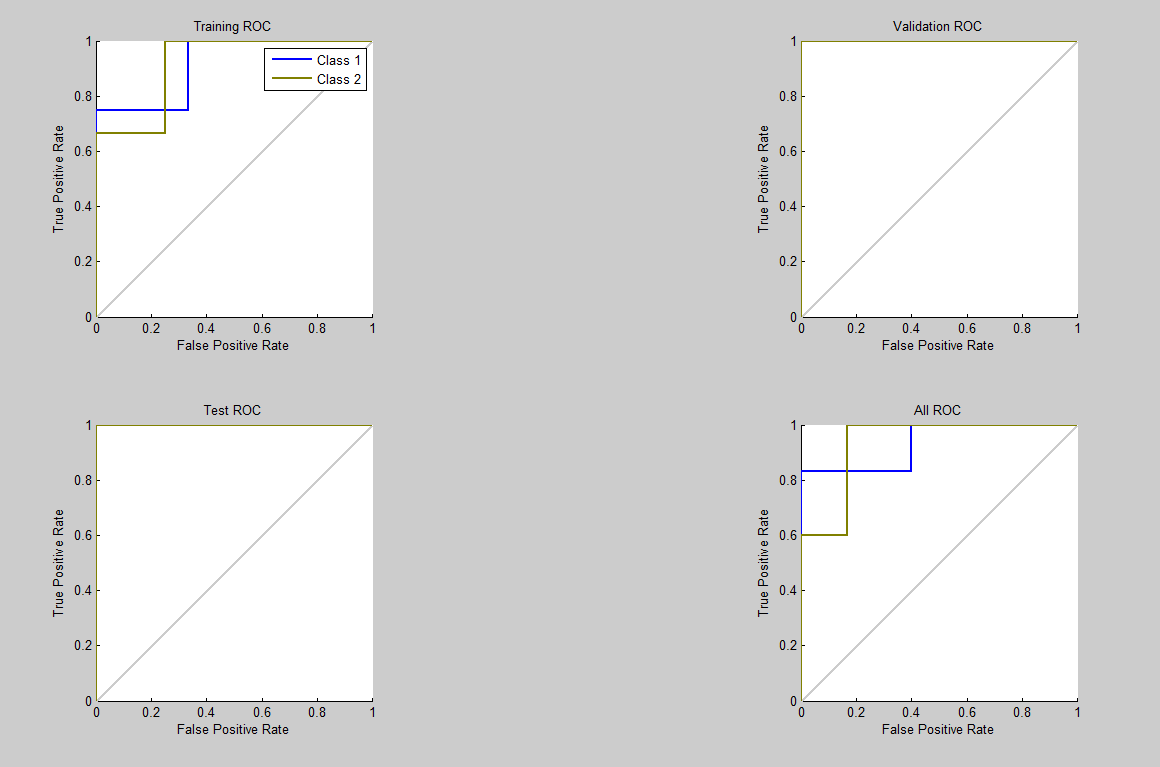
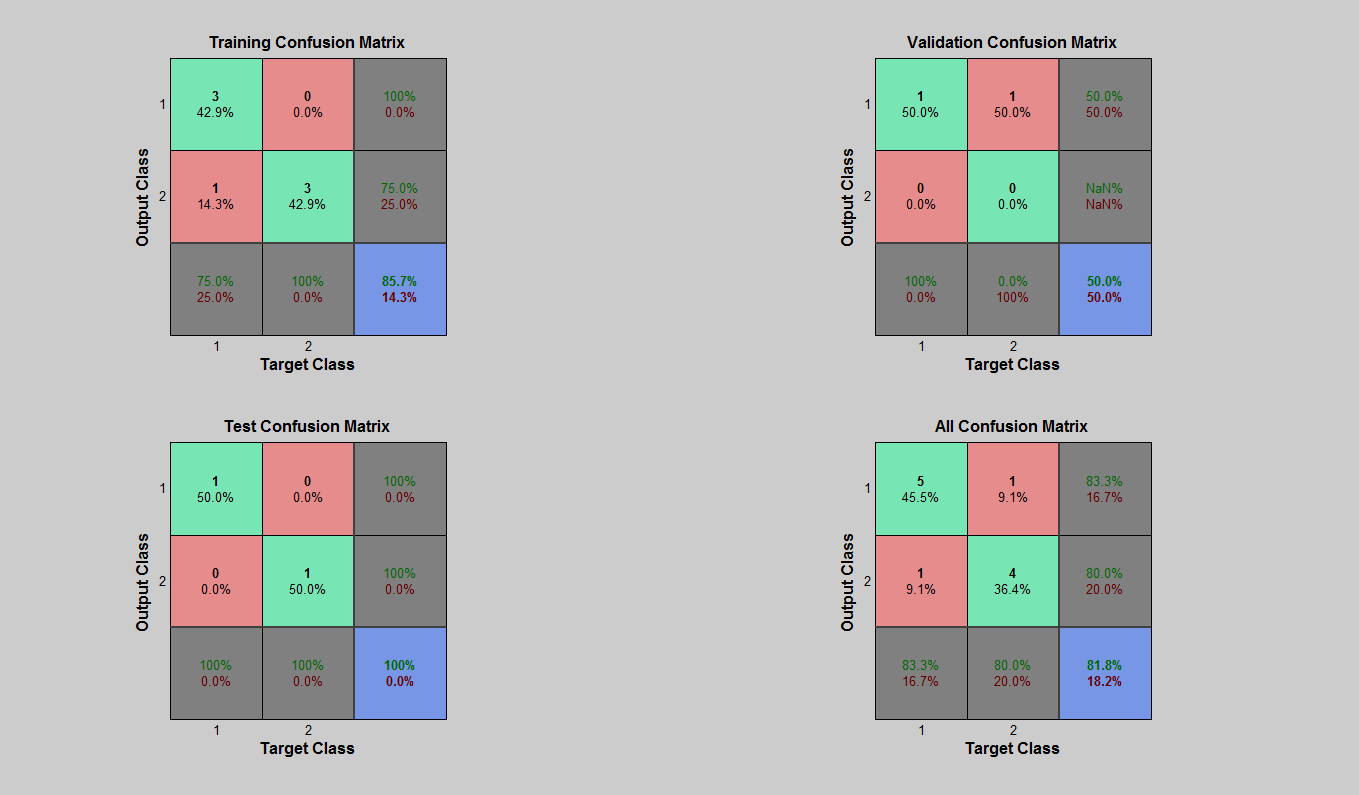
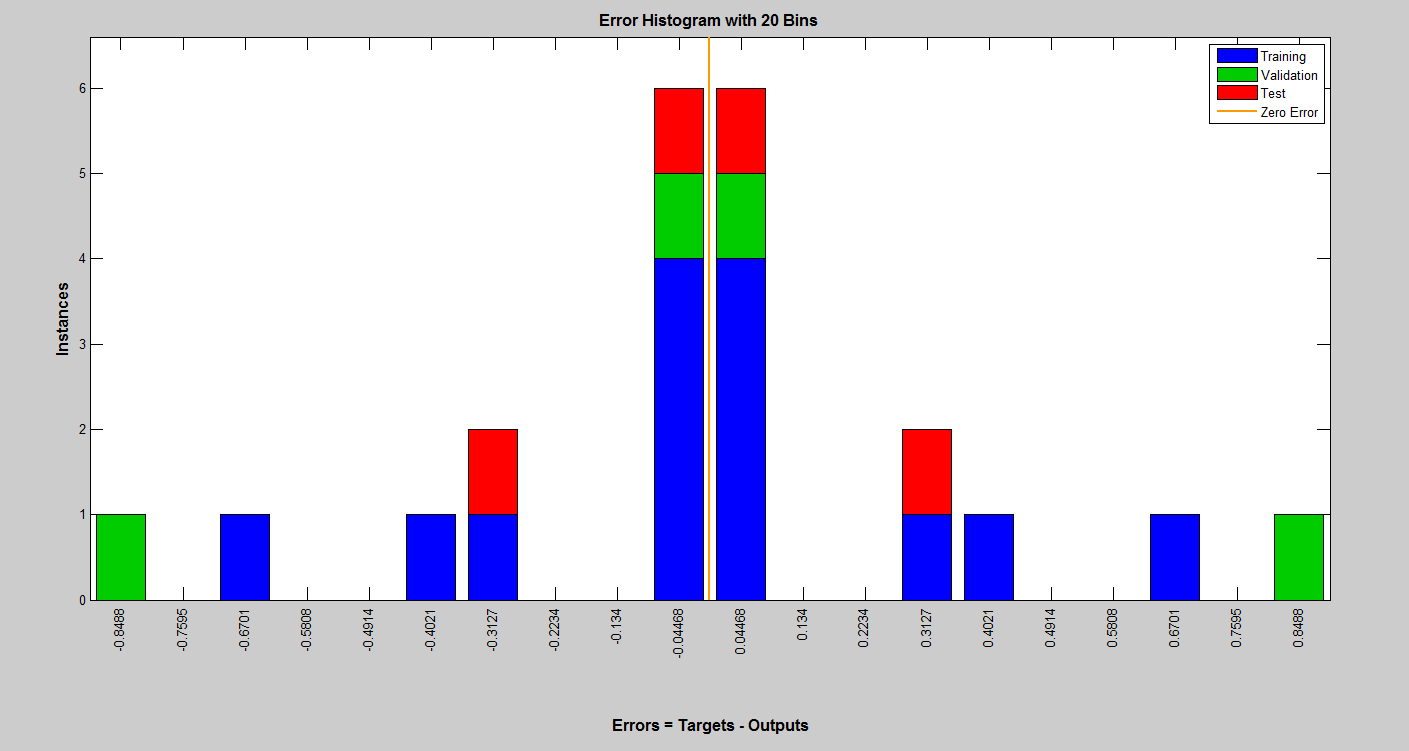
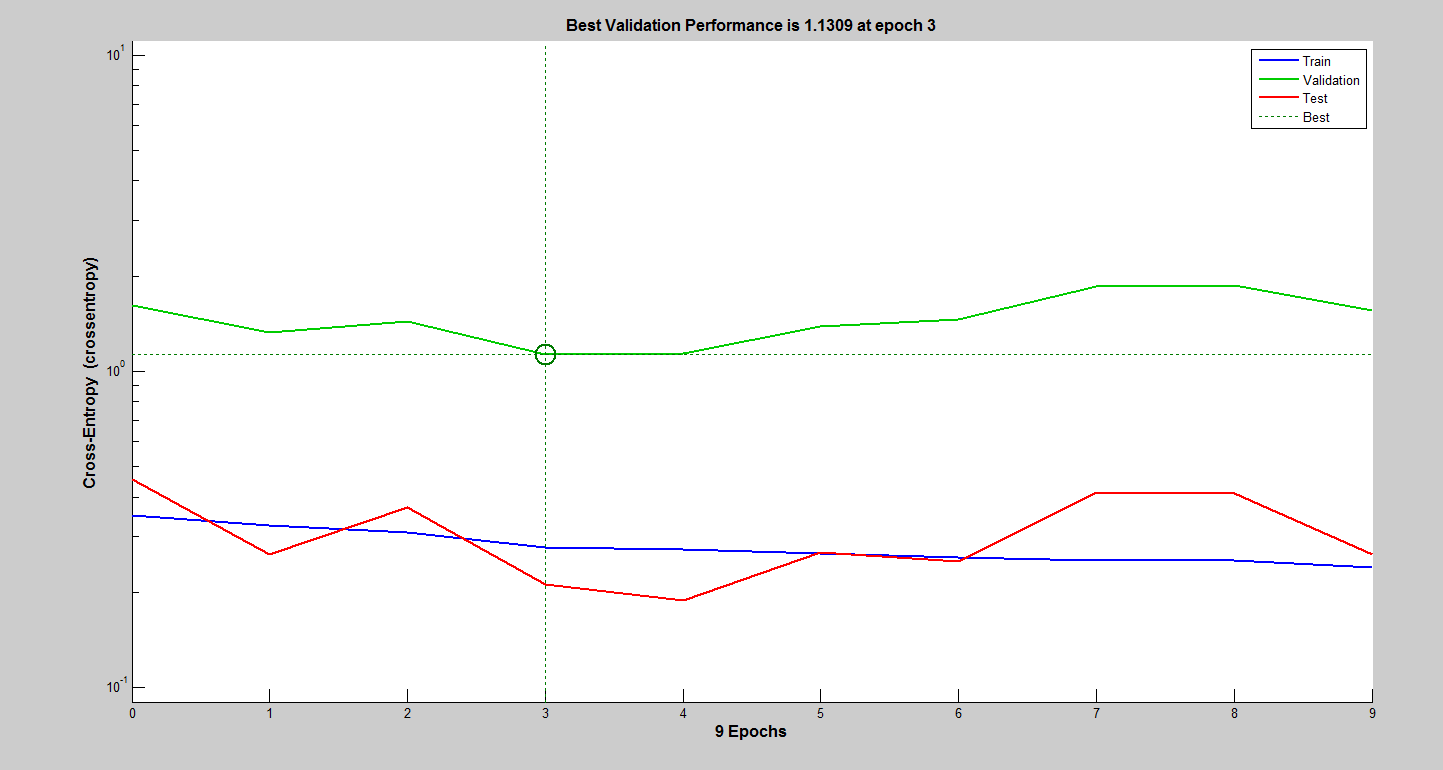
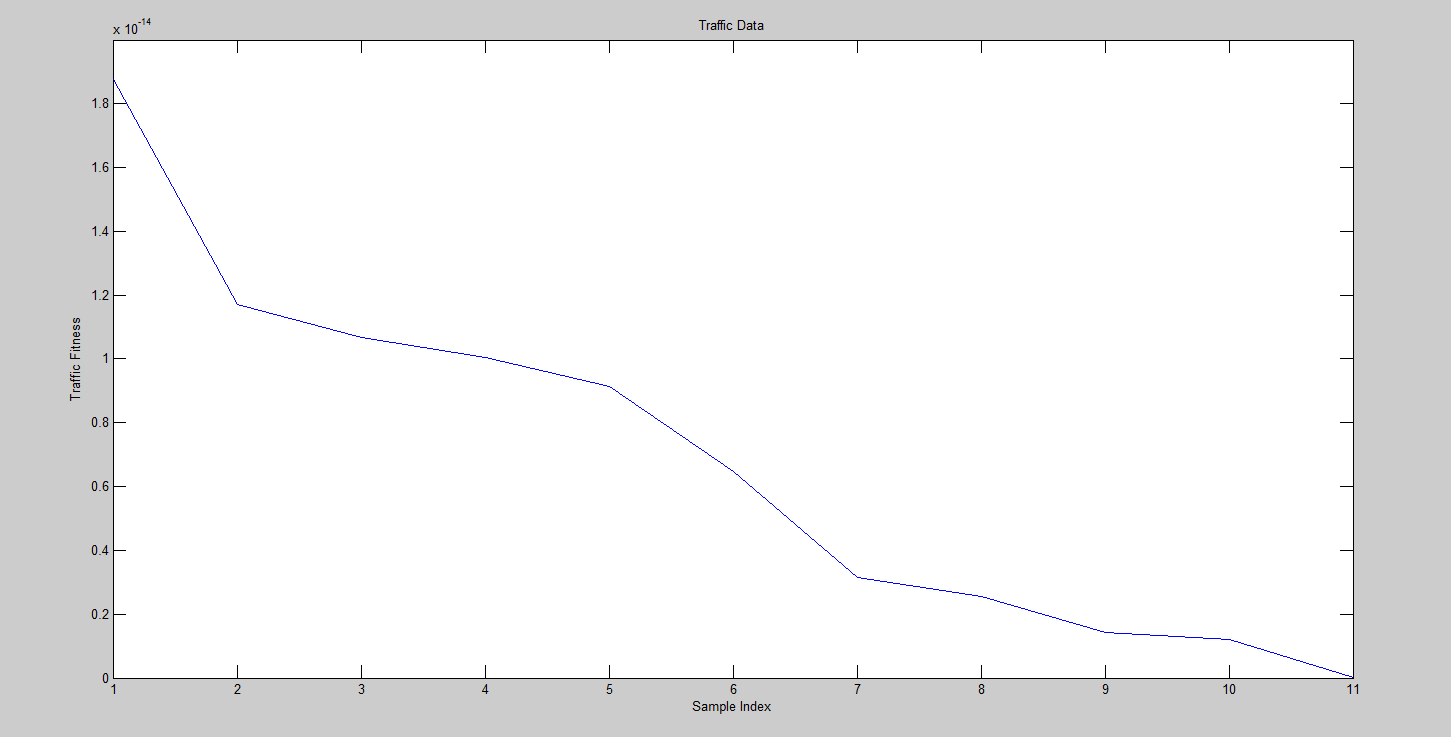
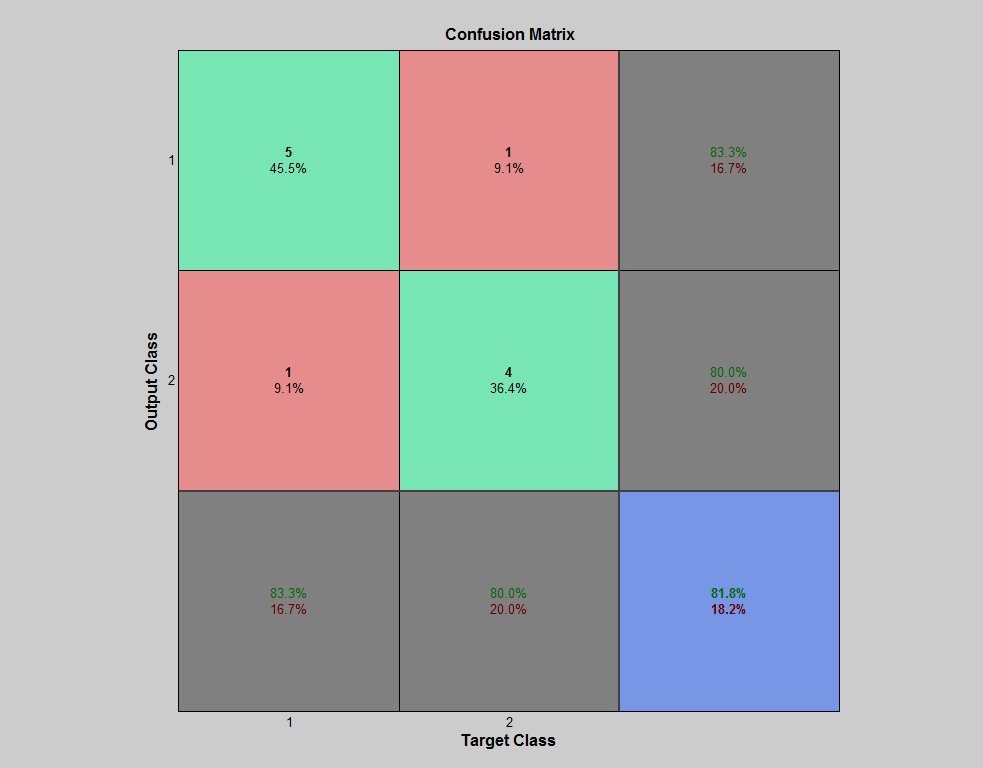
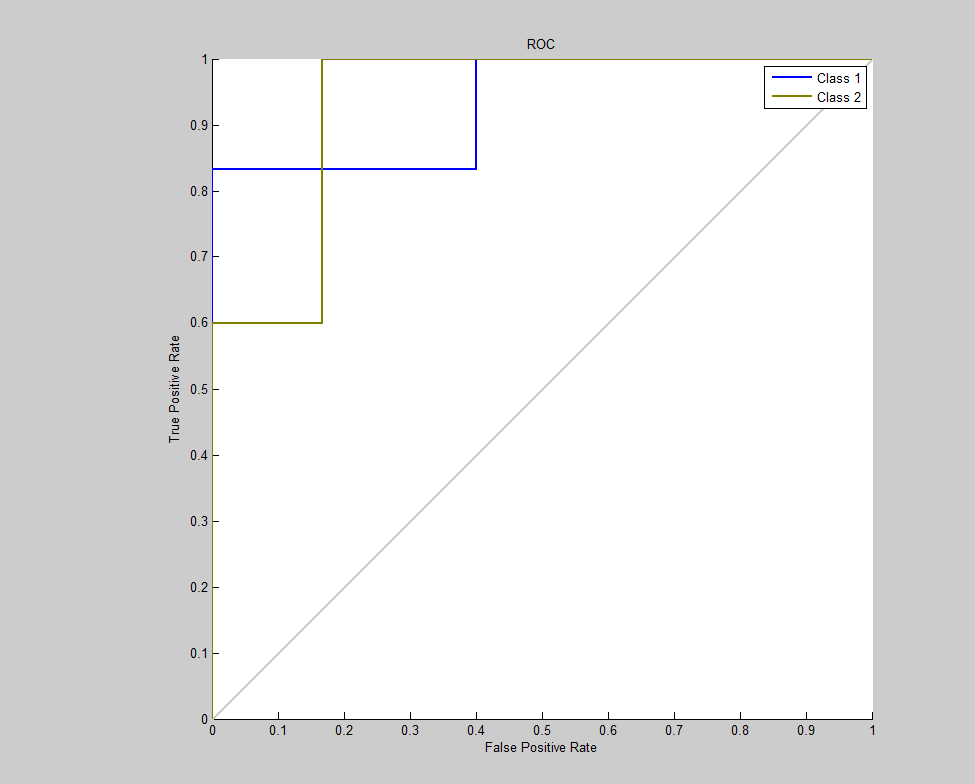
* 1. **ICMP Attack on cloud computing**
     1. **PCA**



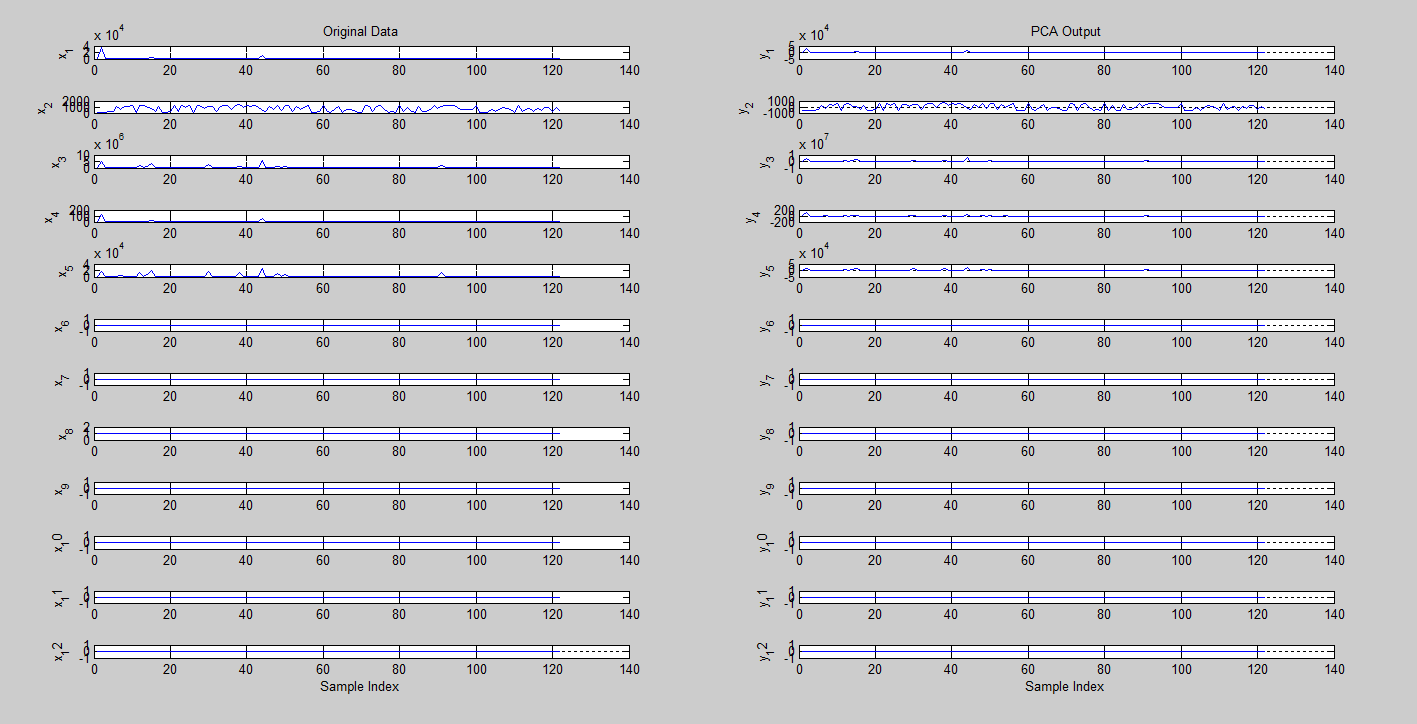
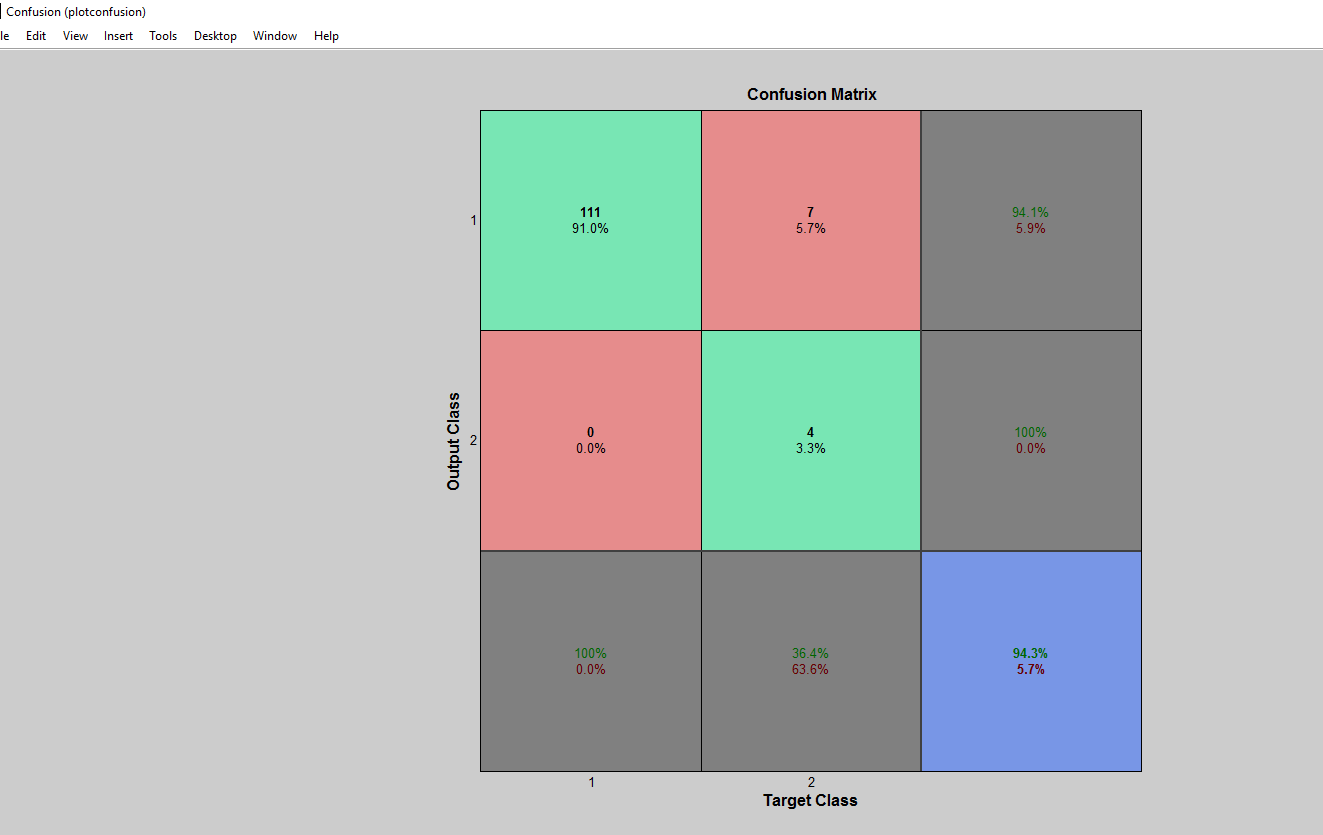
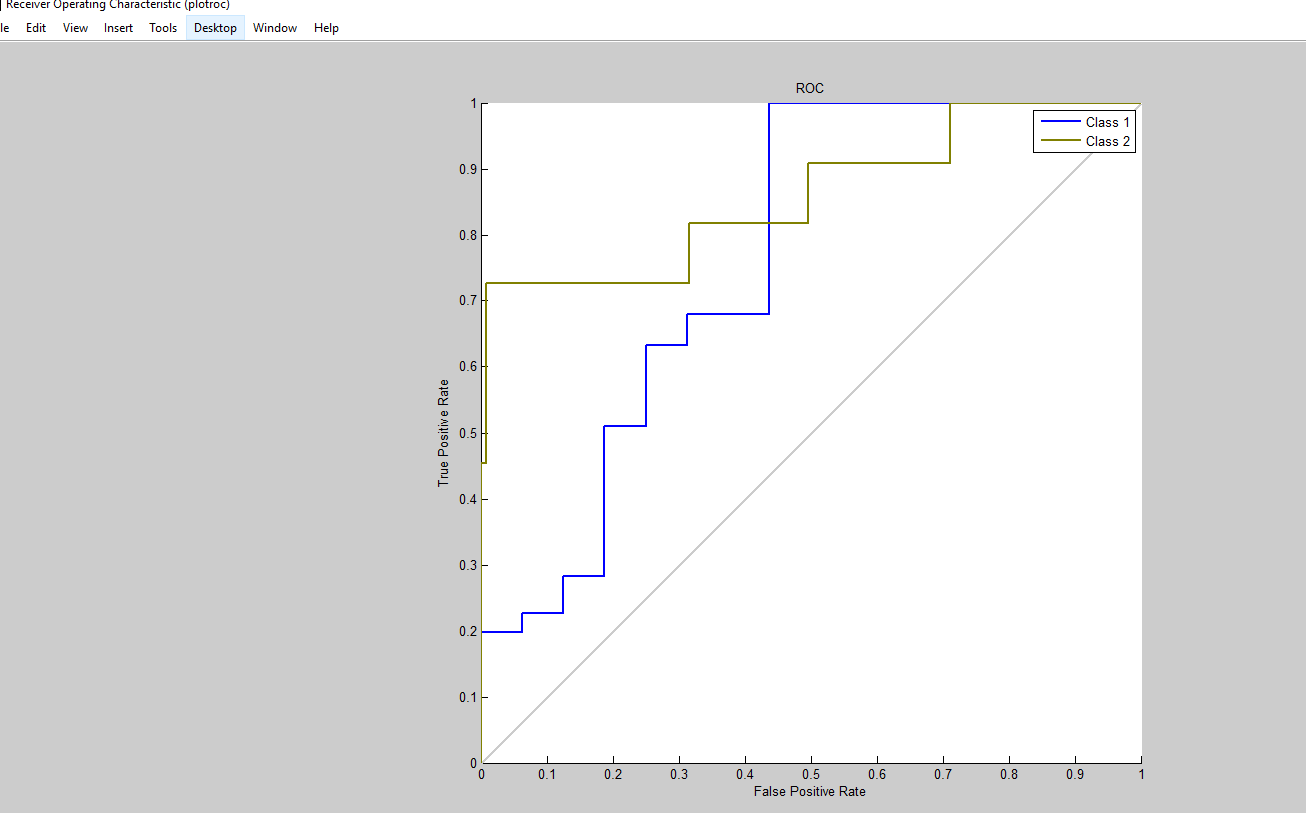
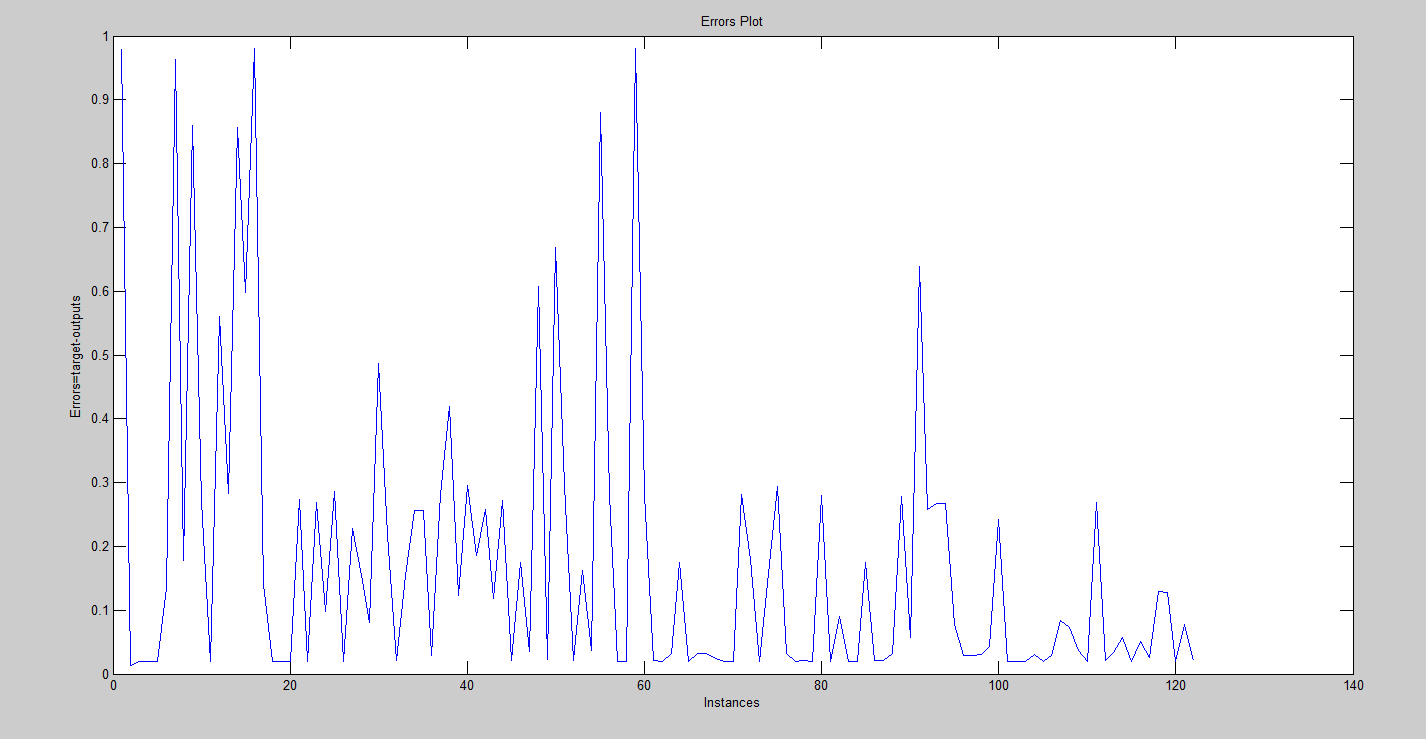
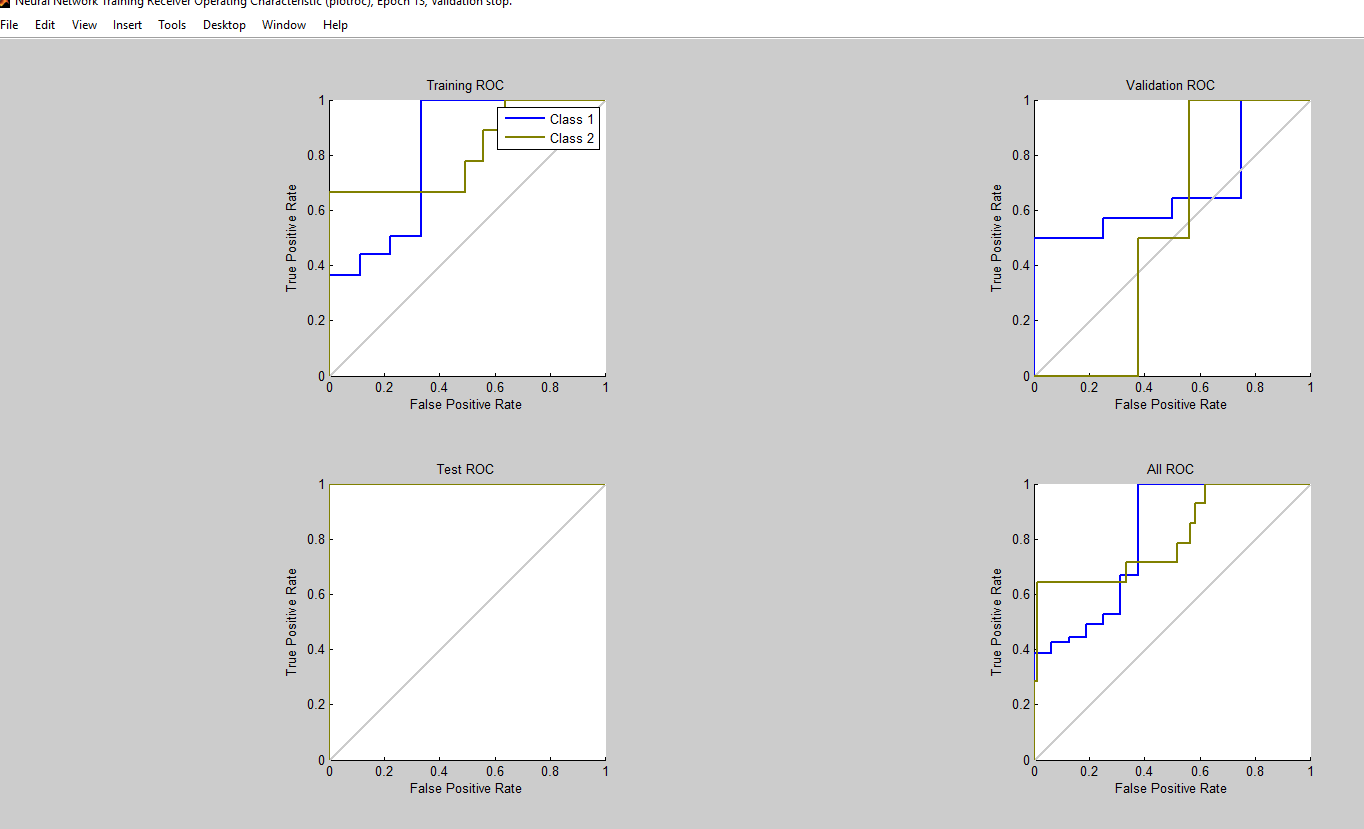
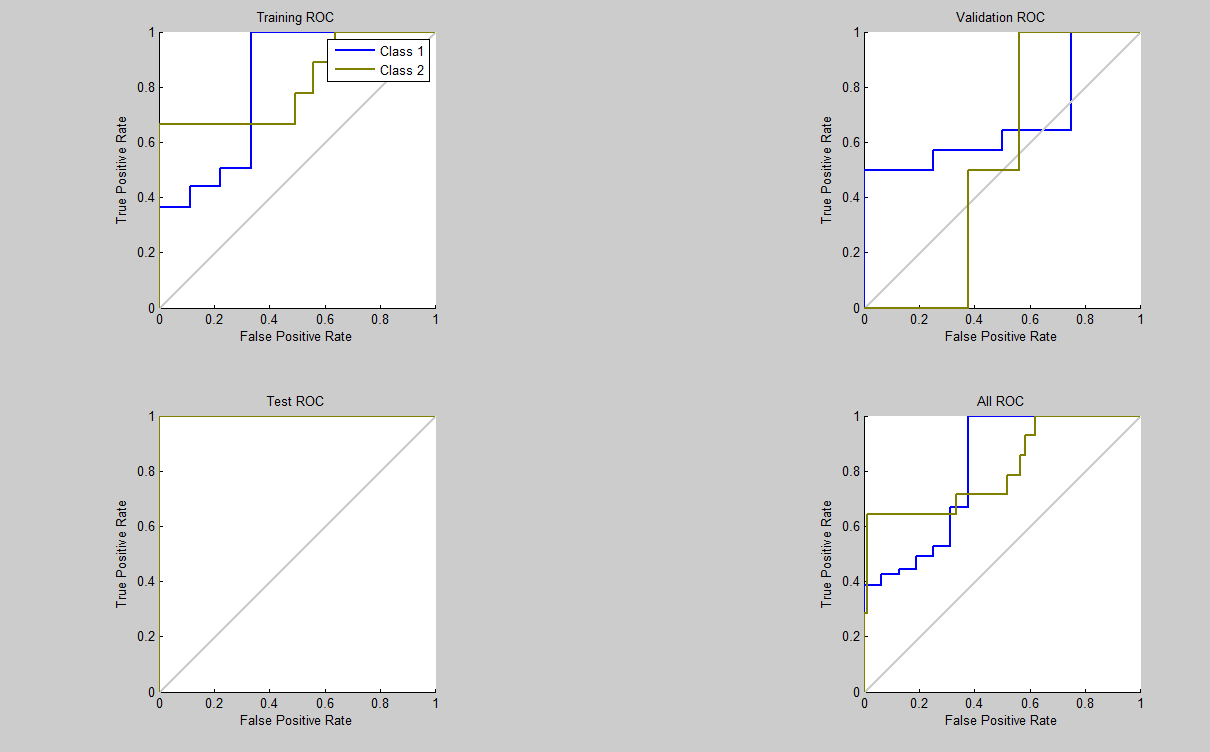
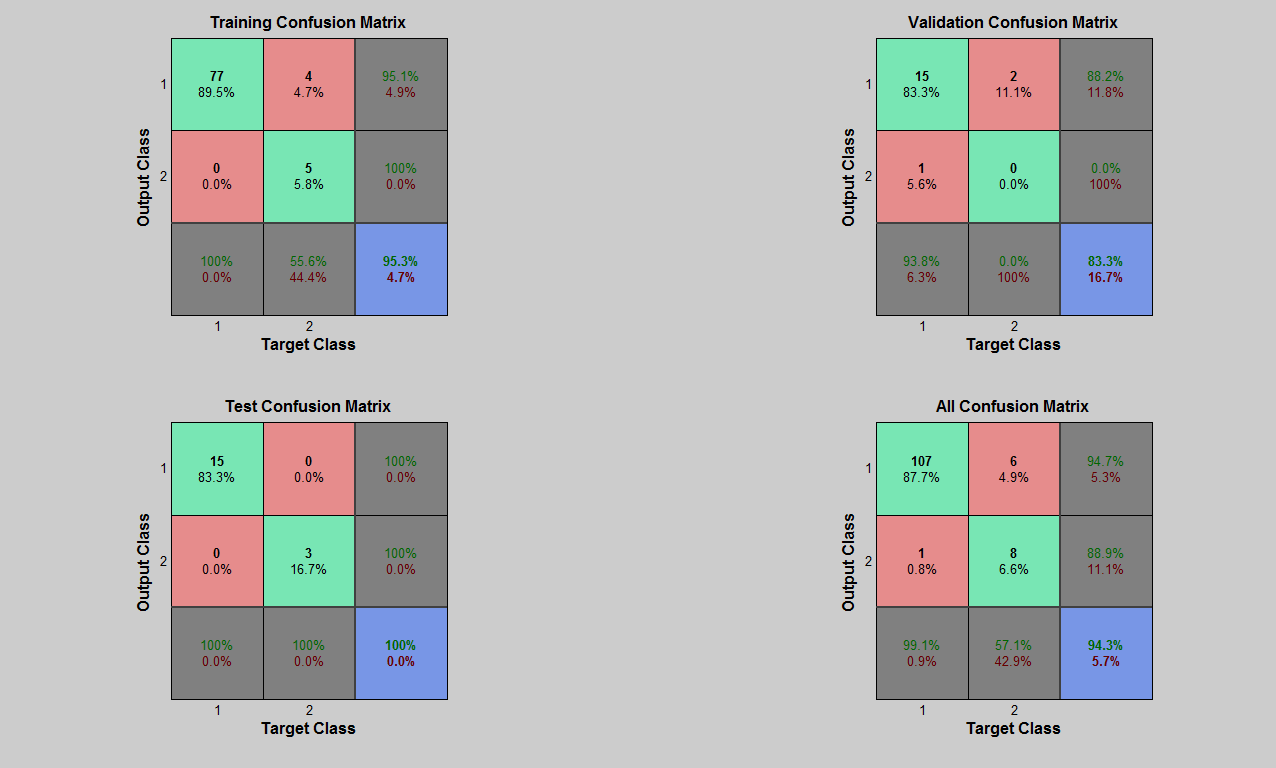
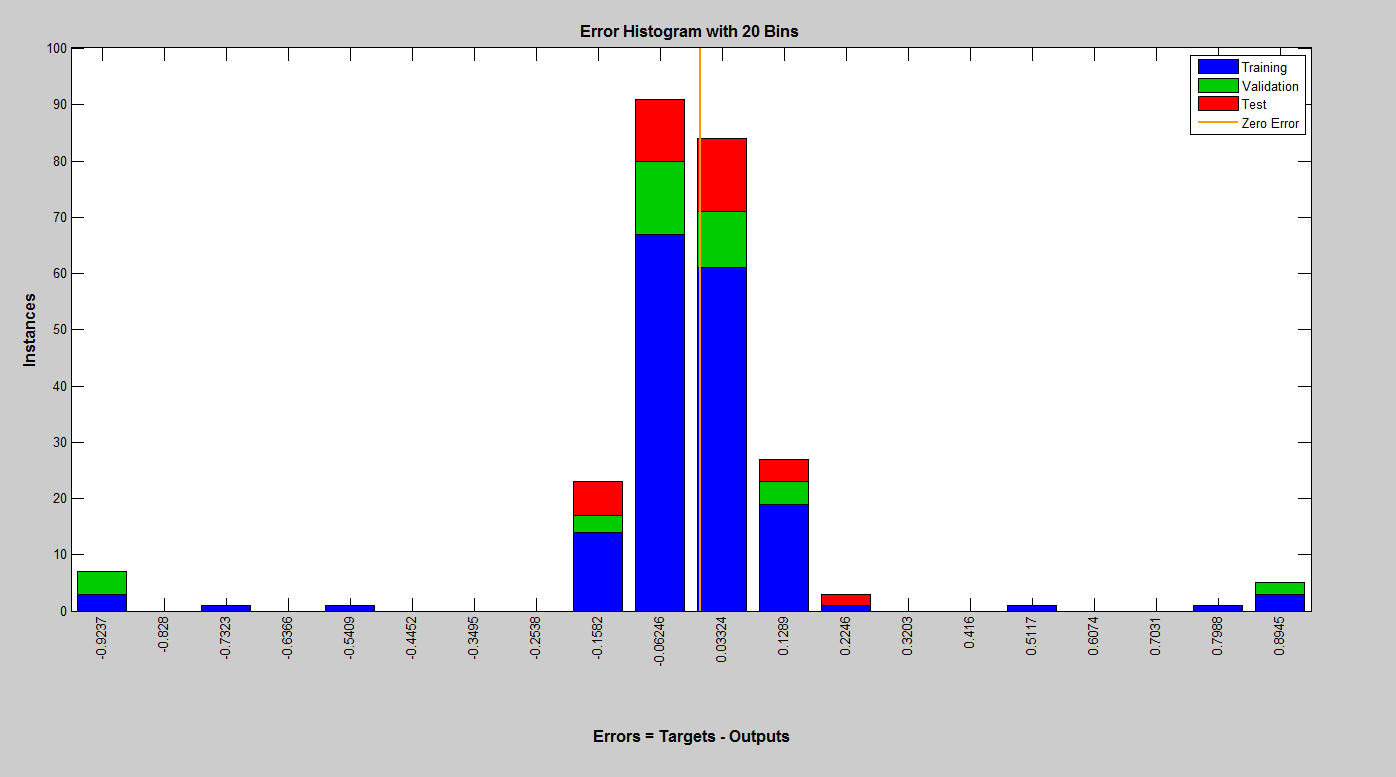
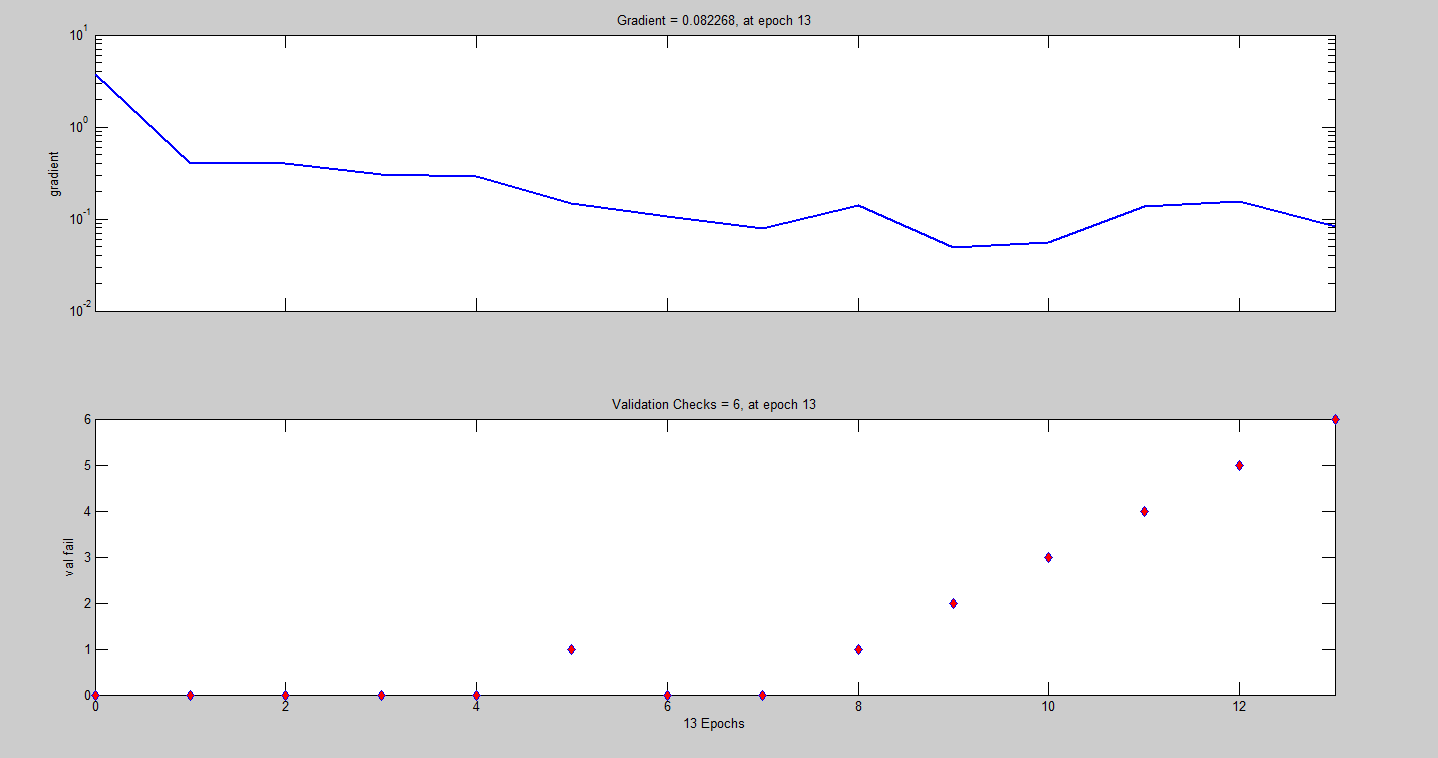
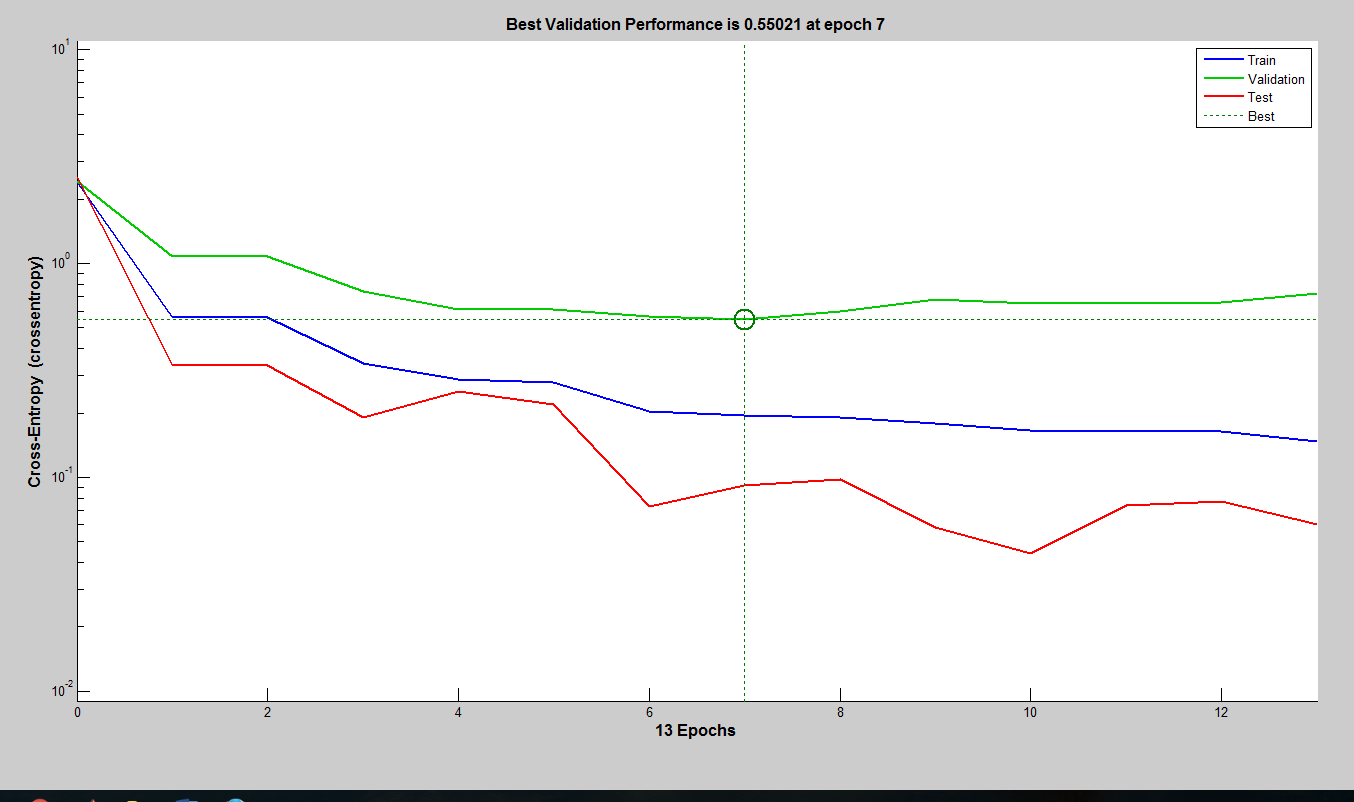
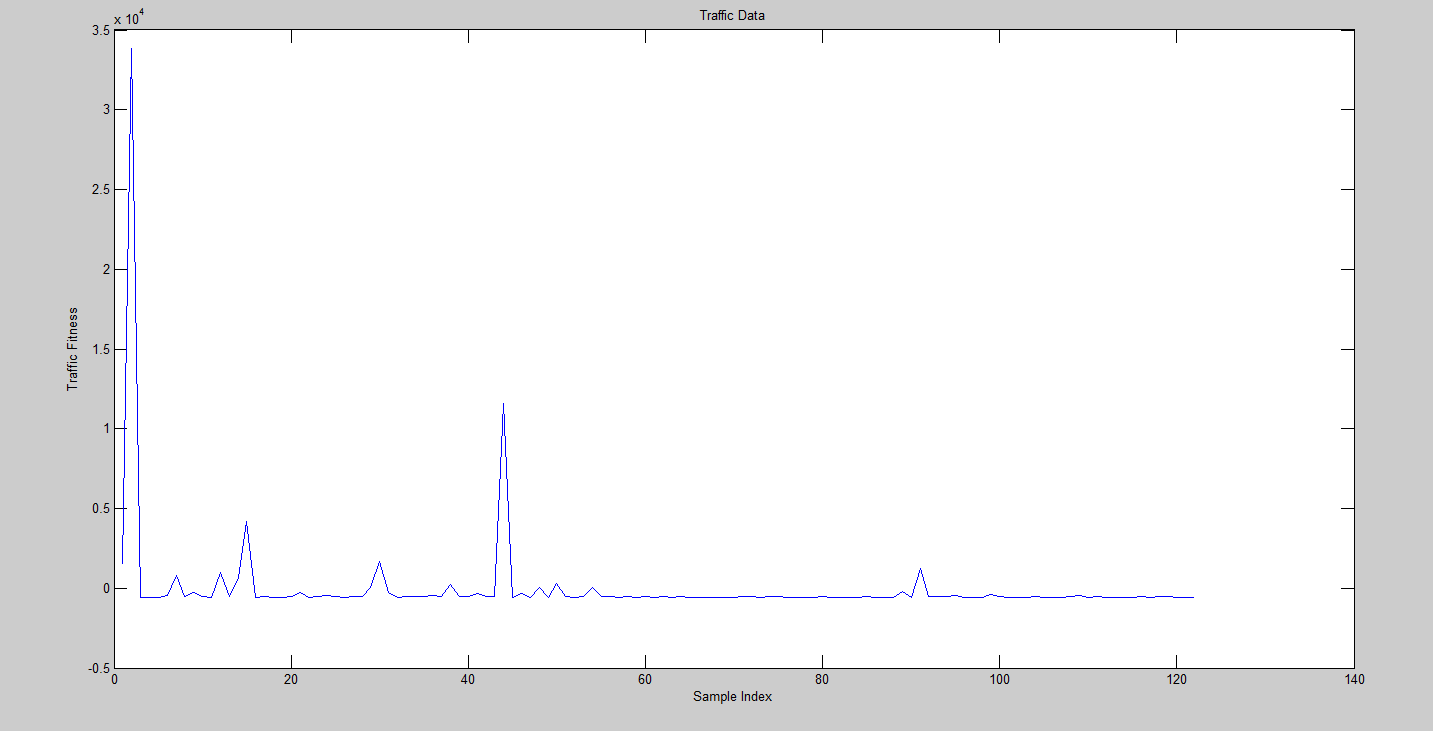
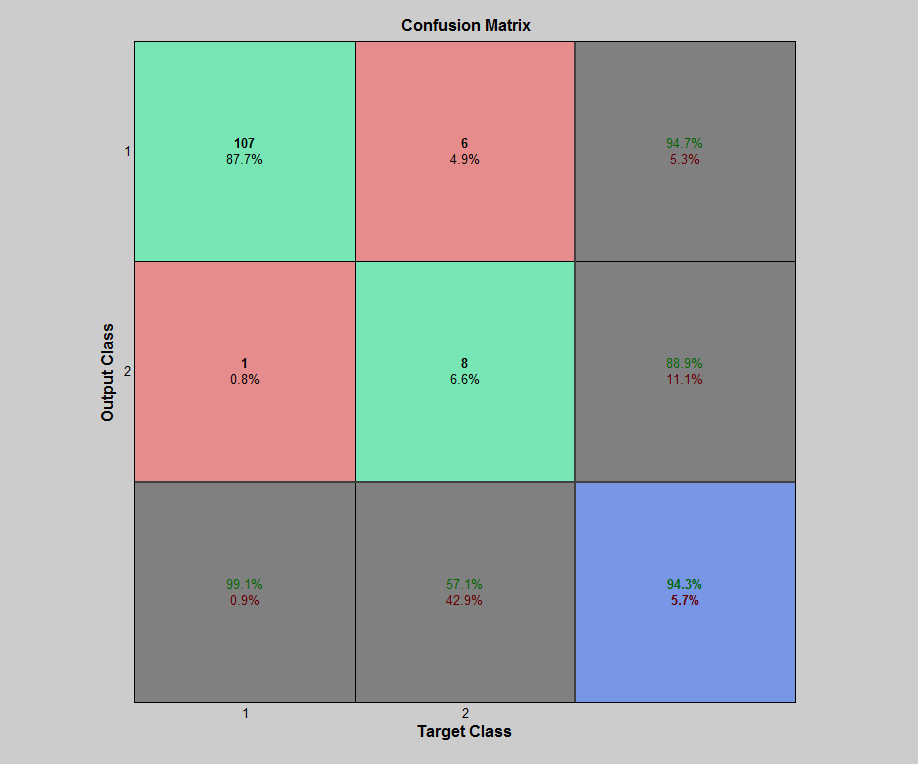
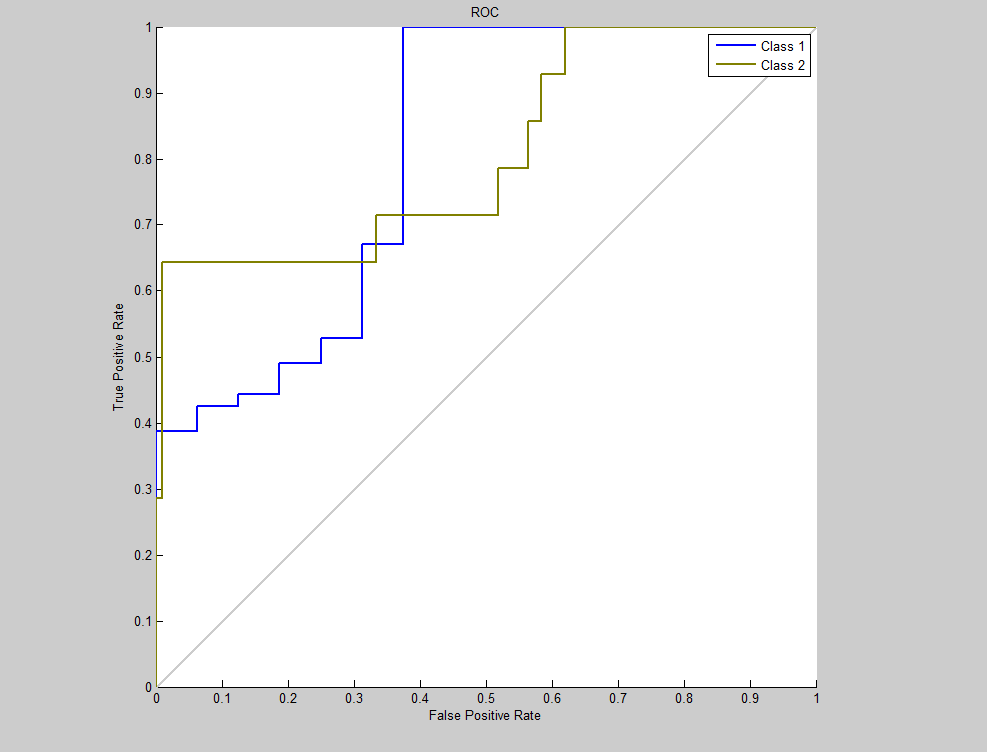
* + 1. **LDA**



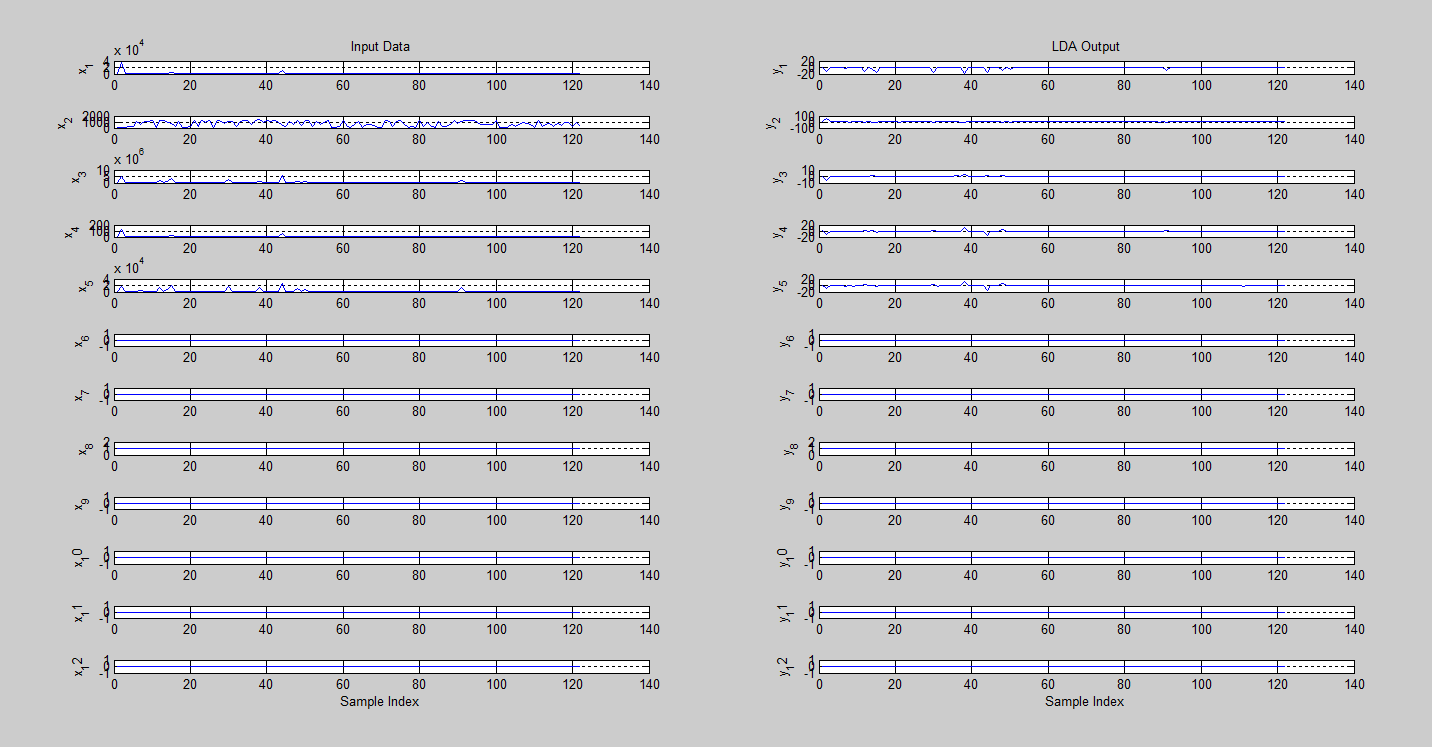
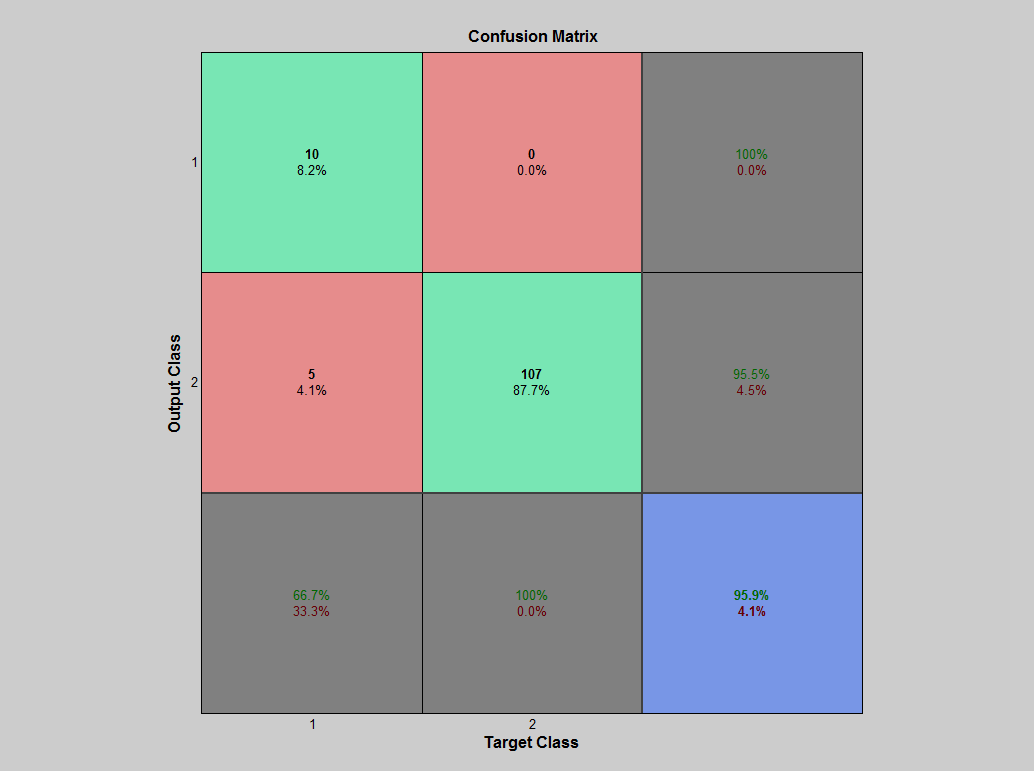
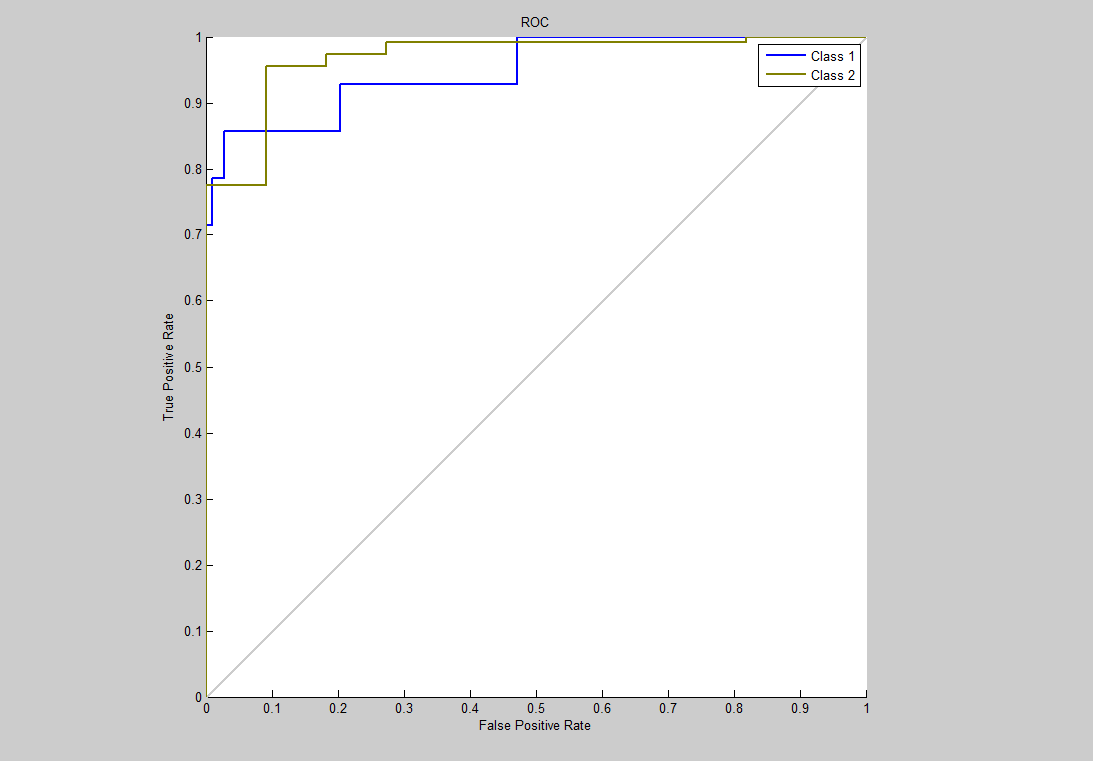
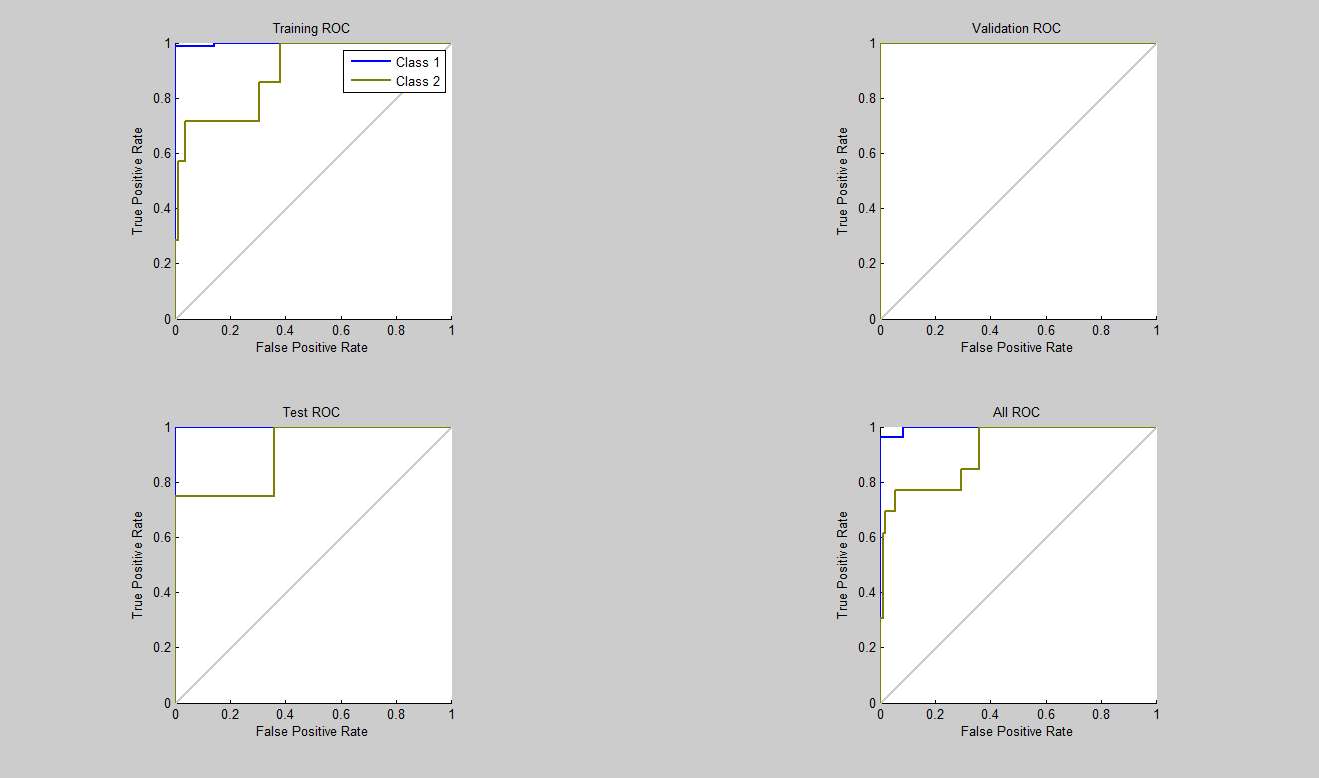
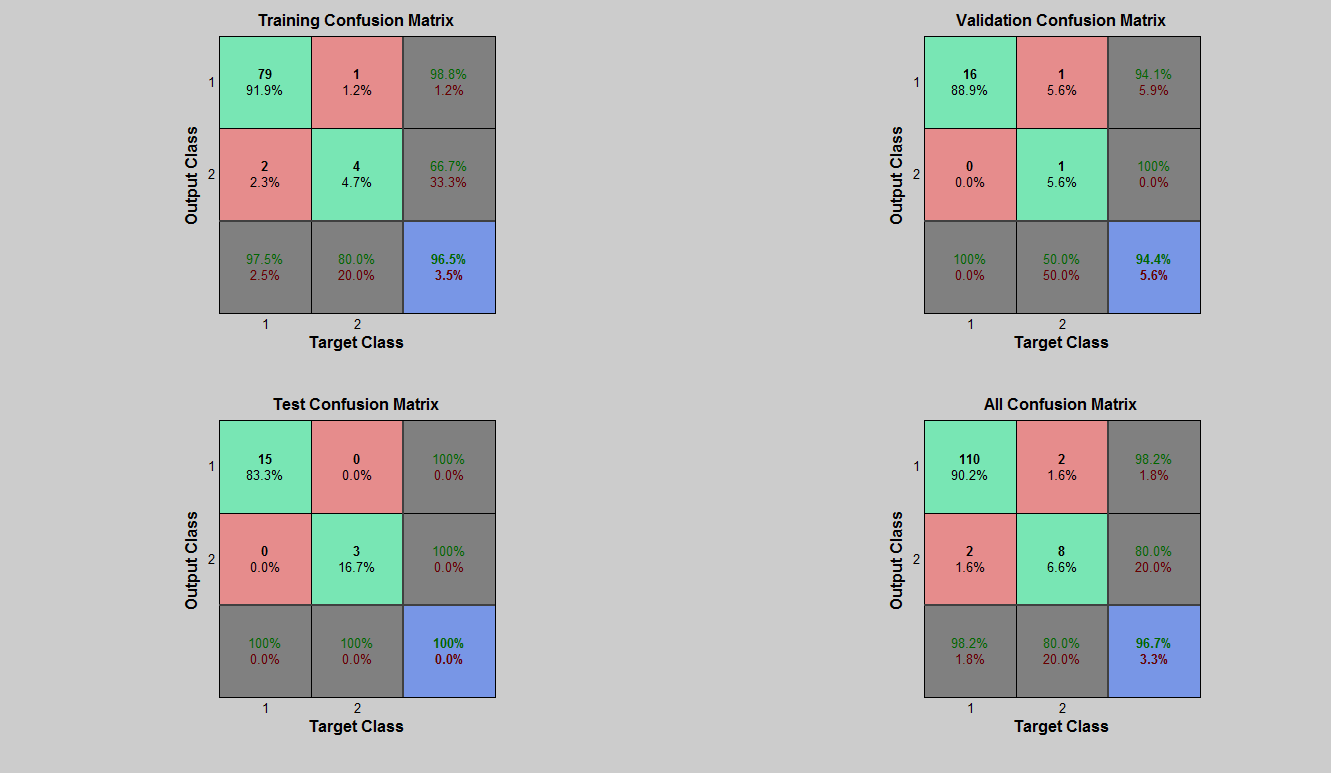
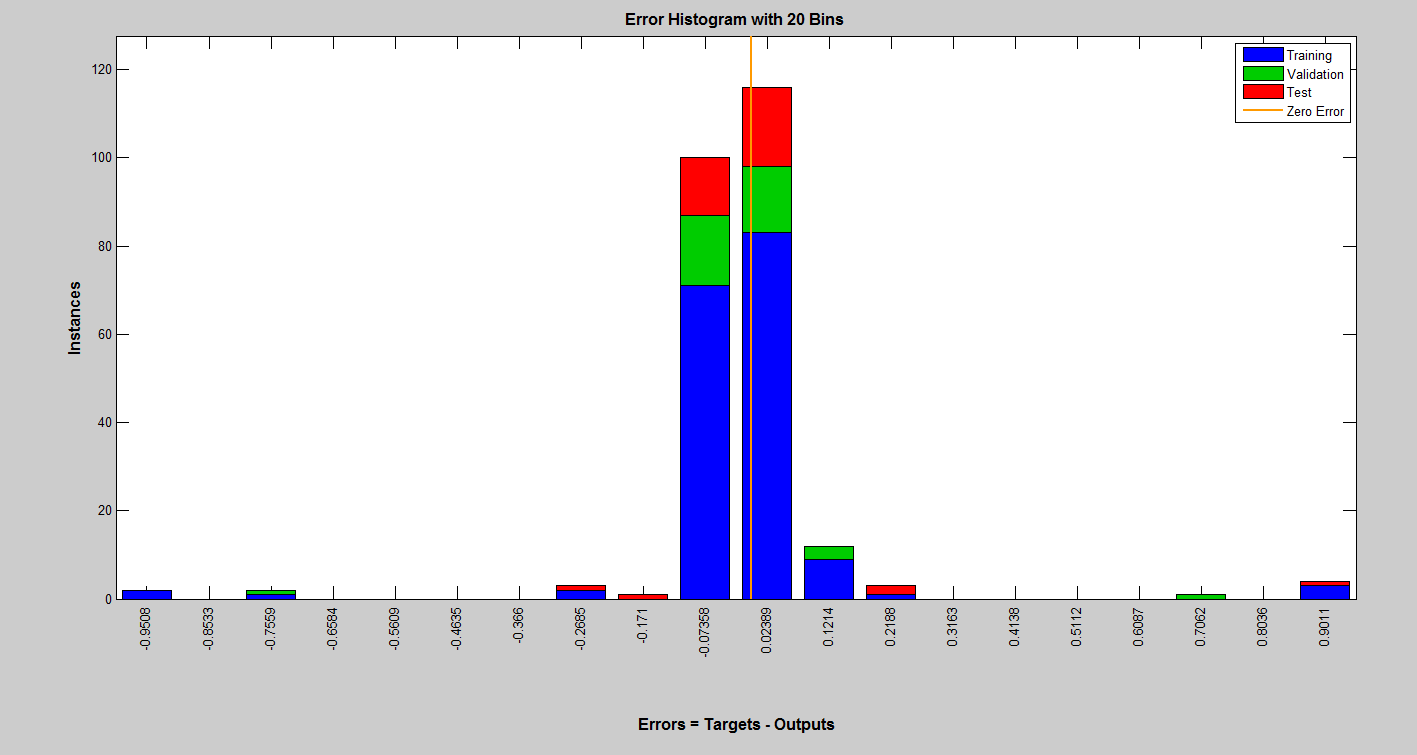
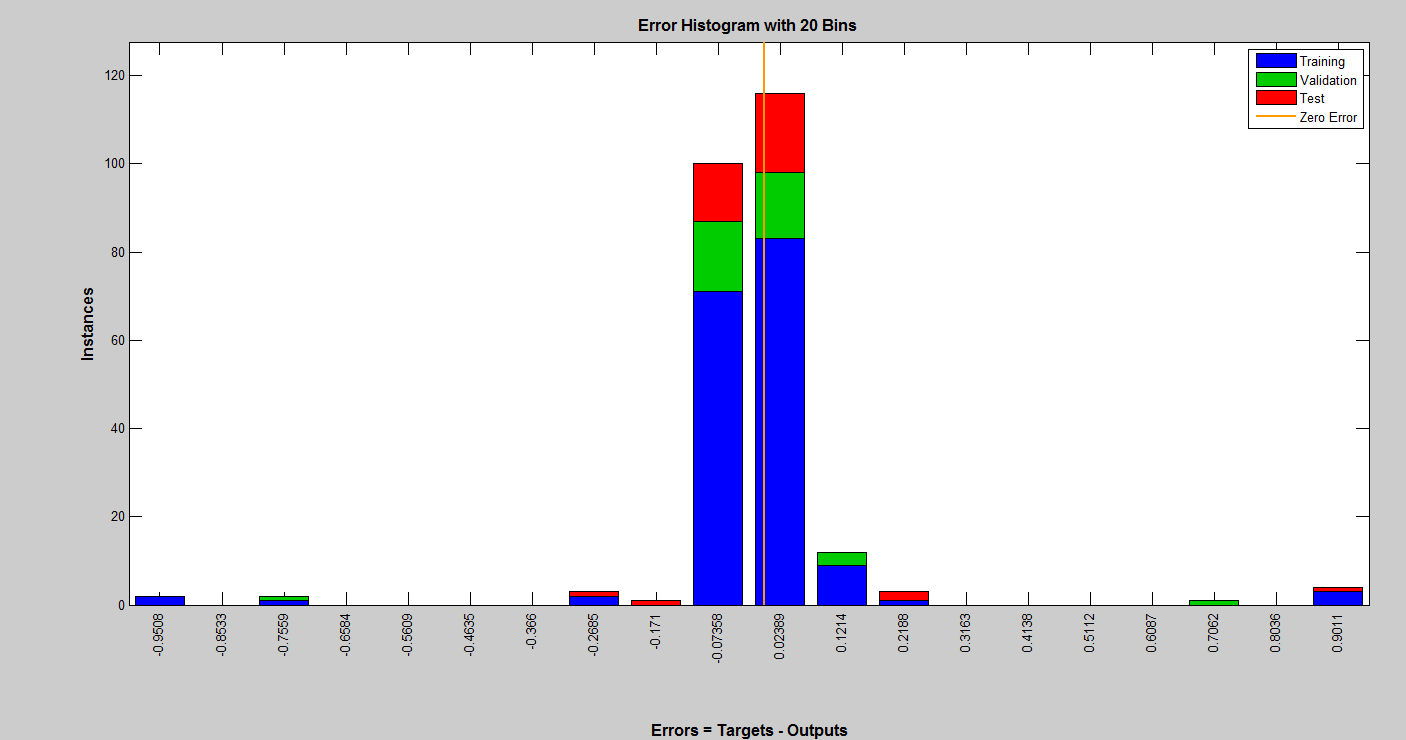
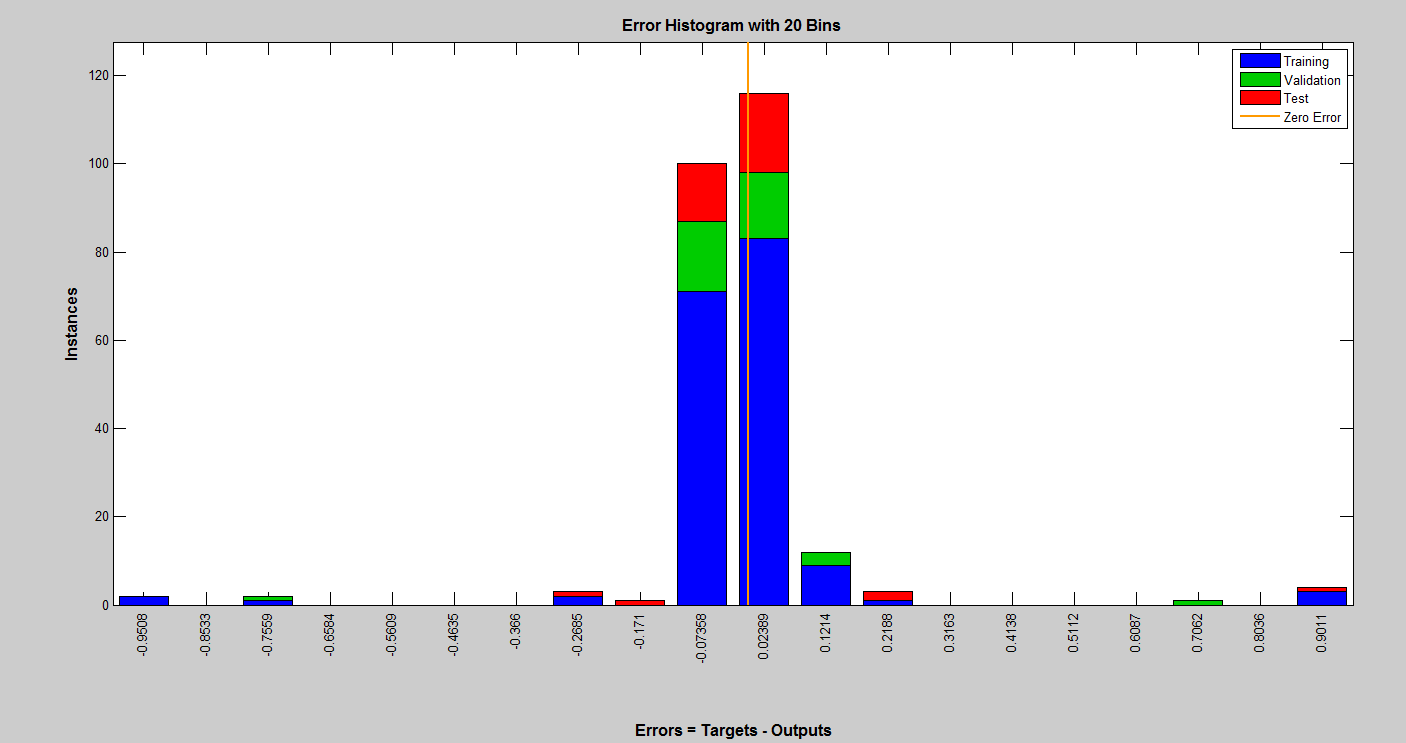
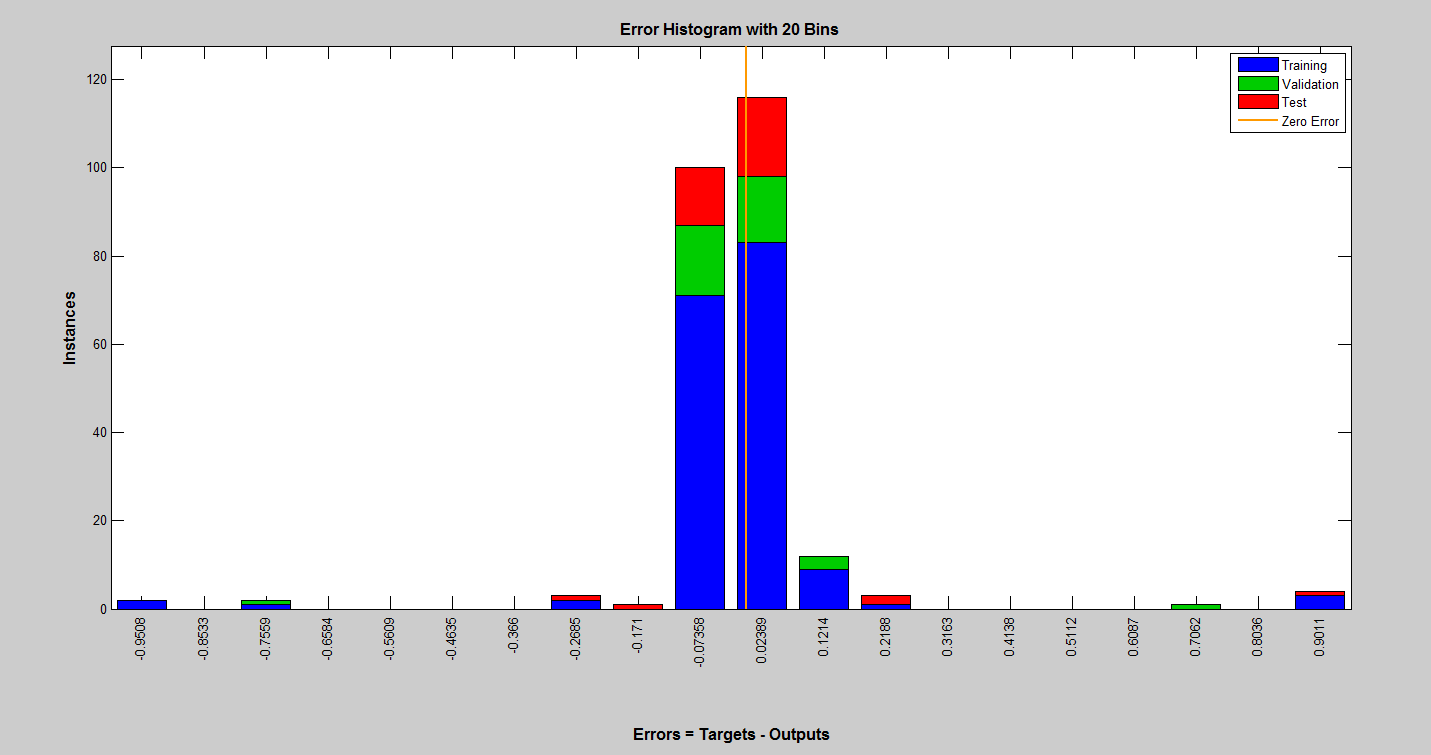
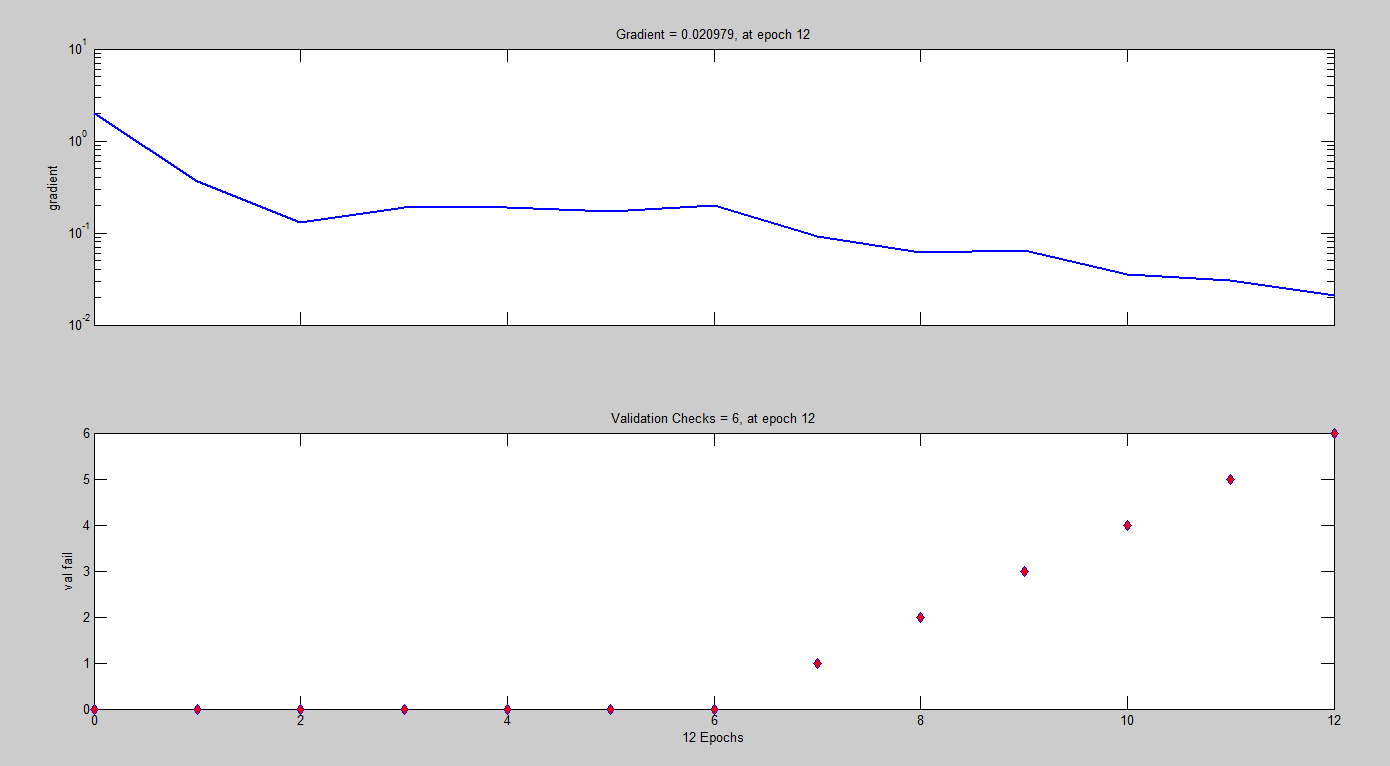
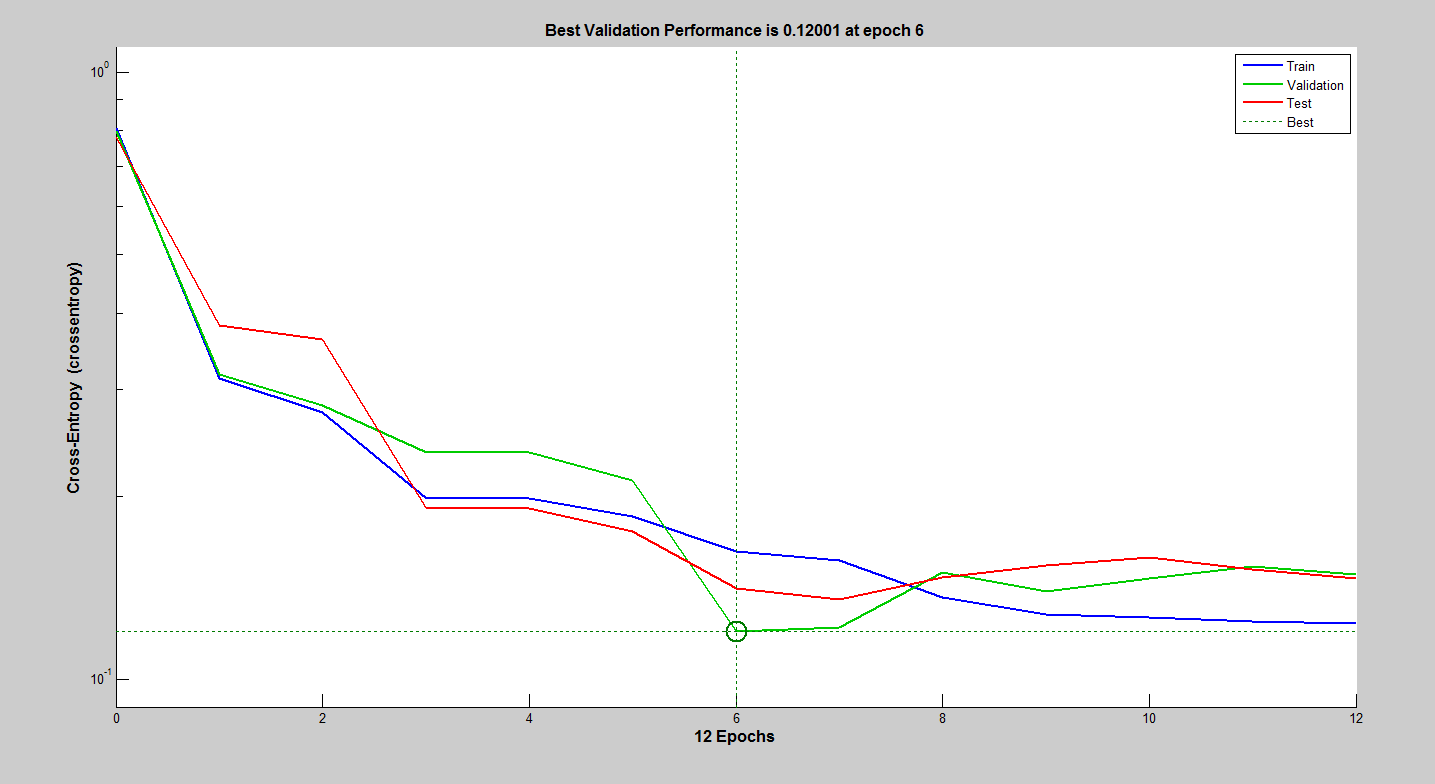
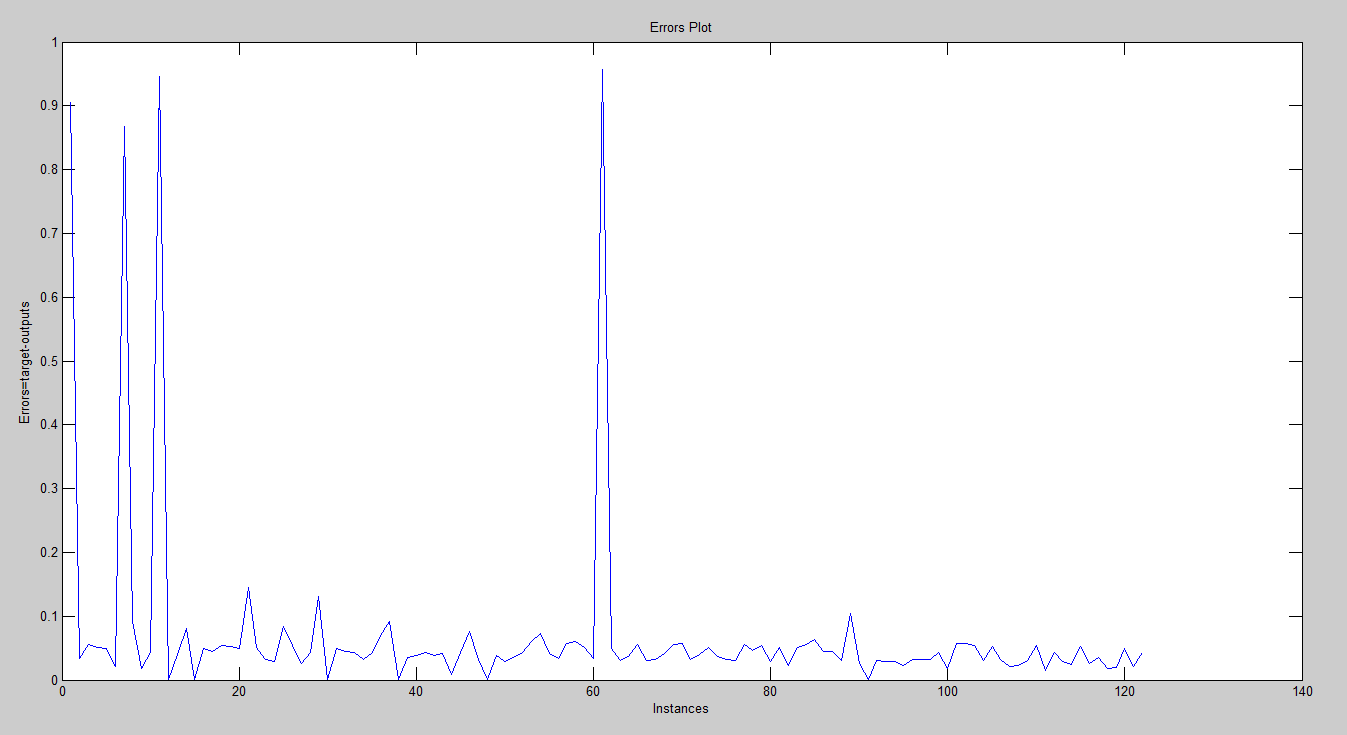
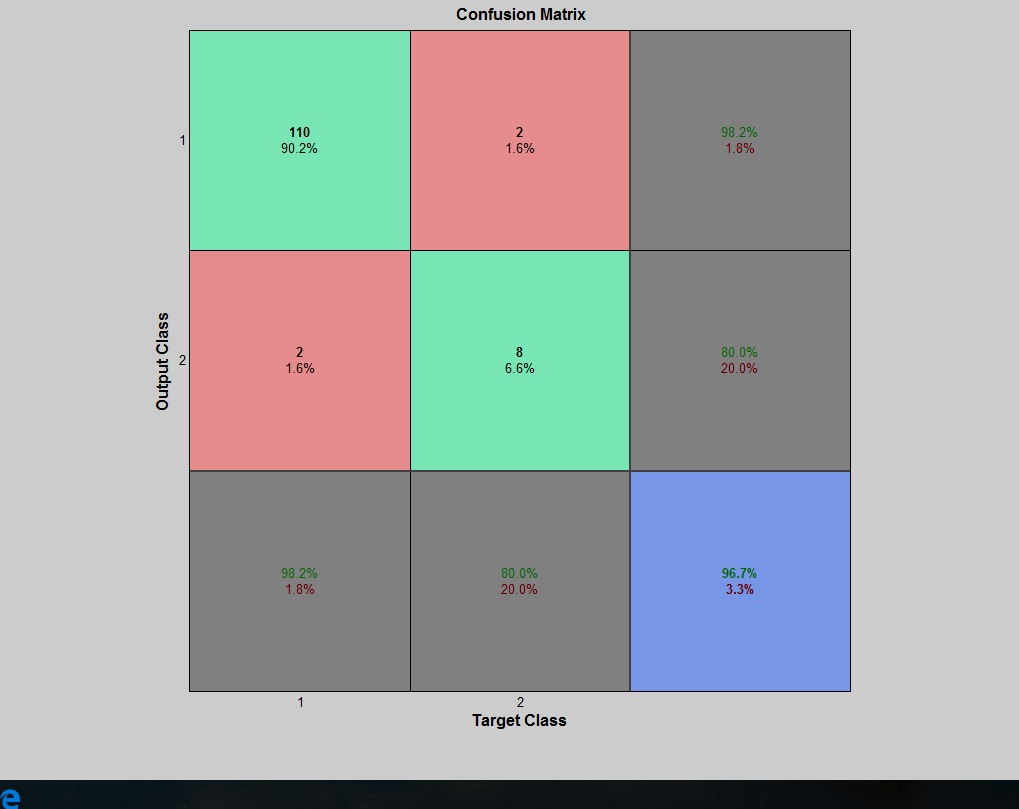
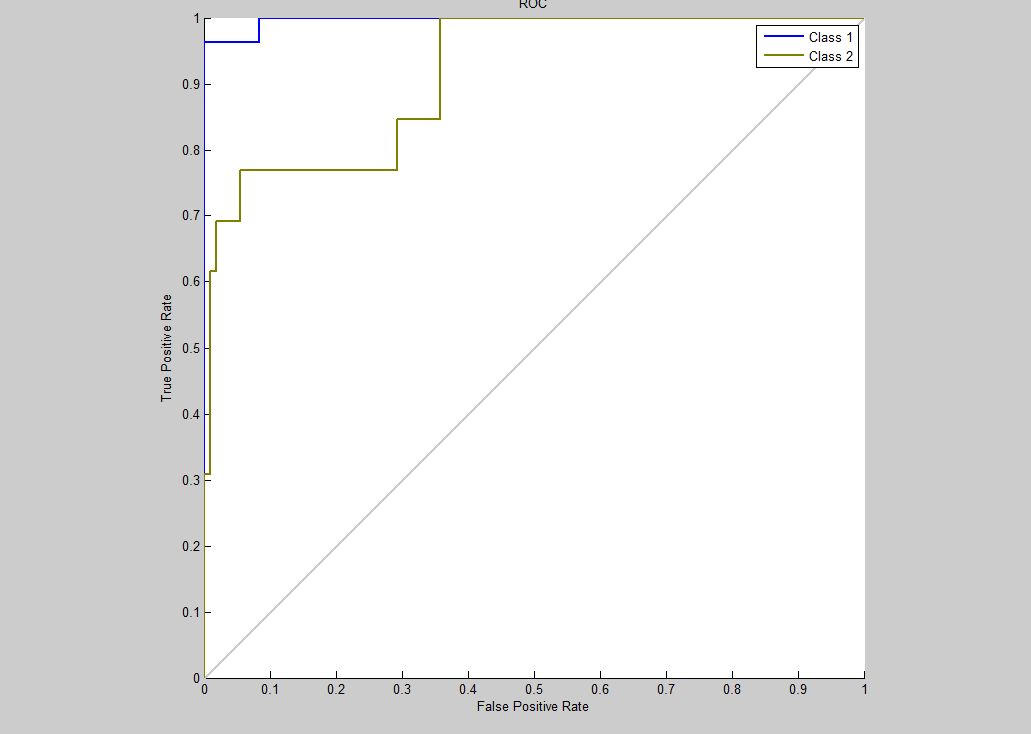
* + 1. **Proposed (PCA+LDA+ALO)**



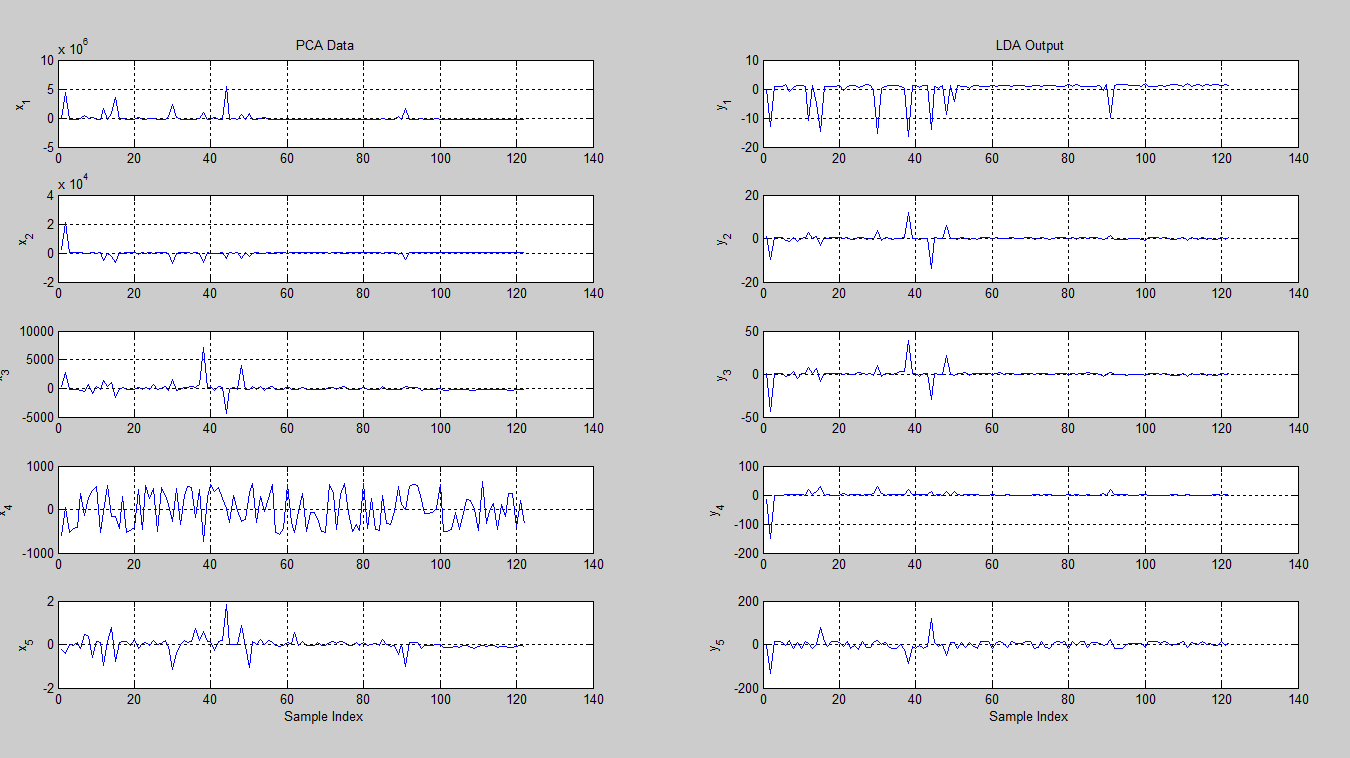
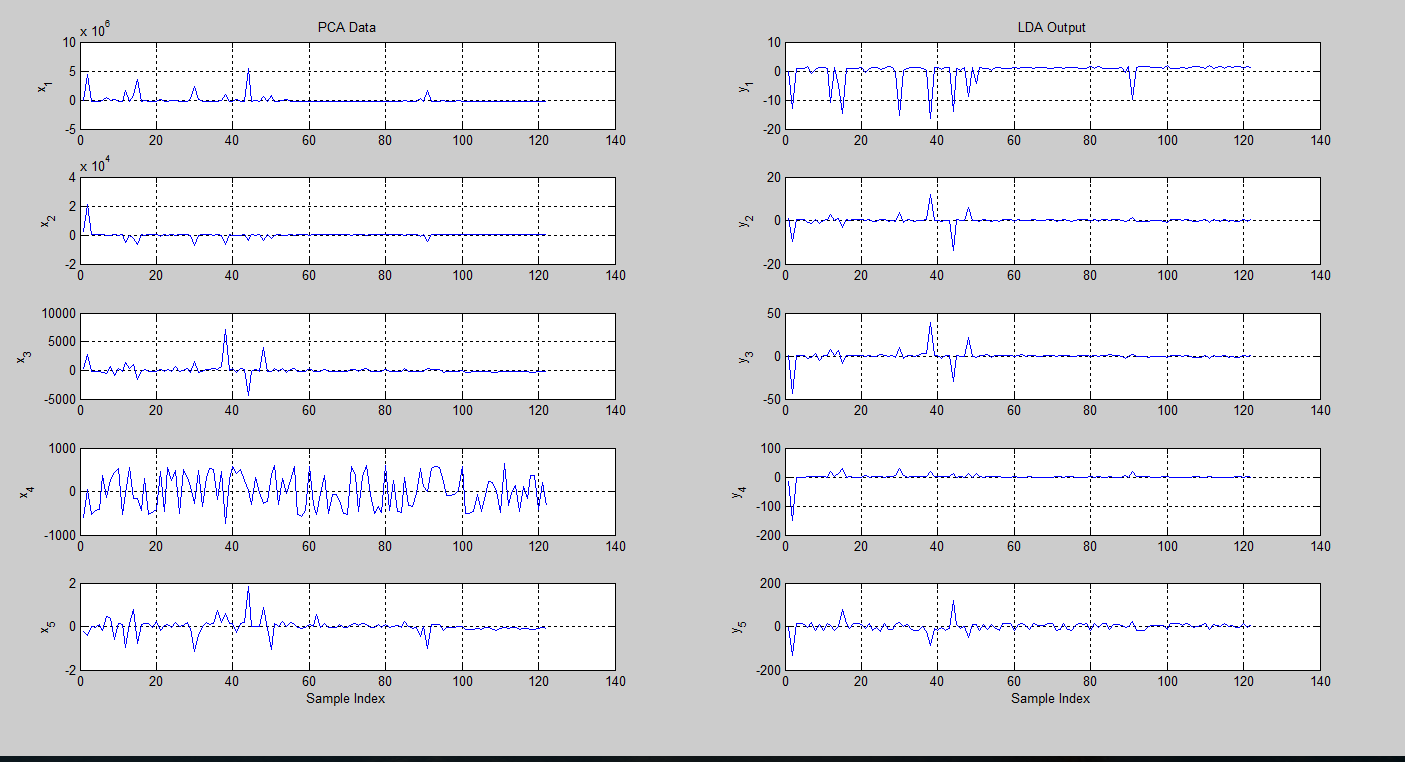
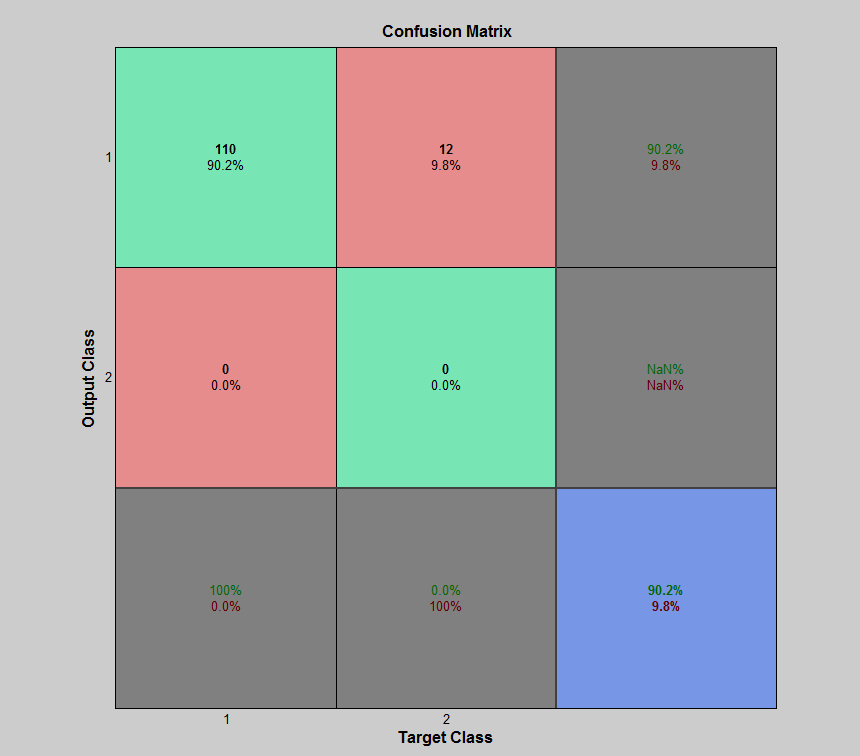
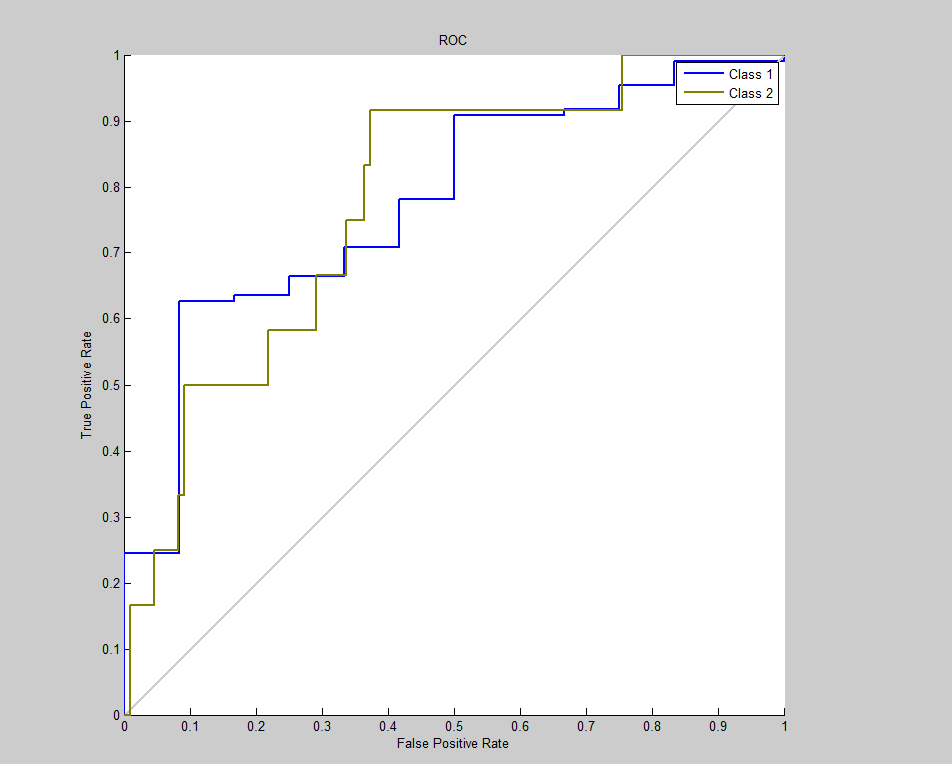
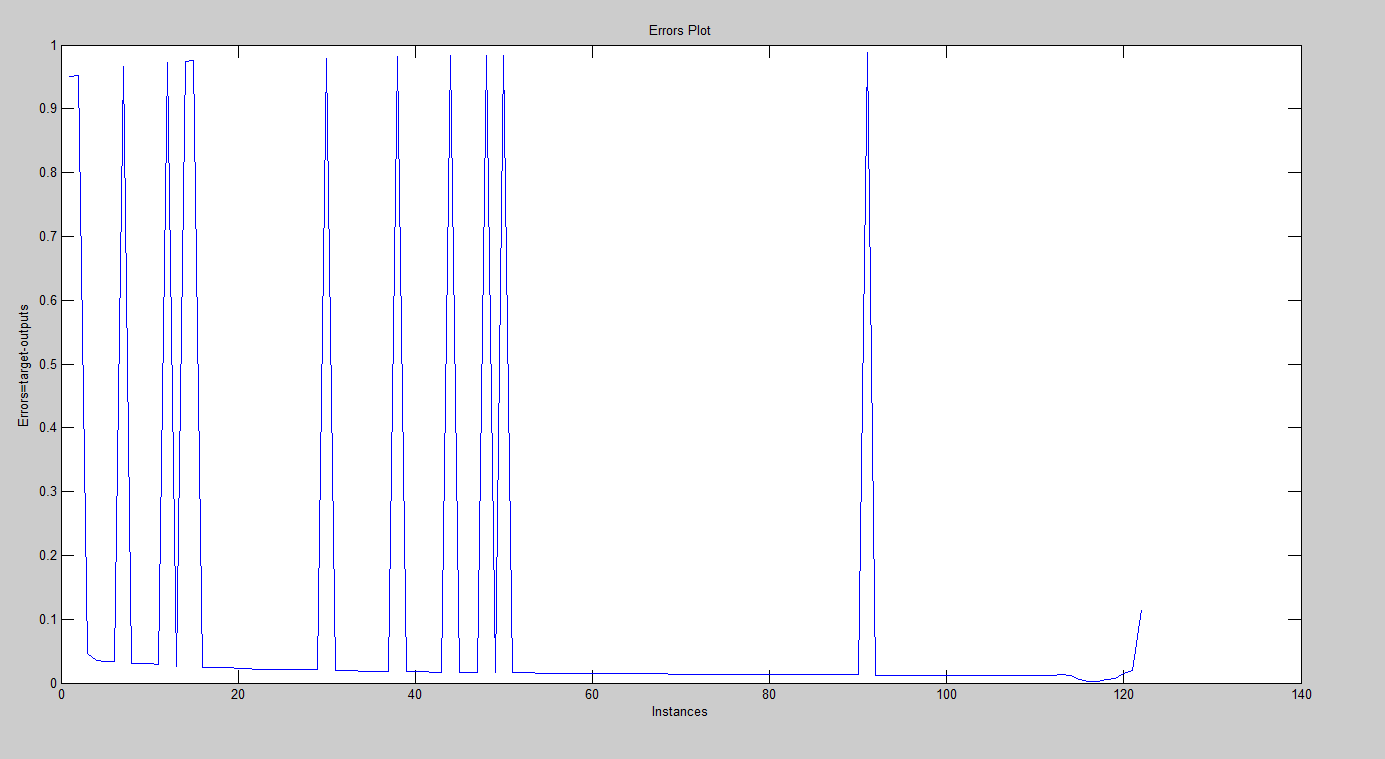
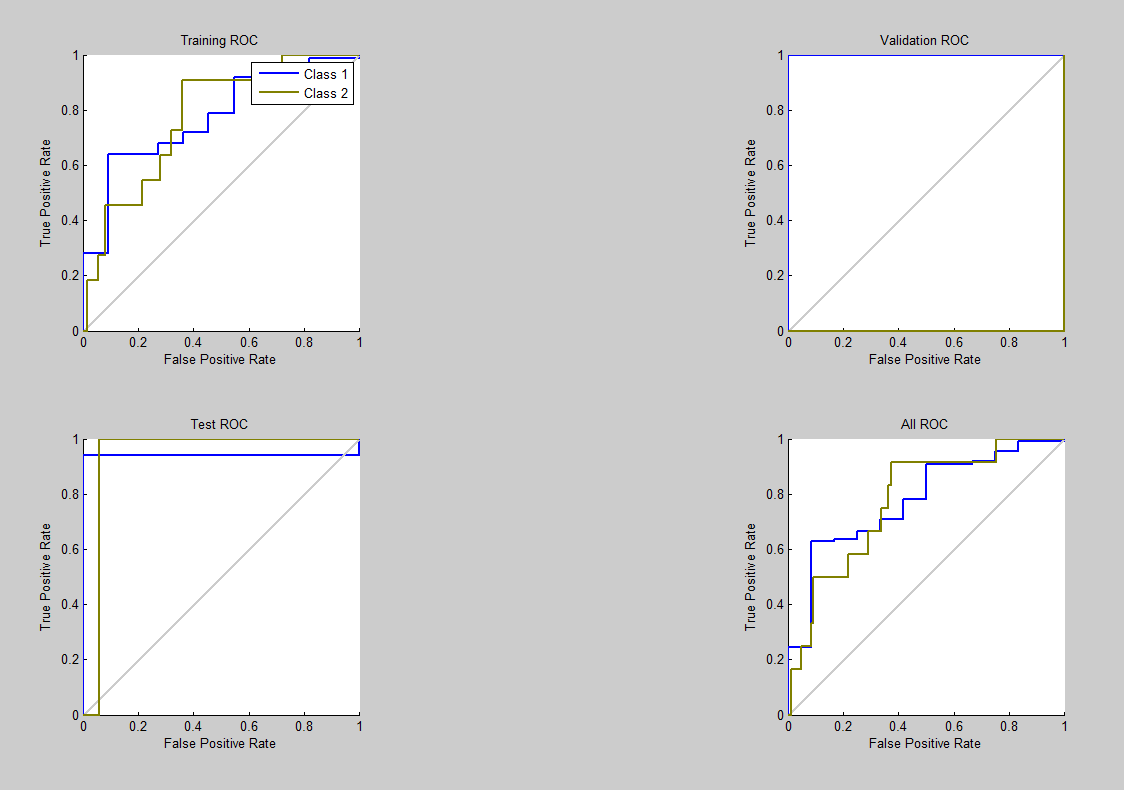
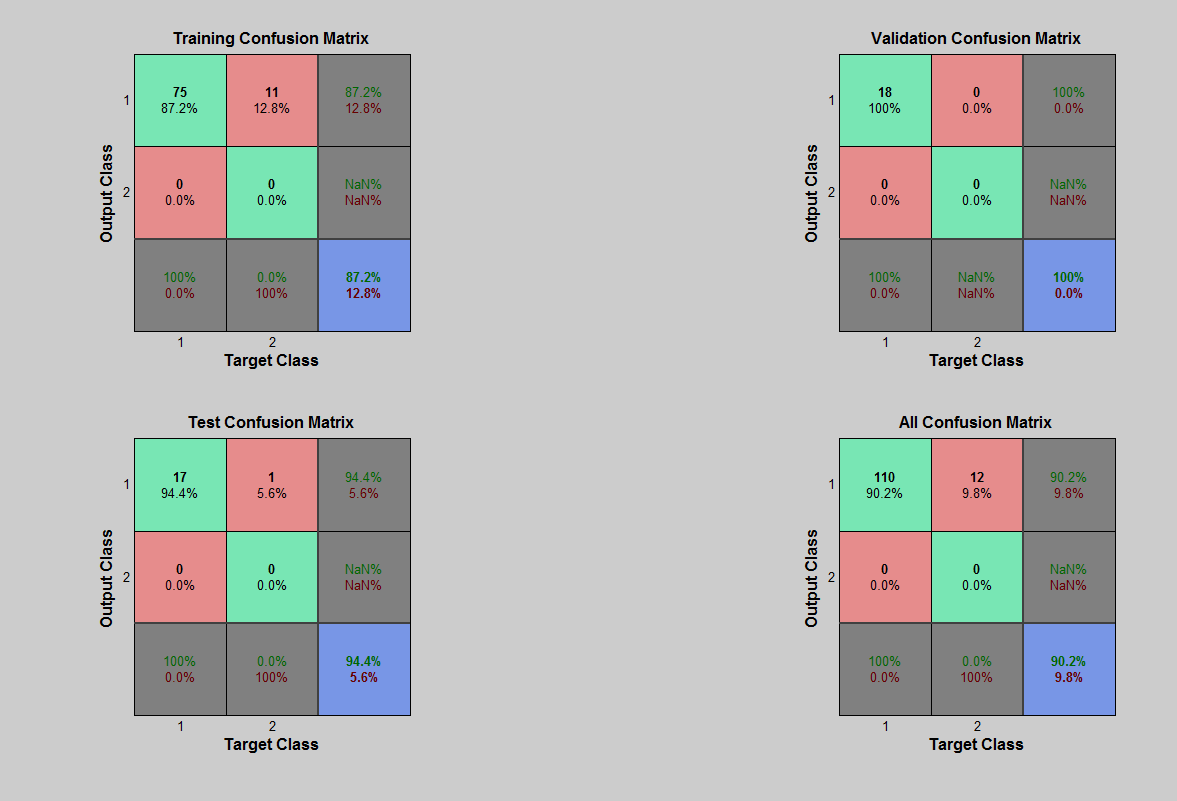
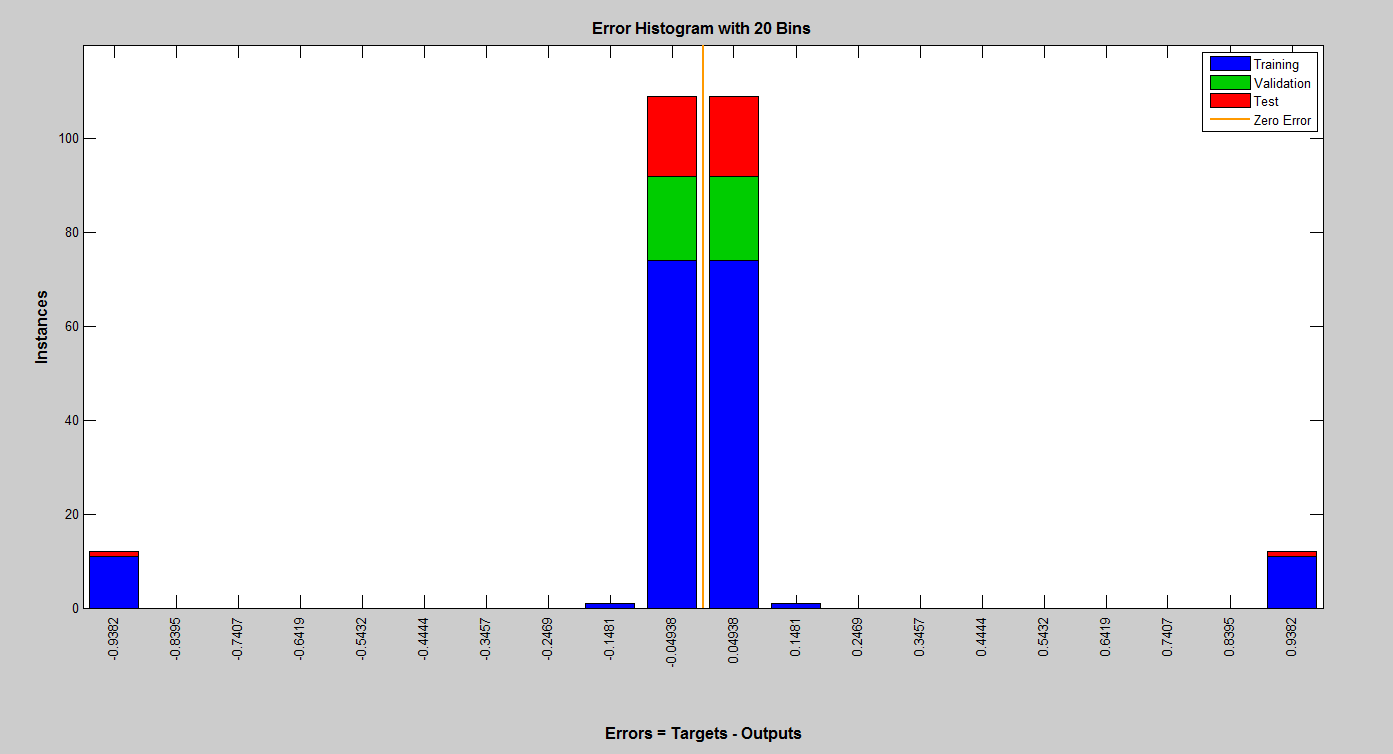
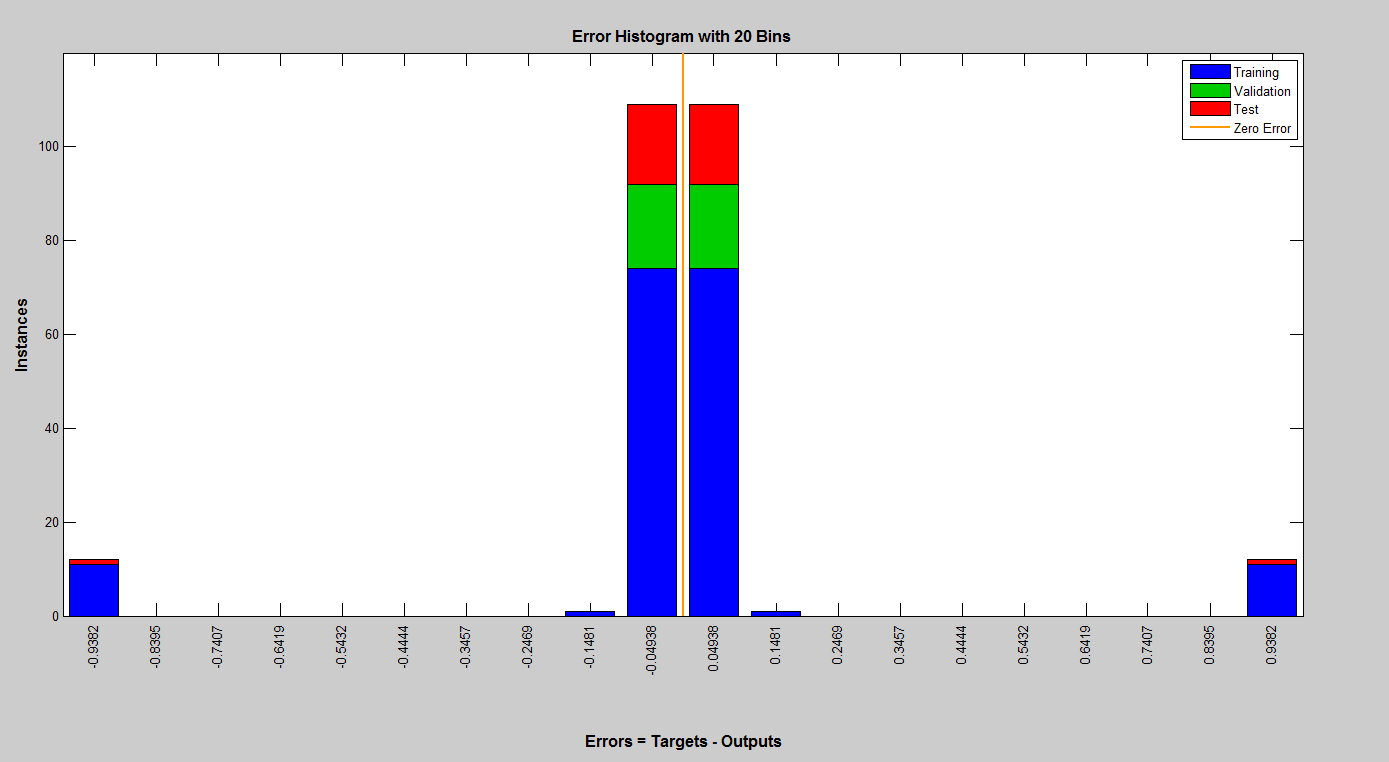
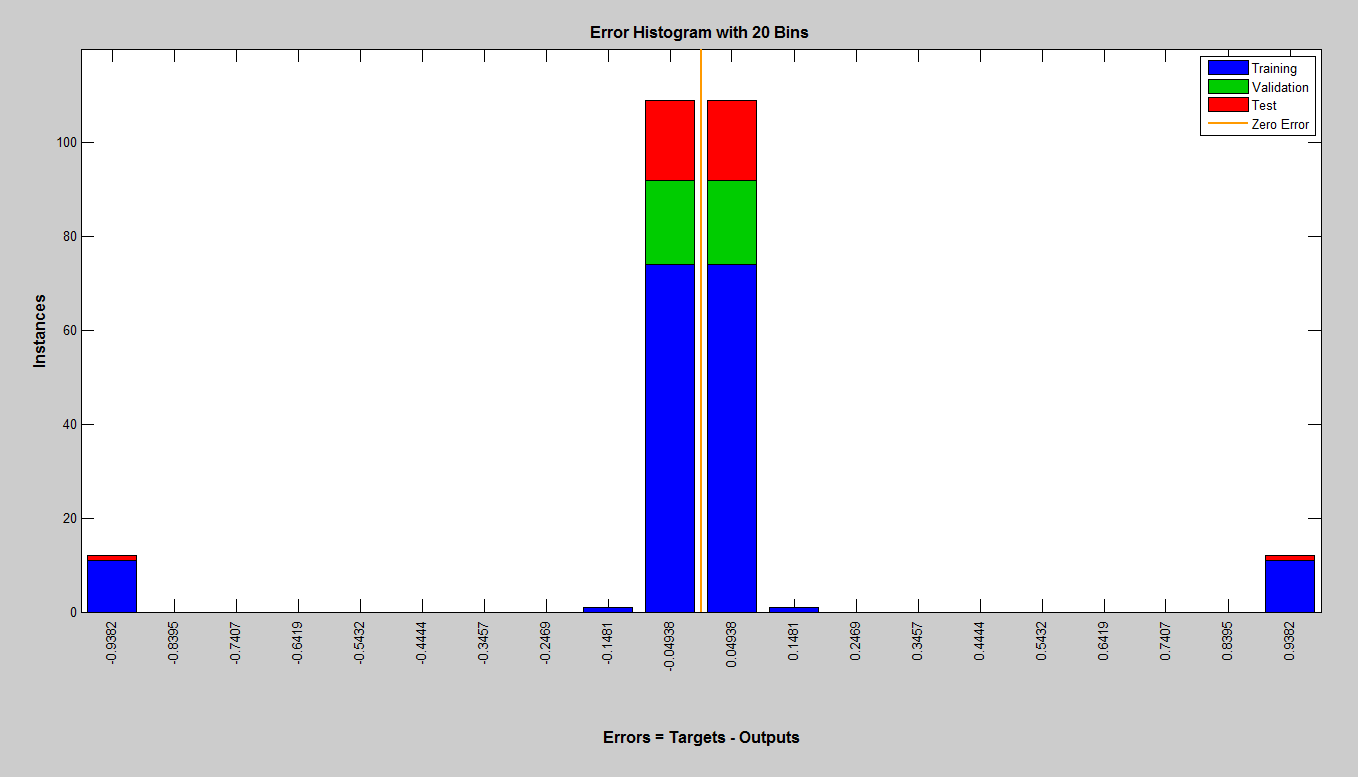
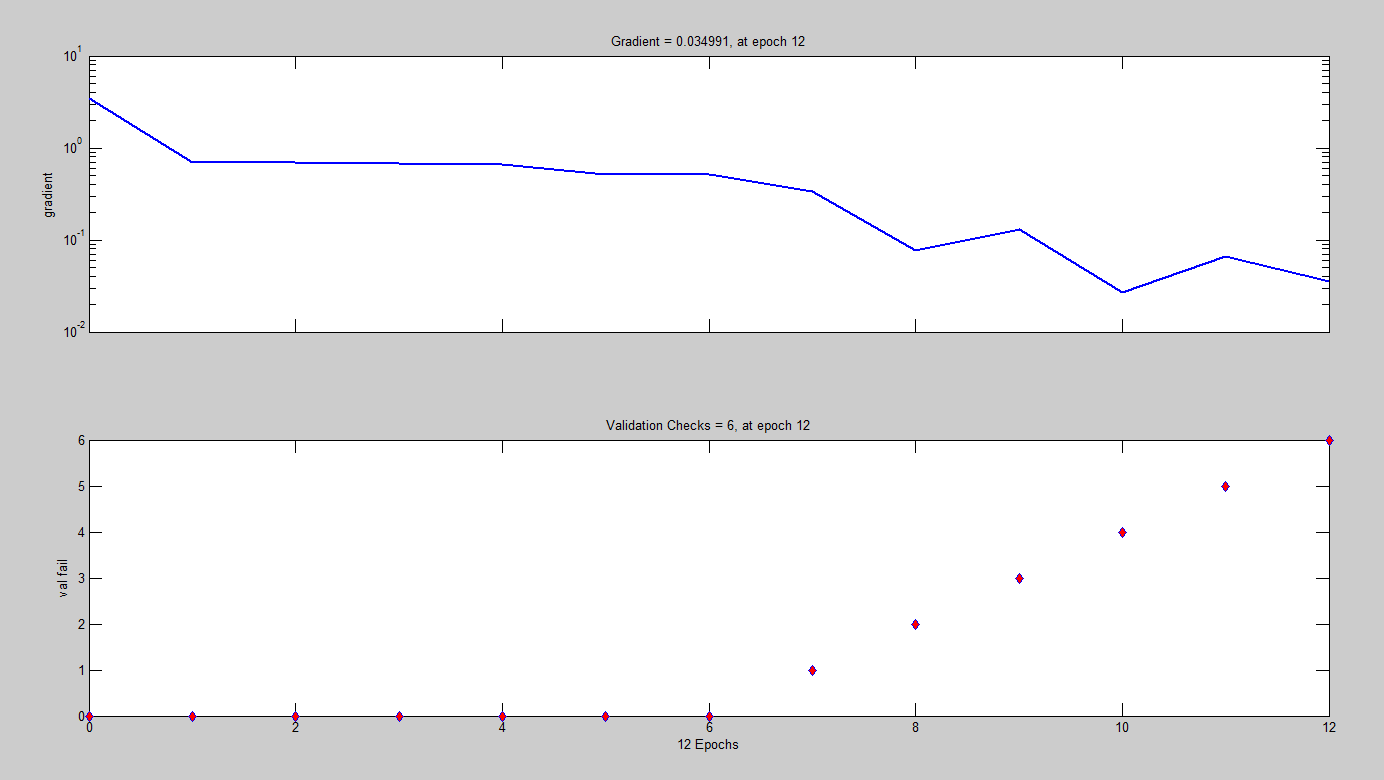
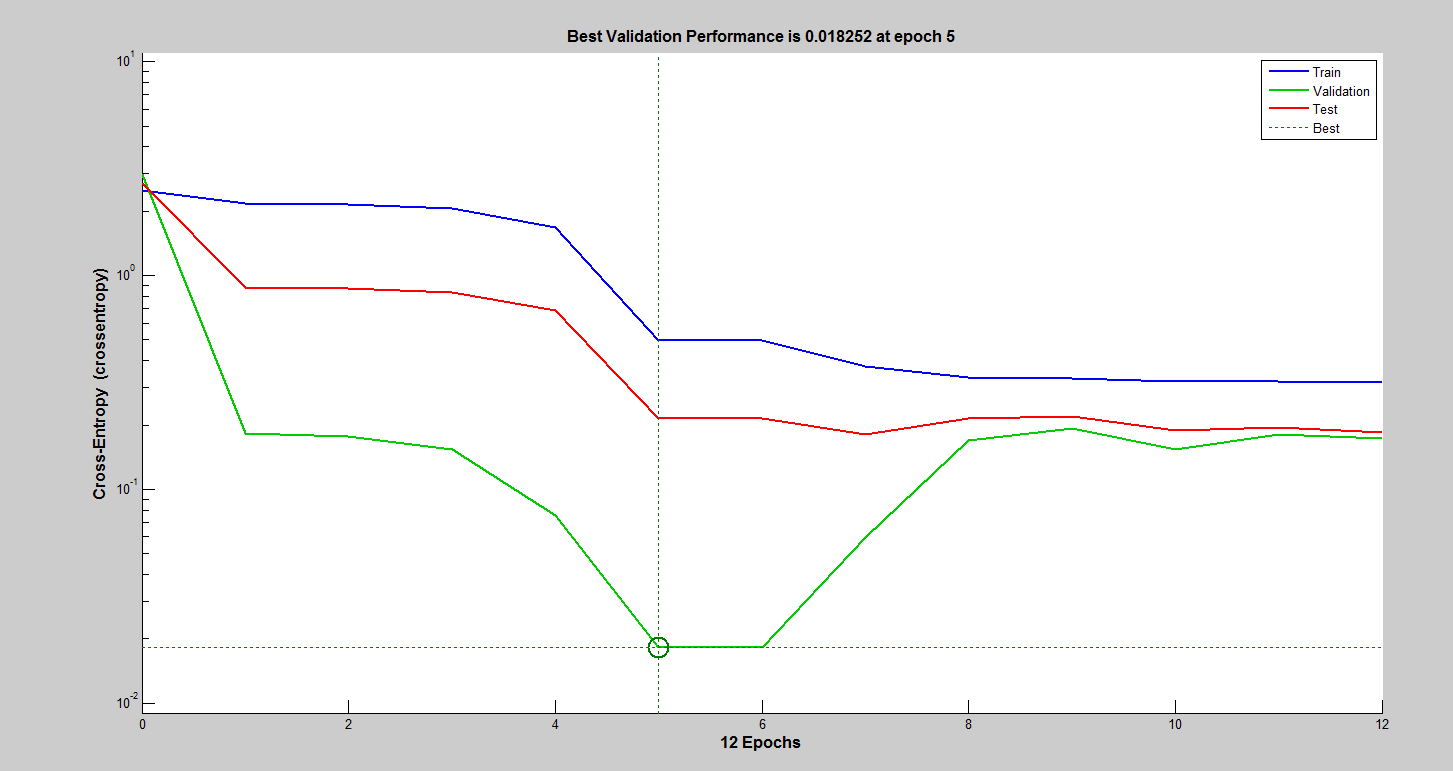
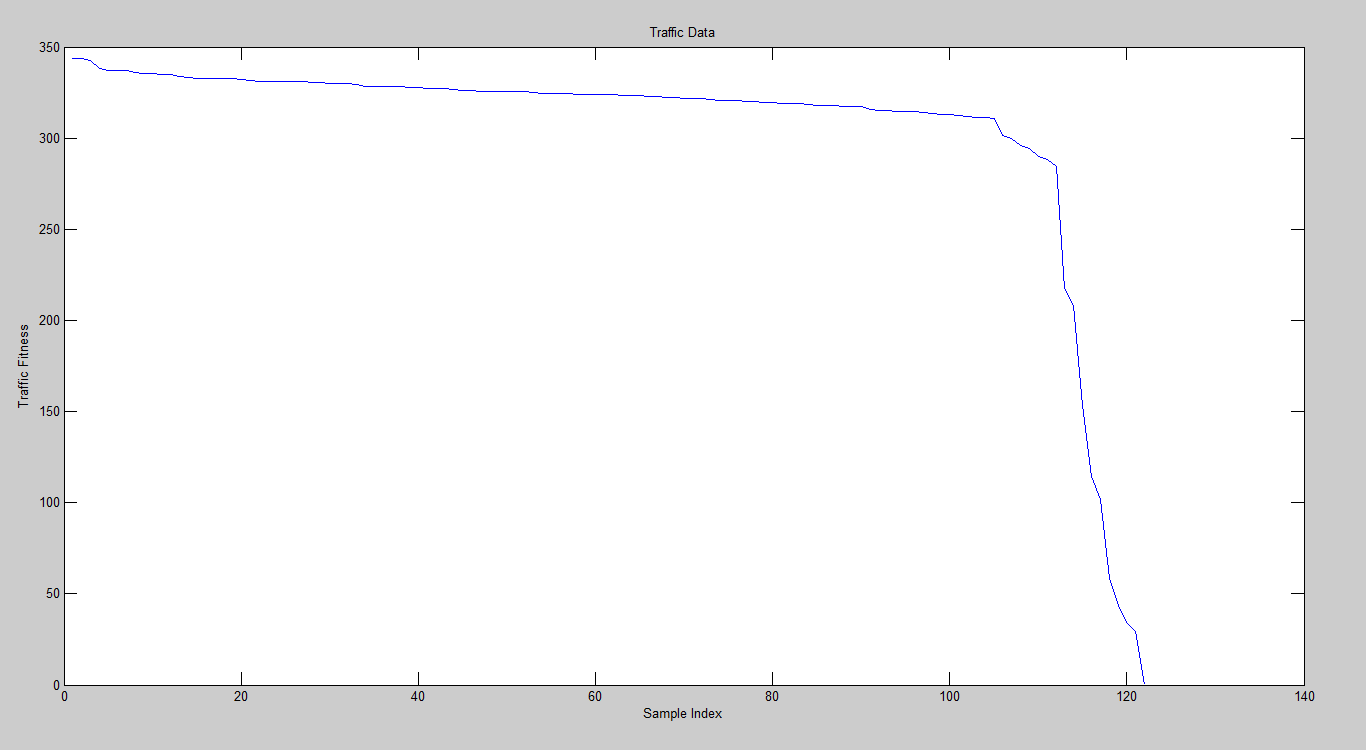
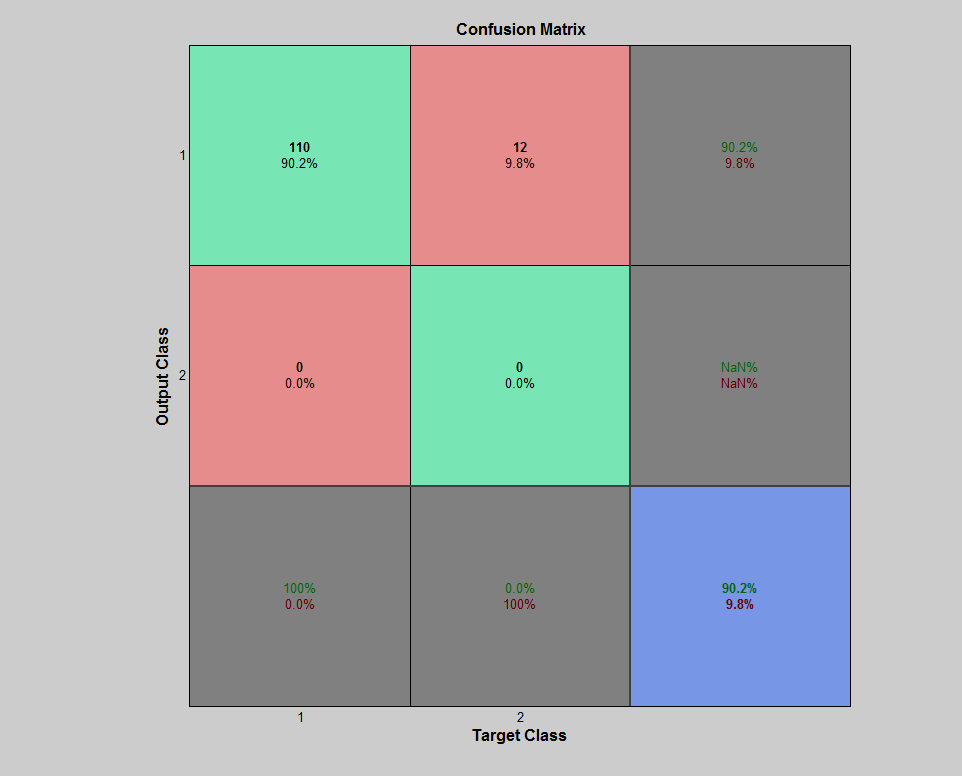
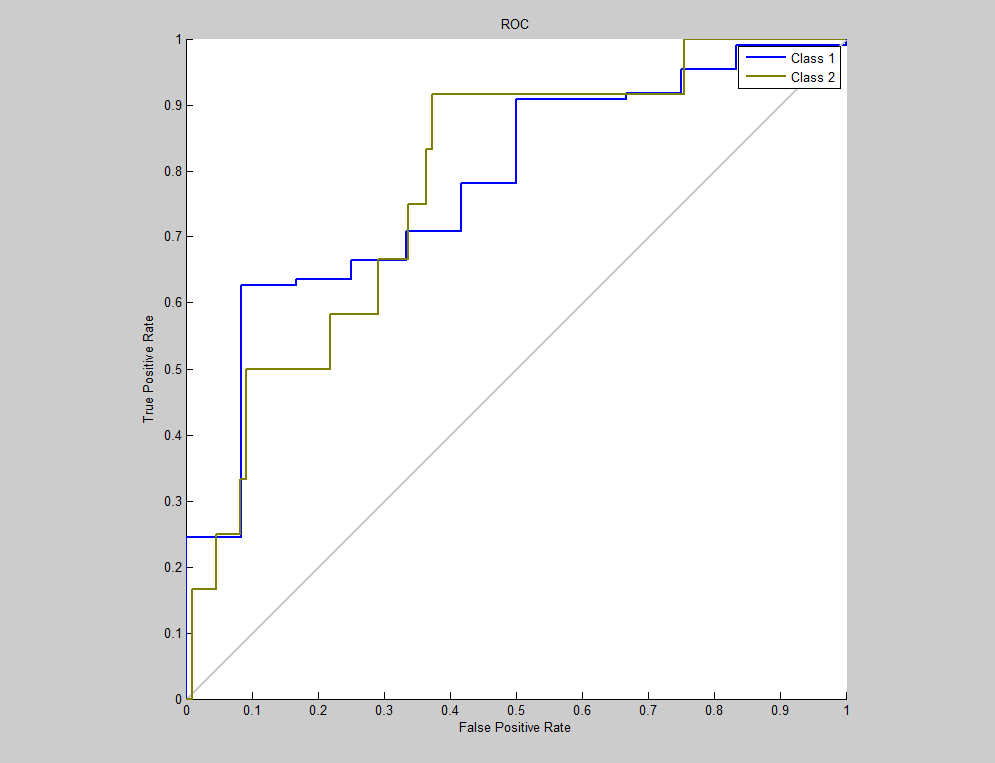
* 1. **TCP Sync Attack cloud computing**
     1. **PCA**



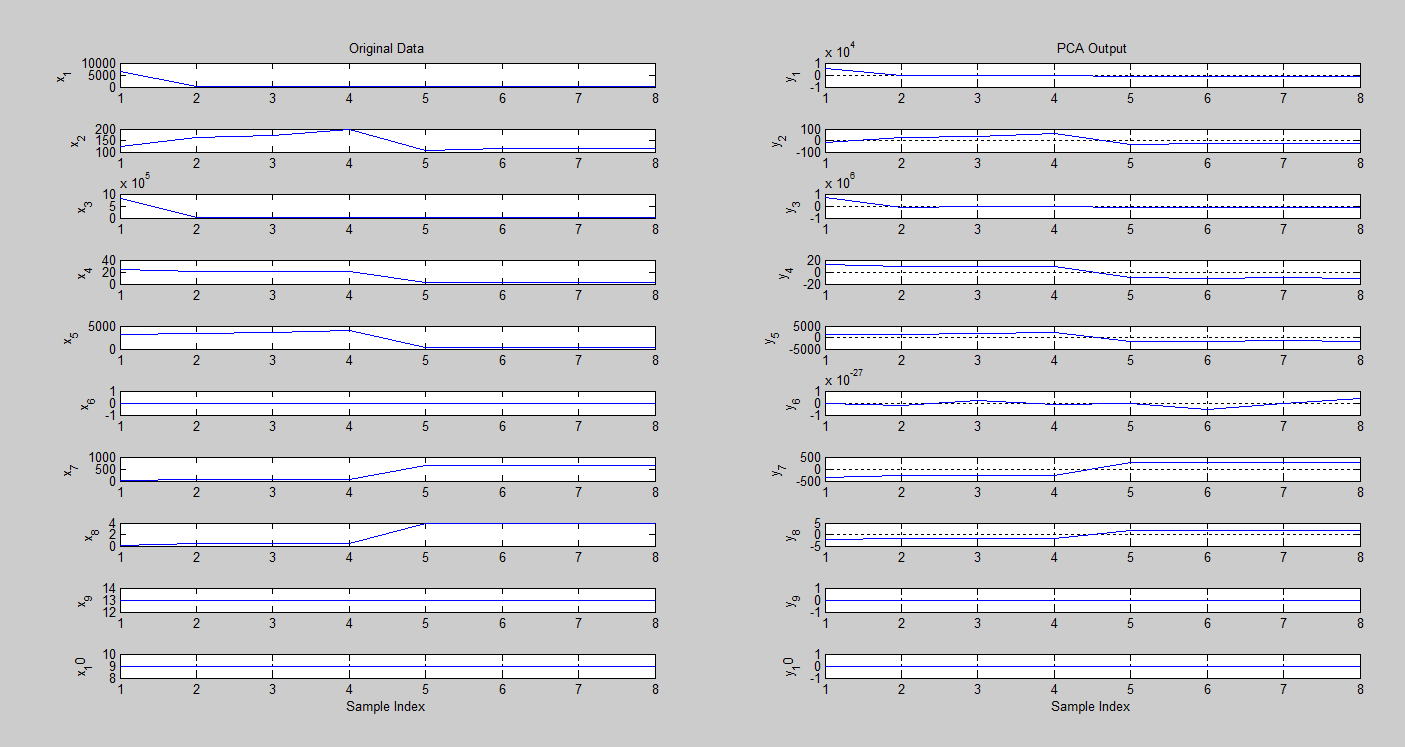
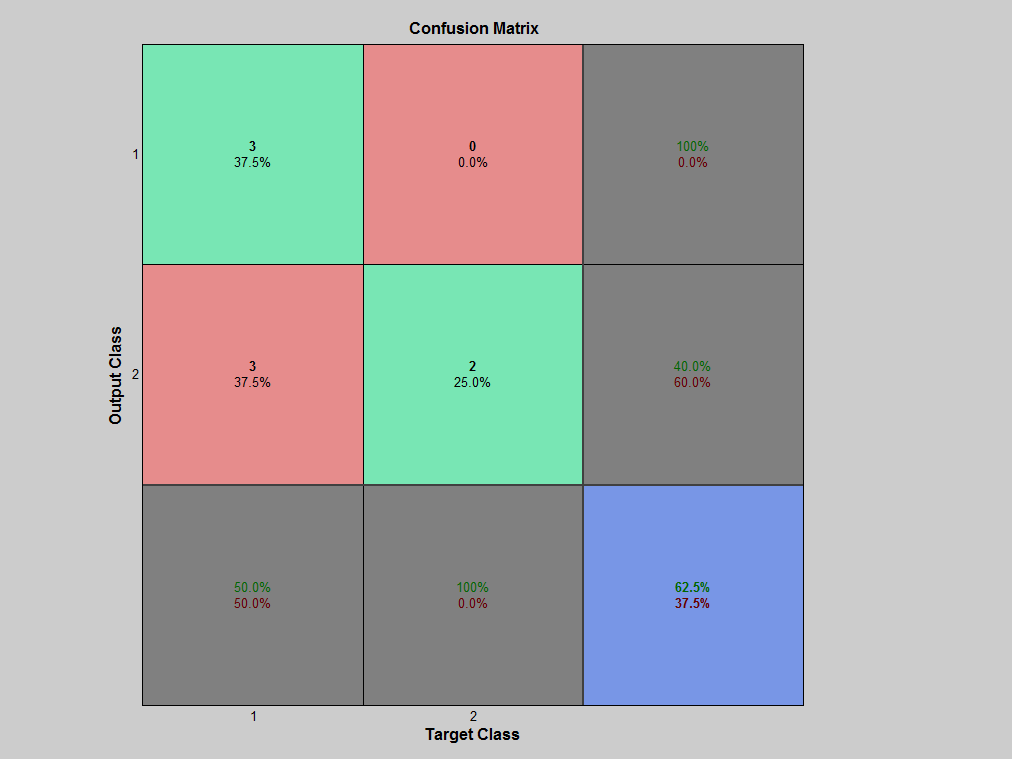
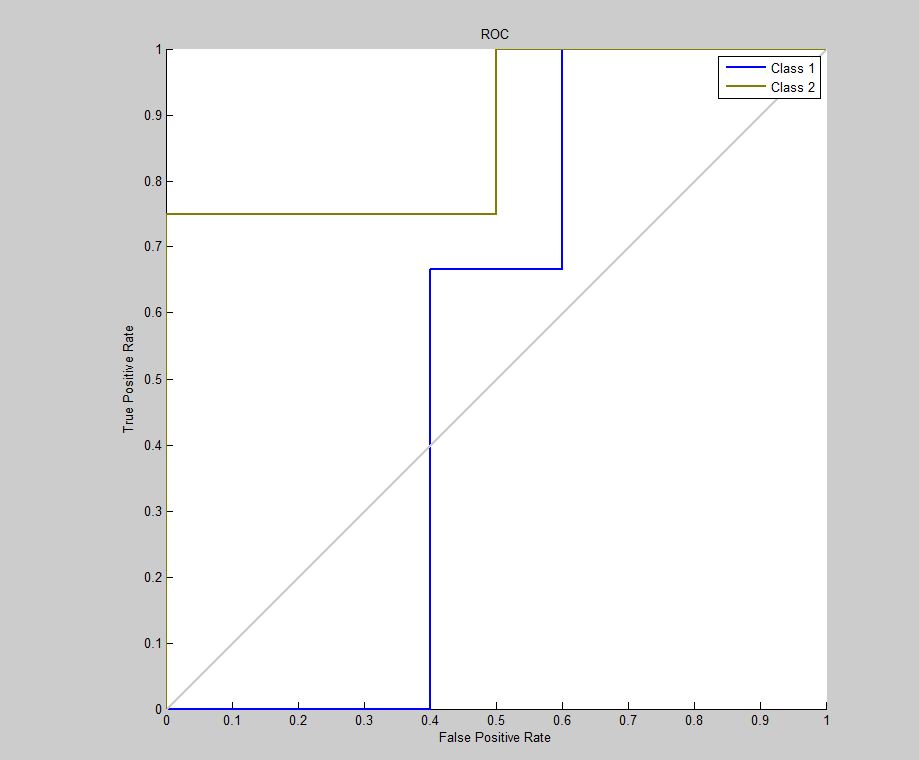
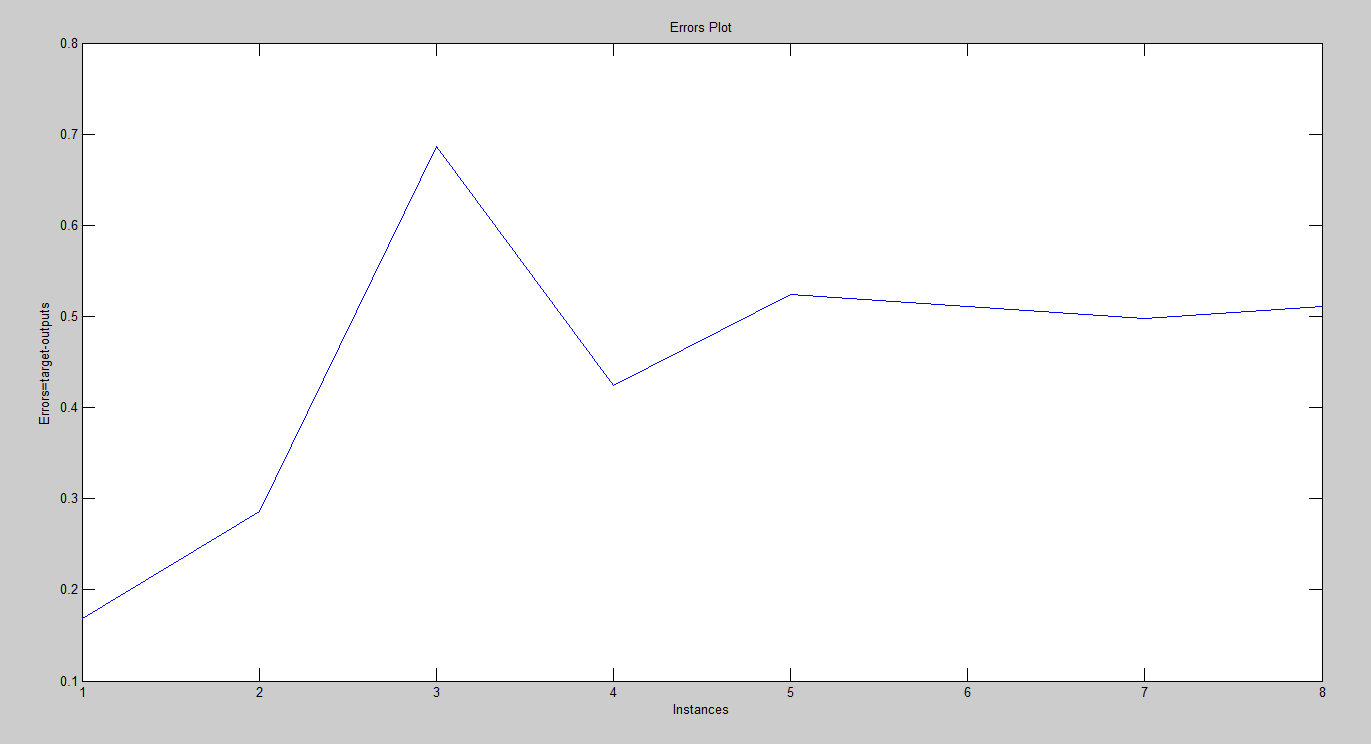
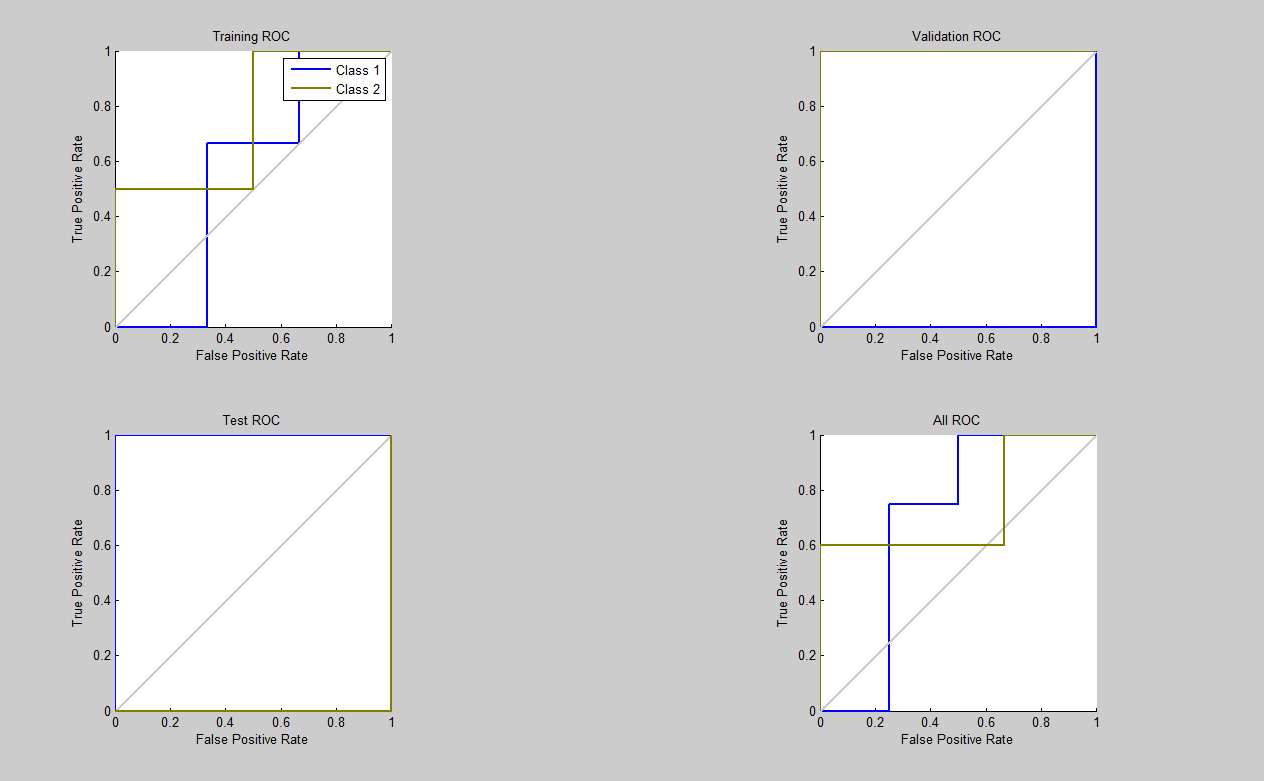
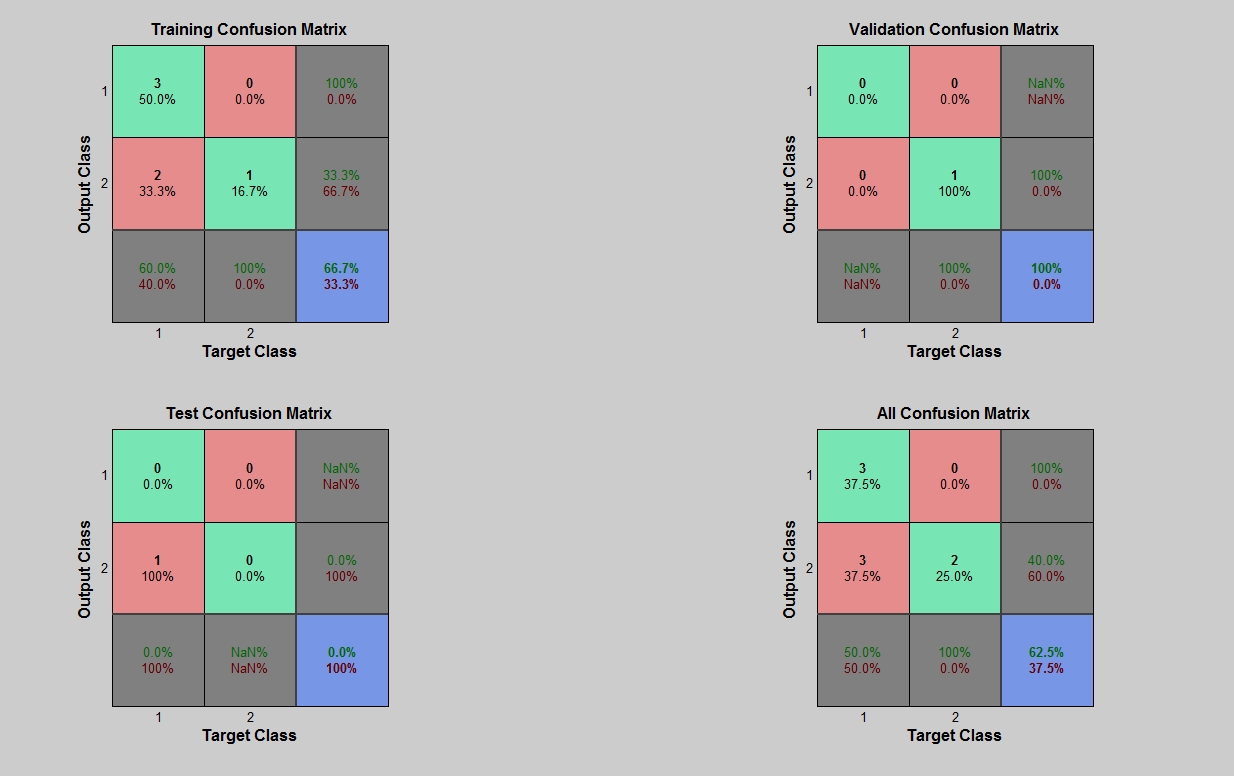
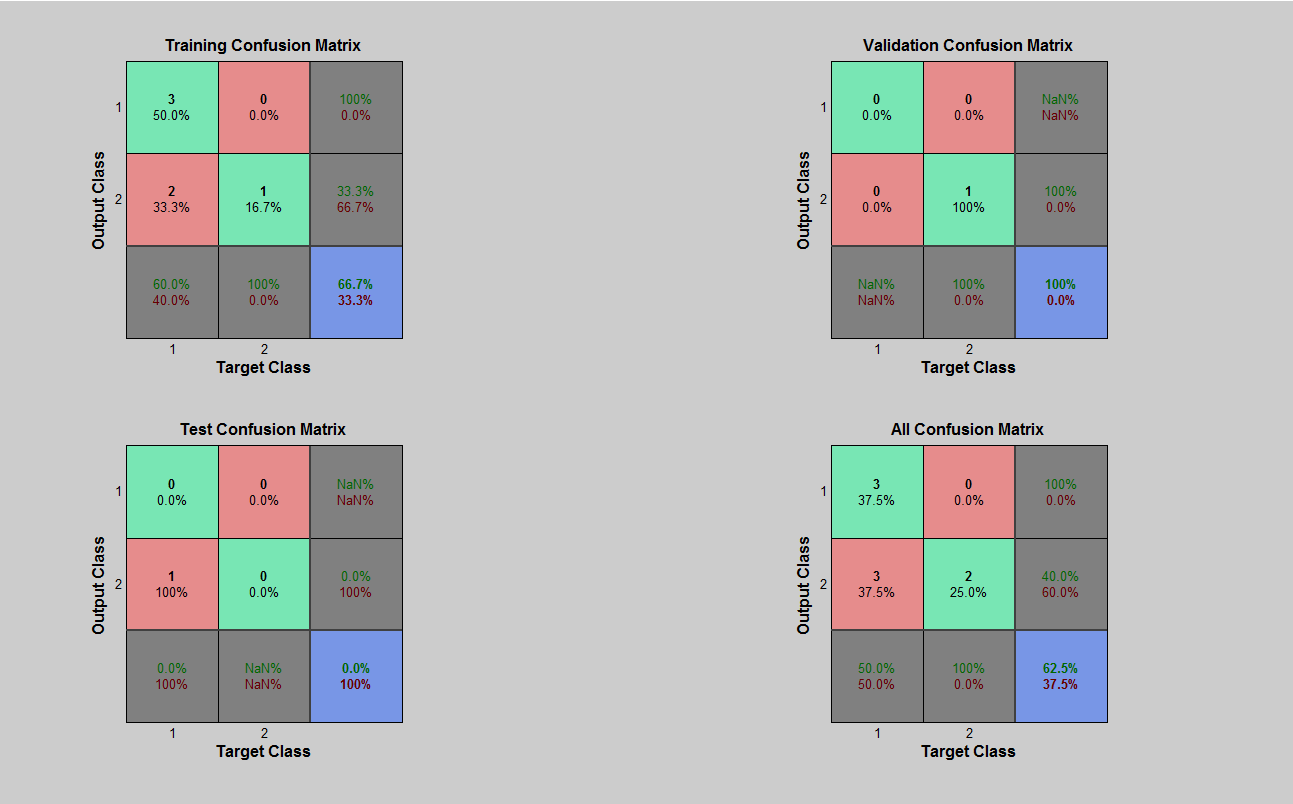
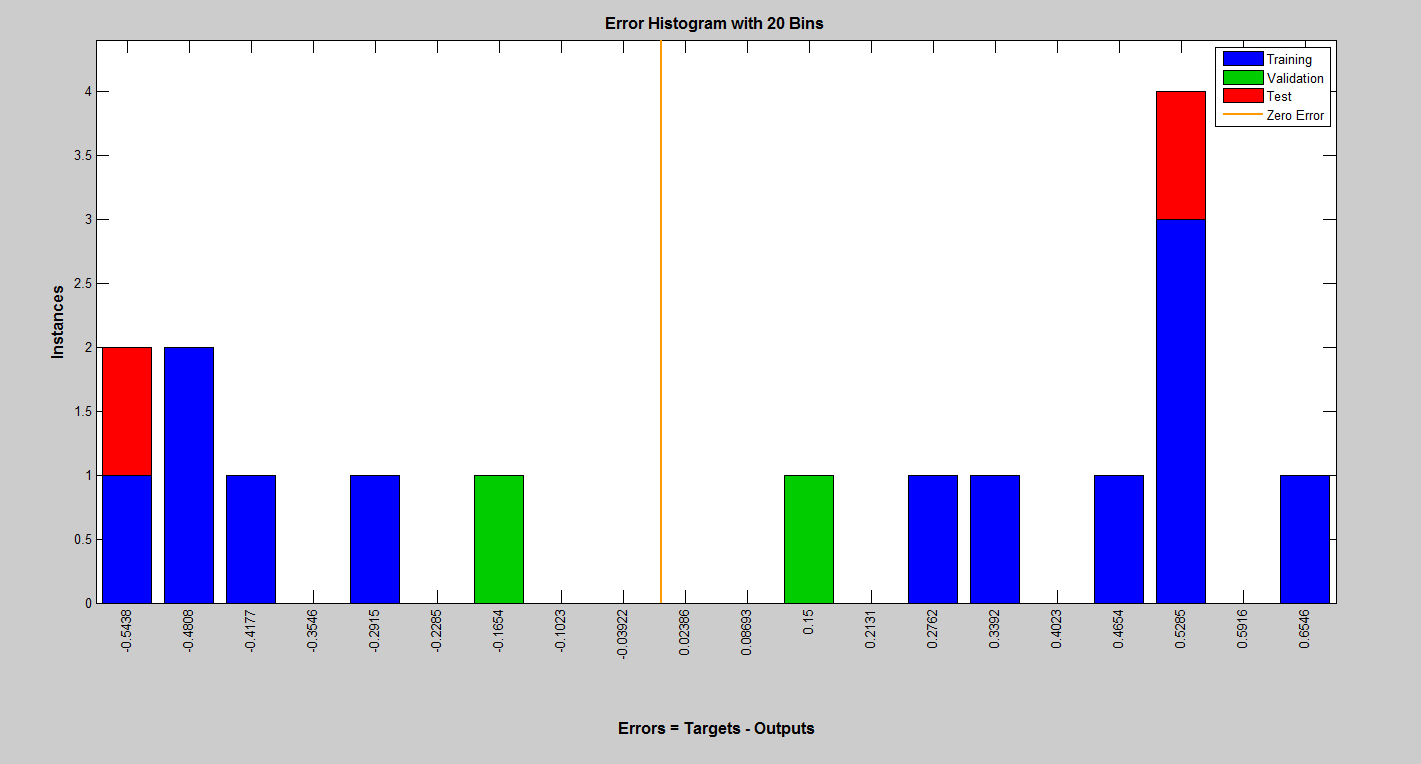
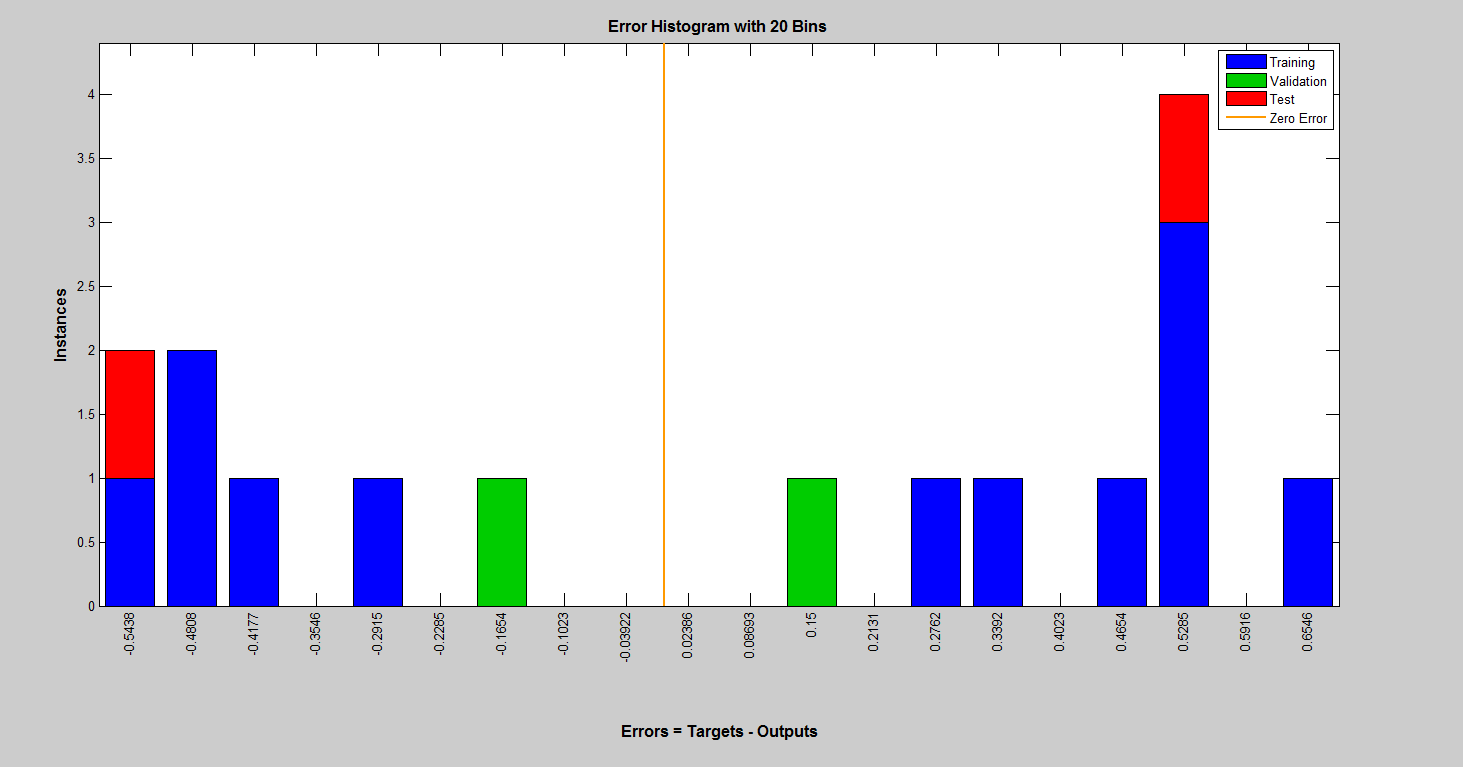
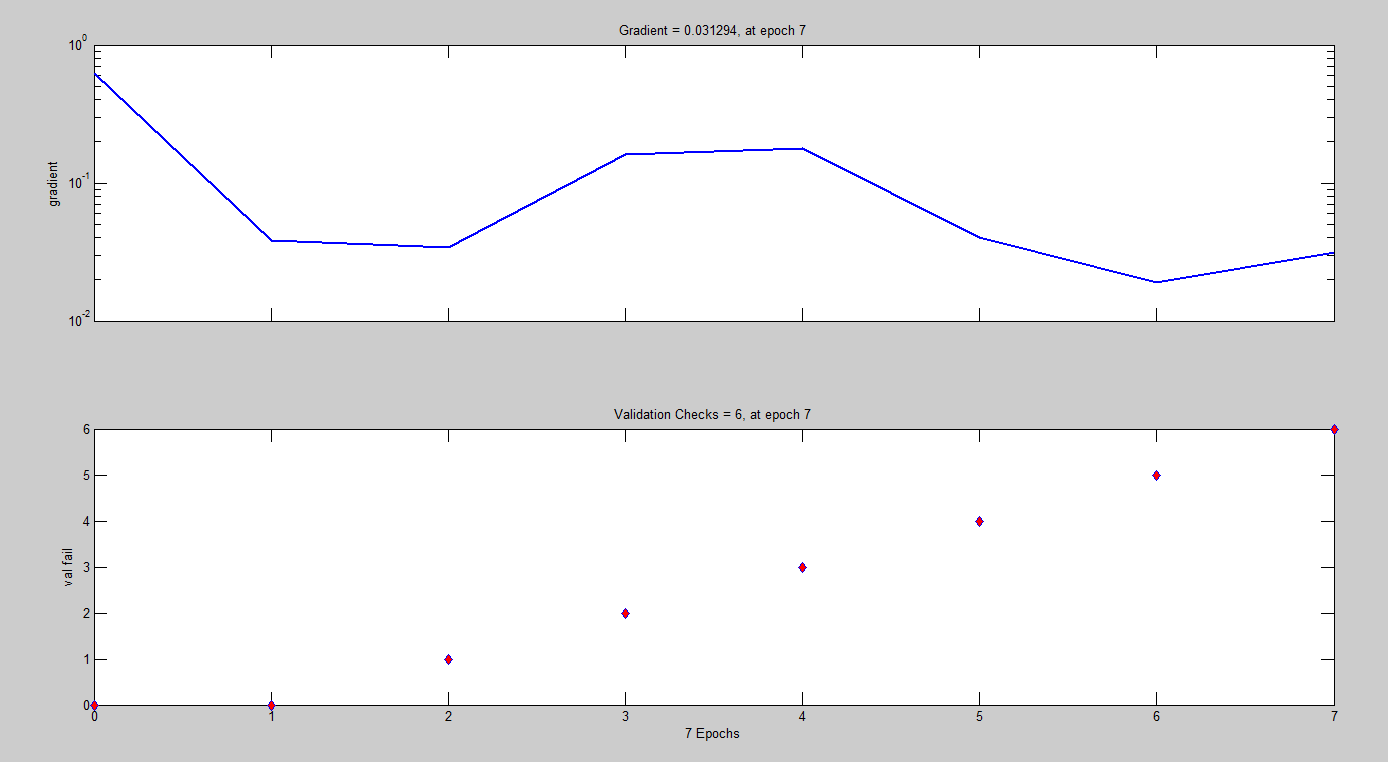
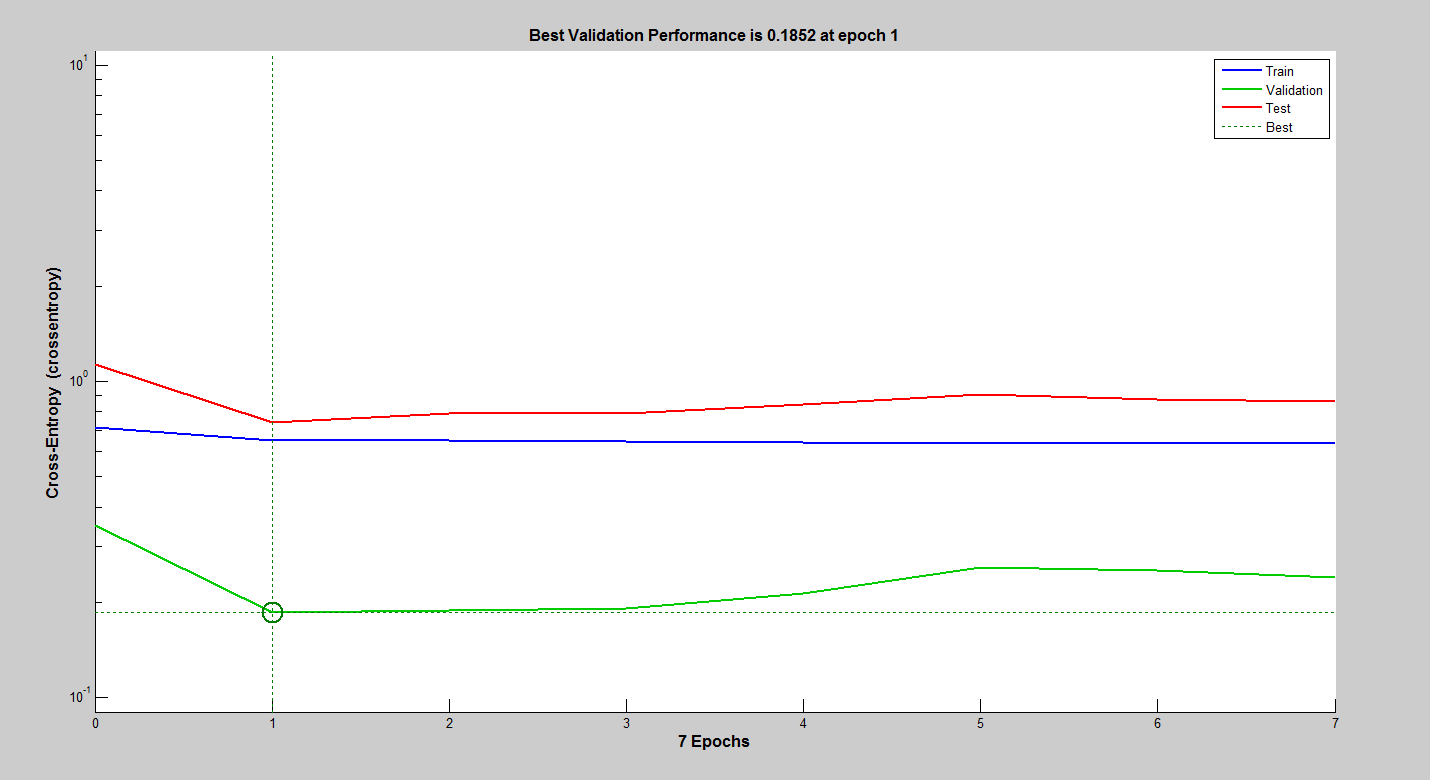
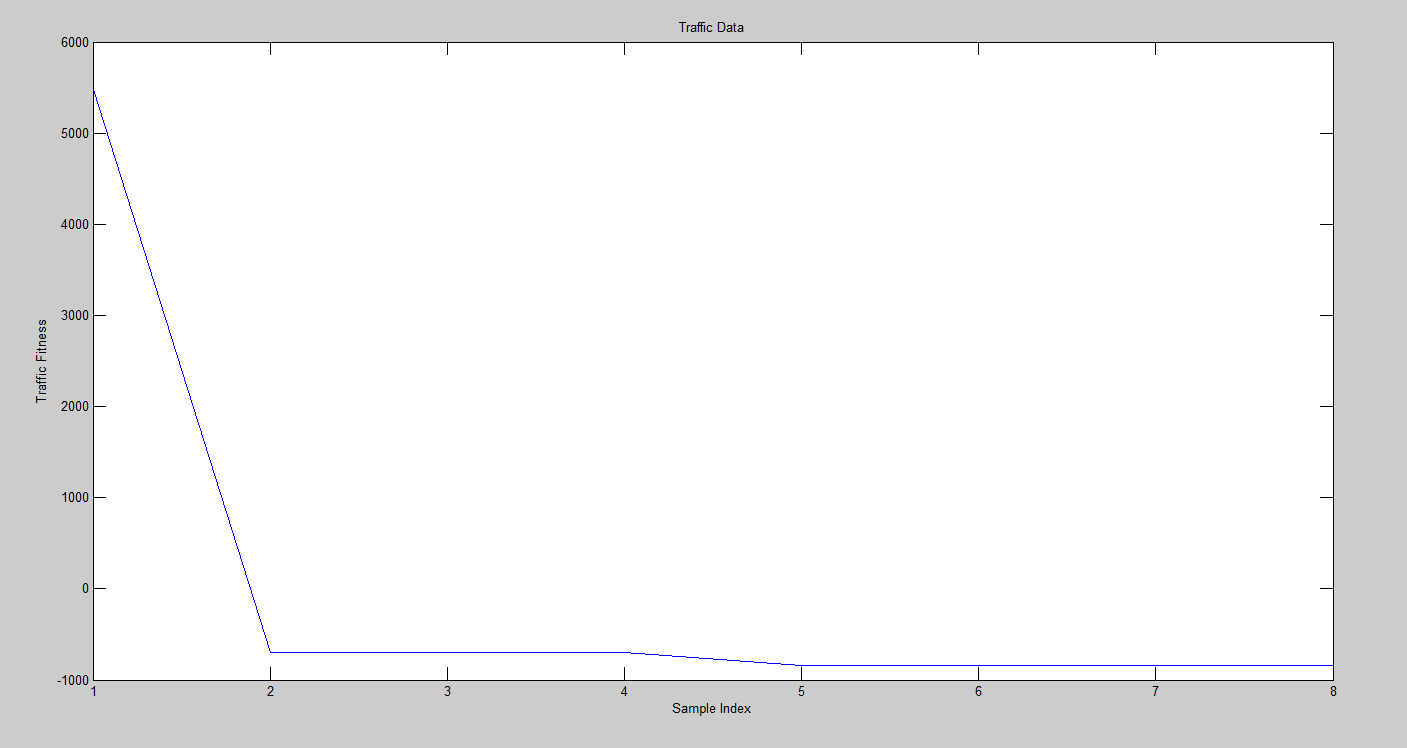
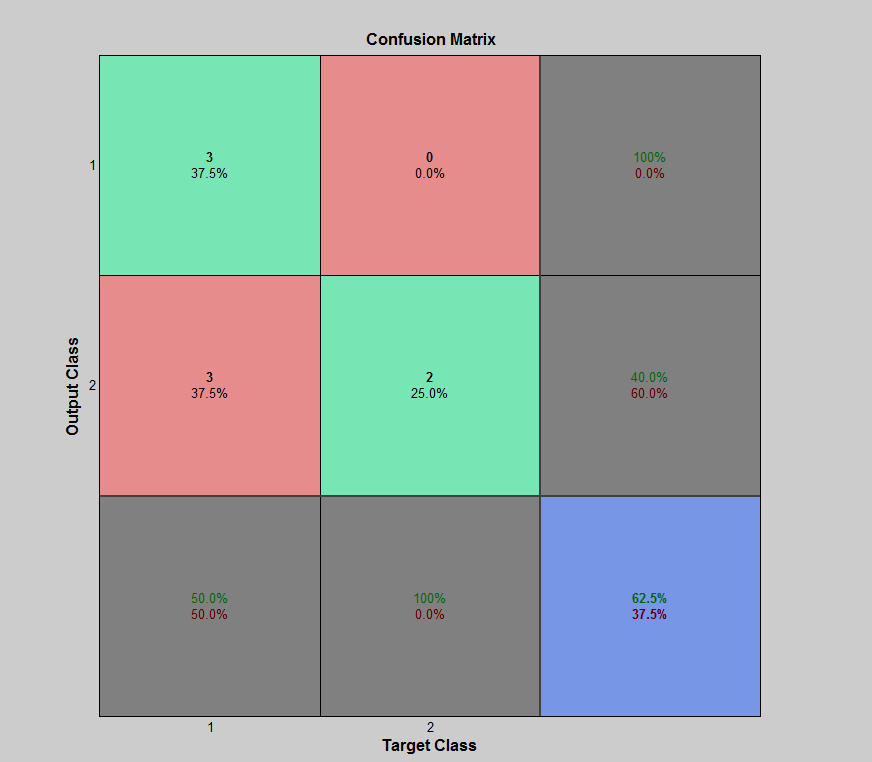
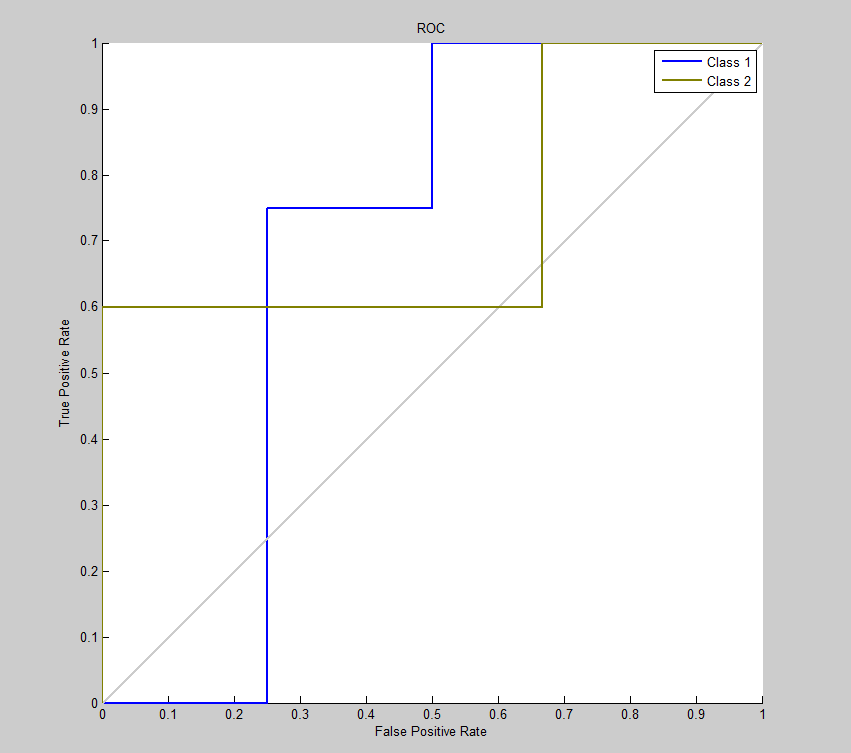
* + 1. **LDA**



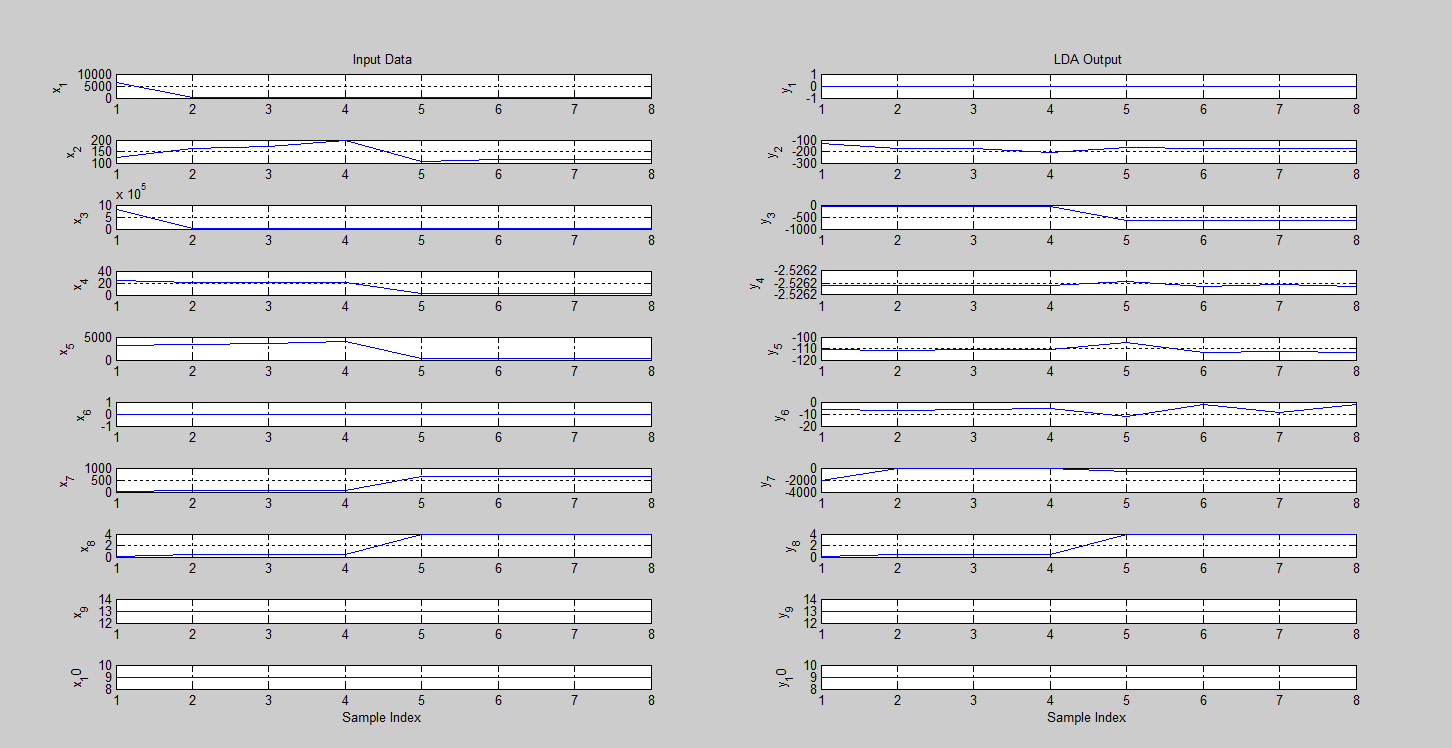
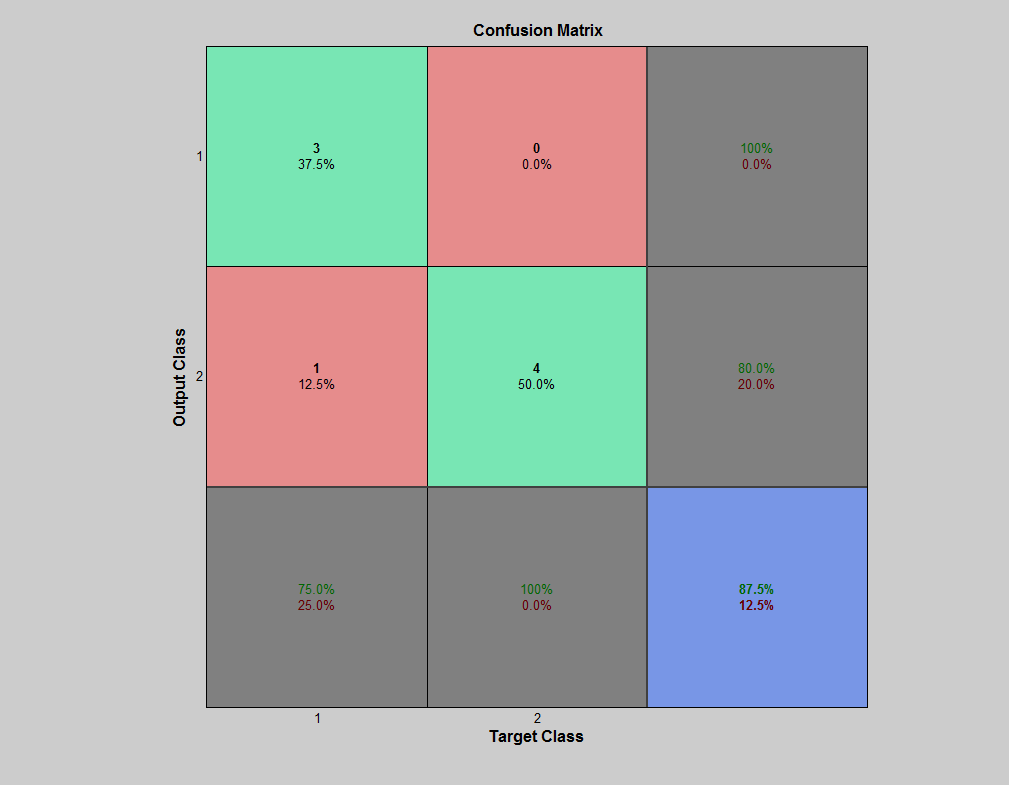
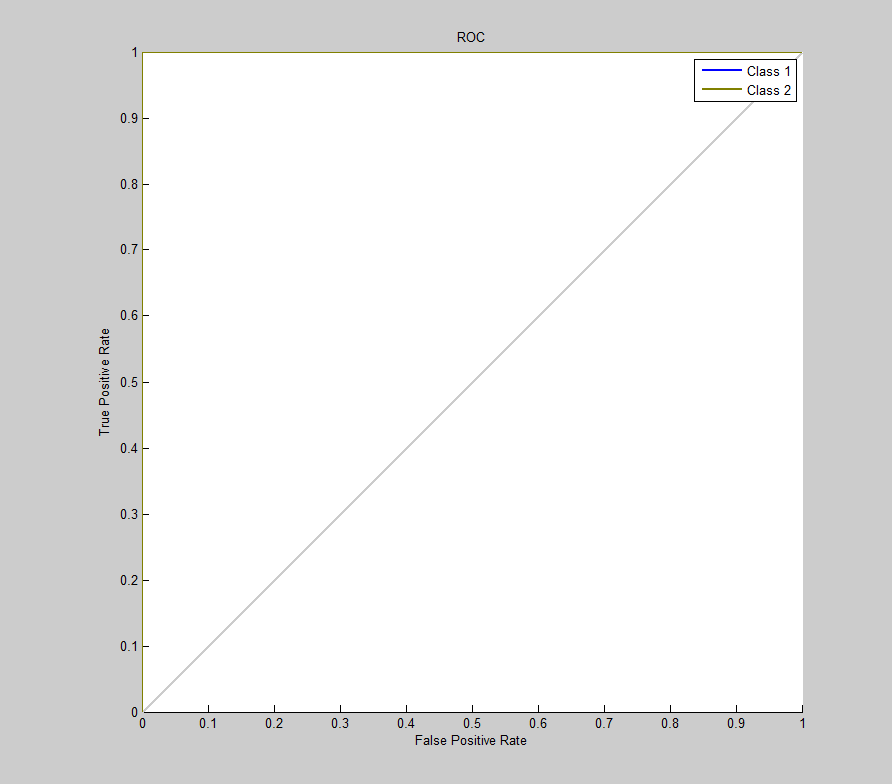
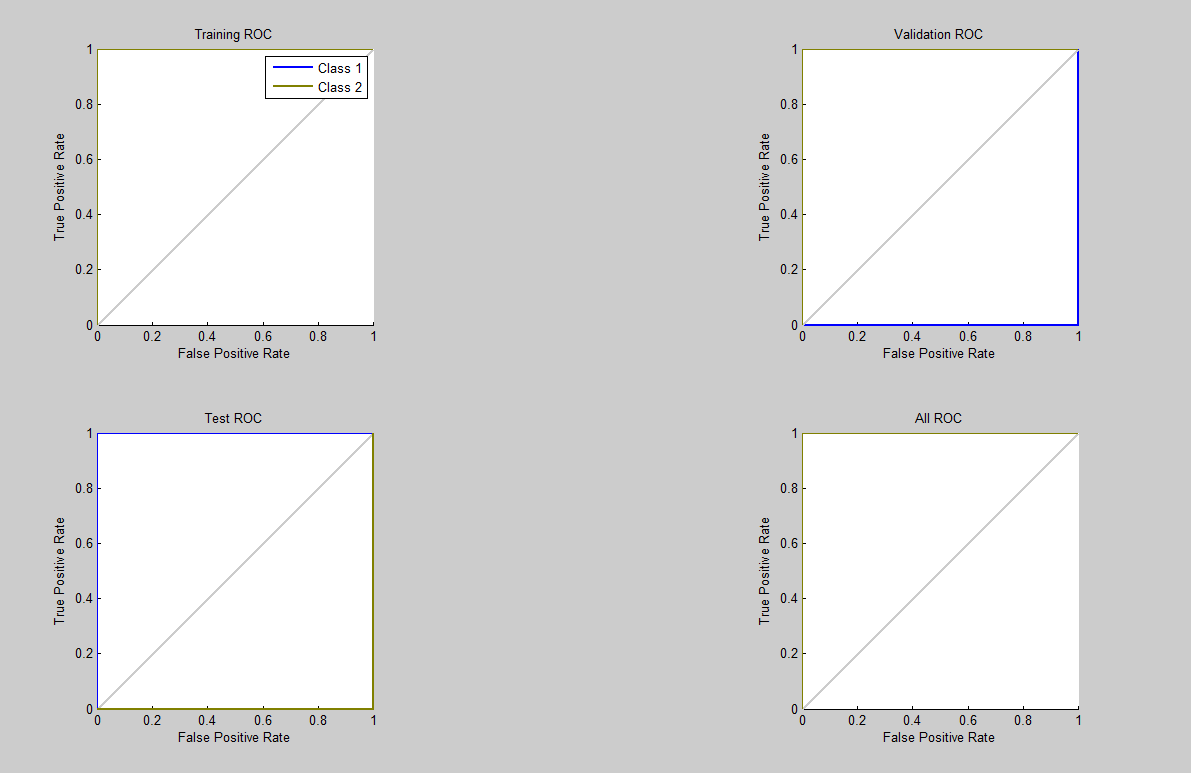
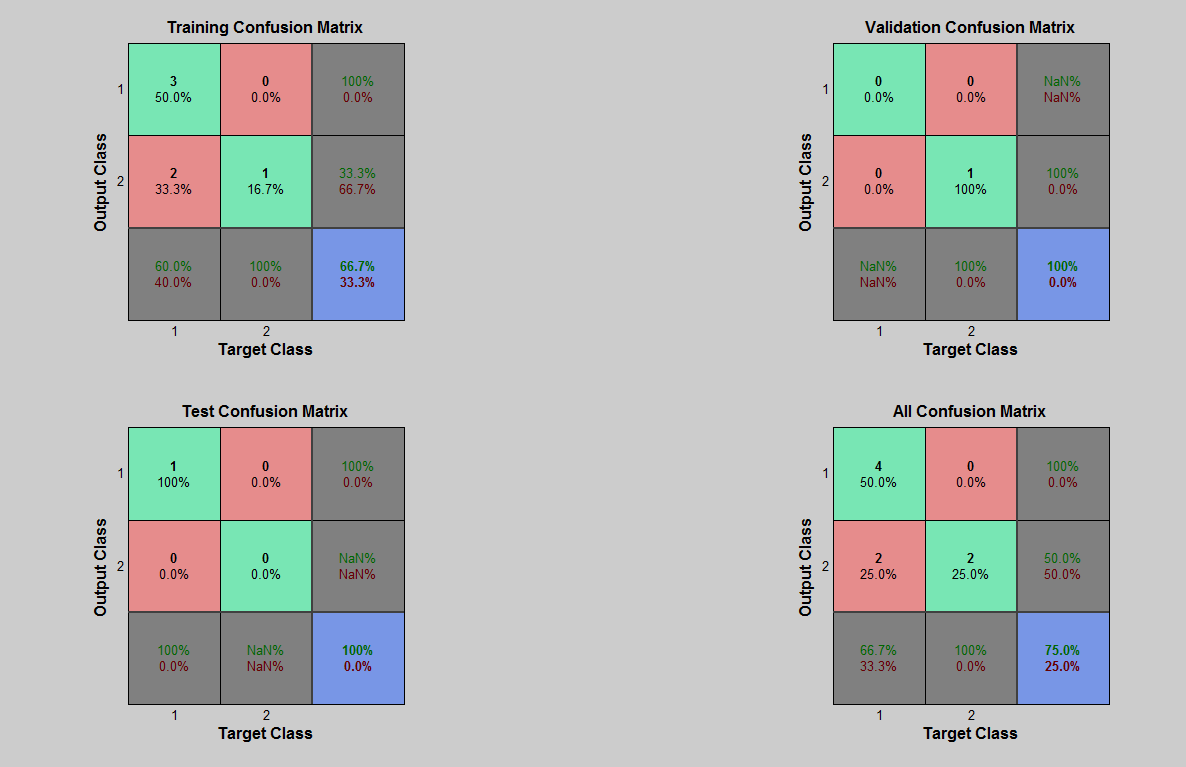
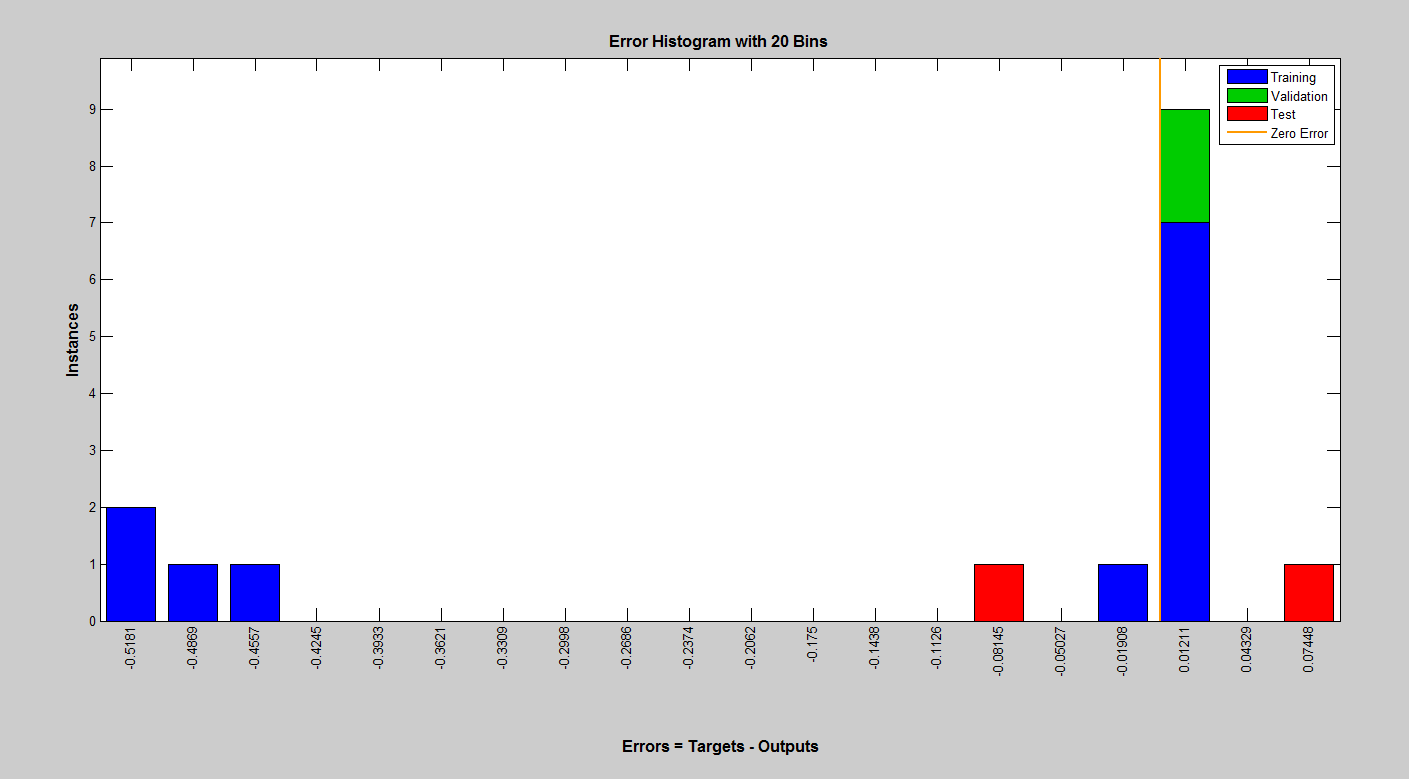
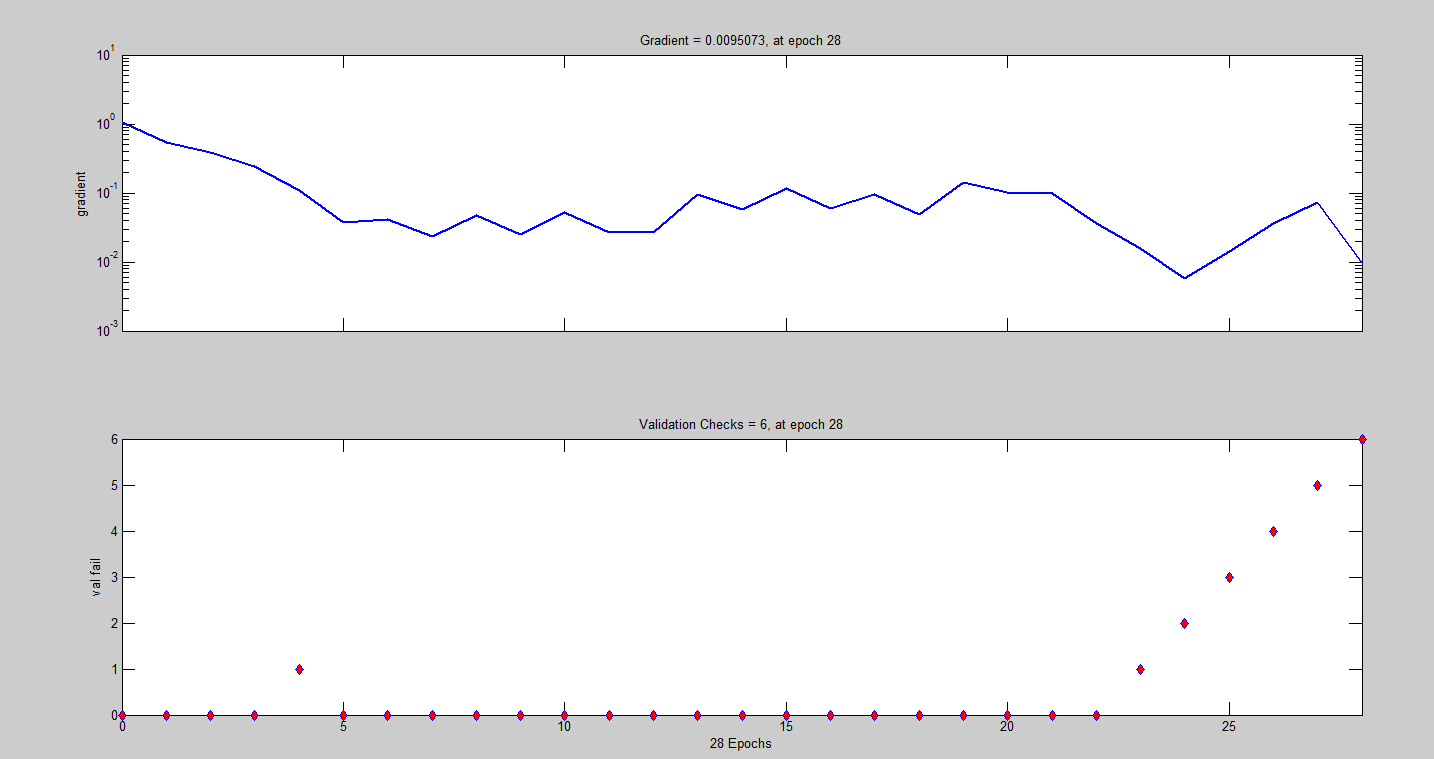
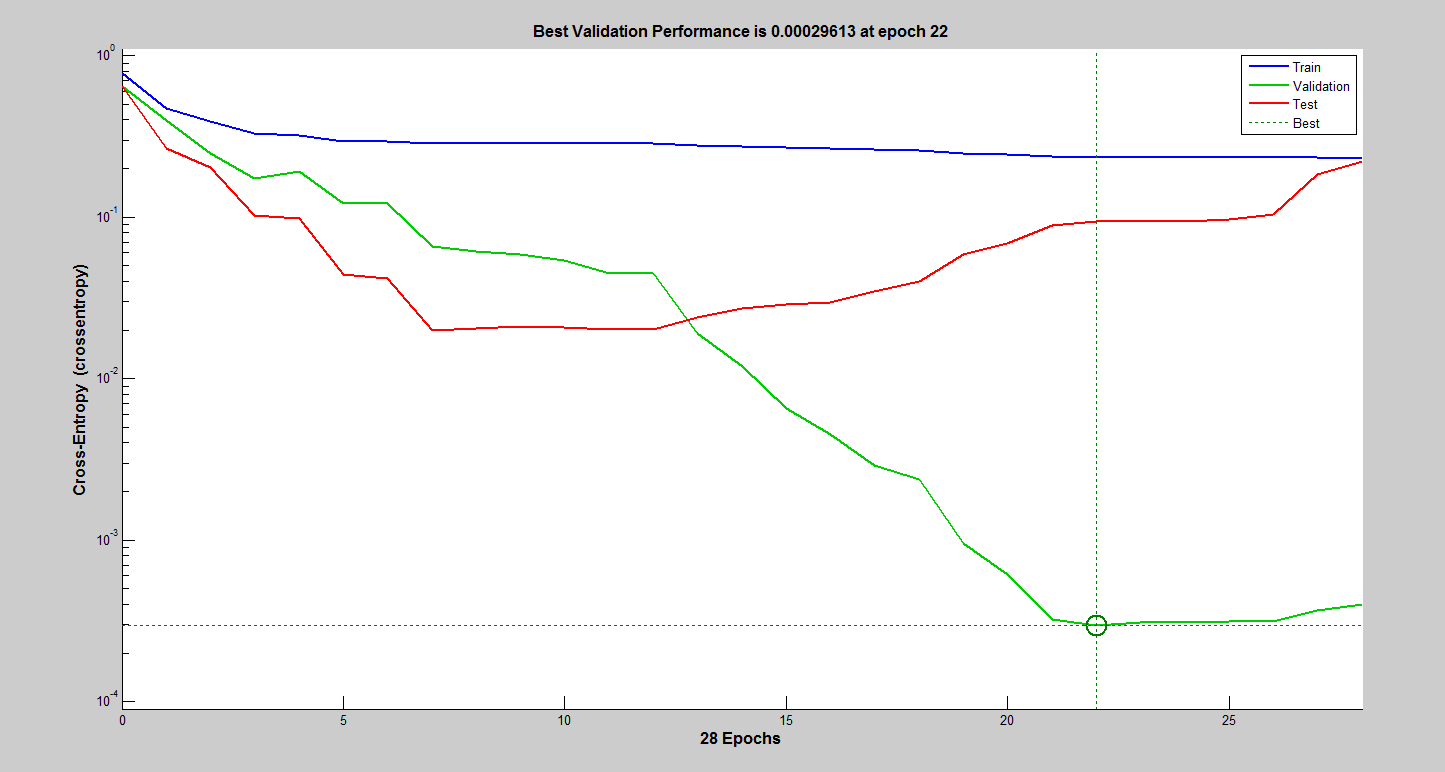
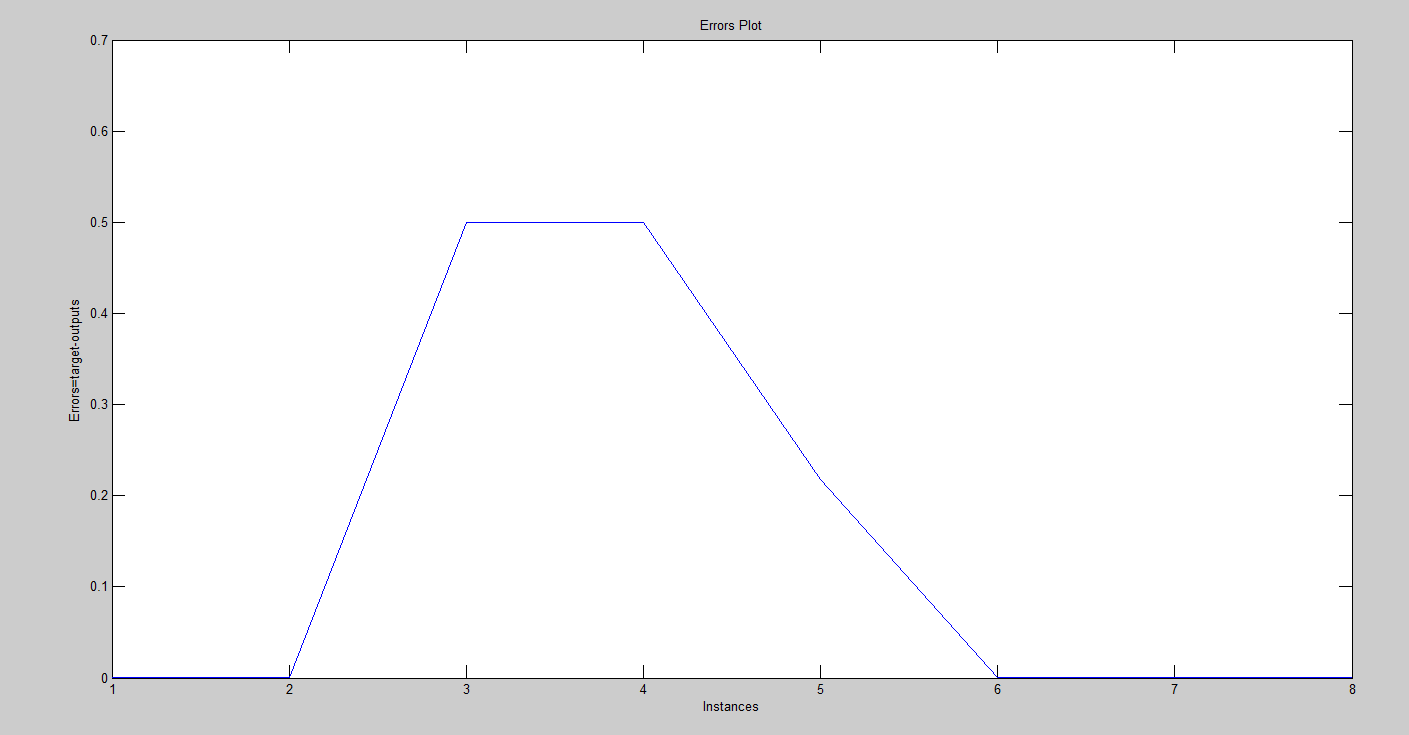
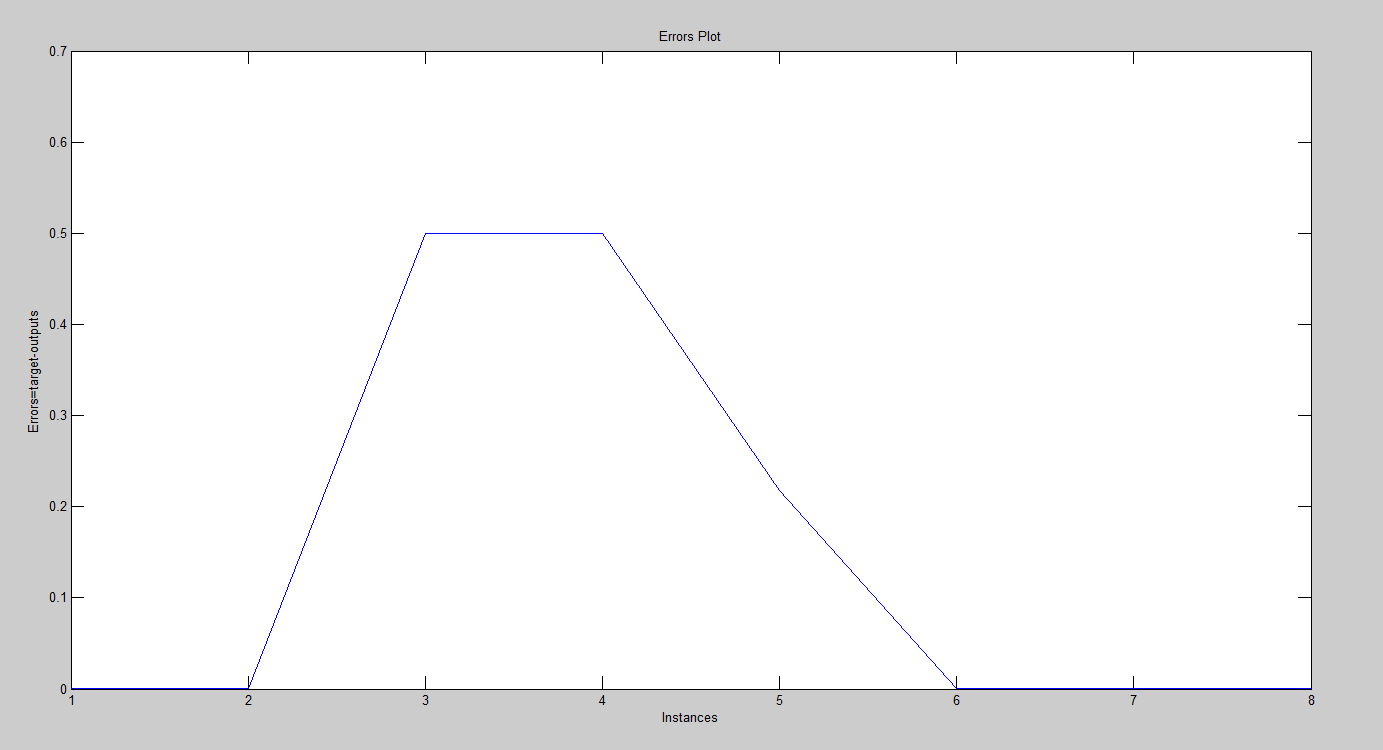
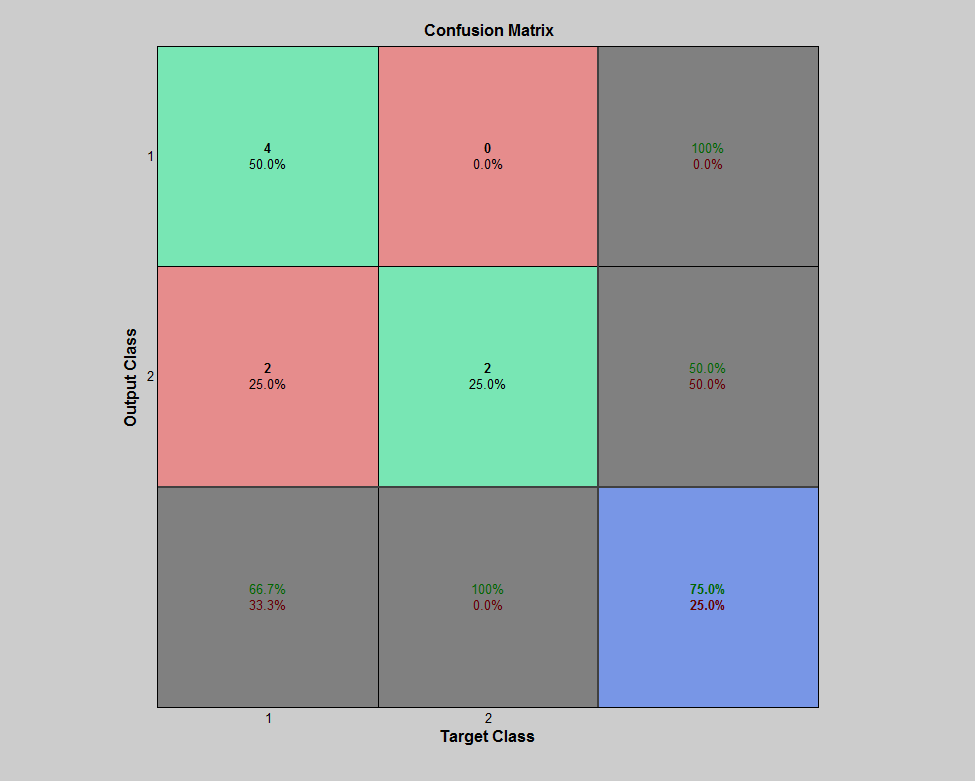
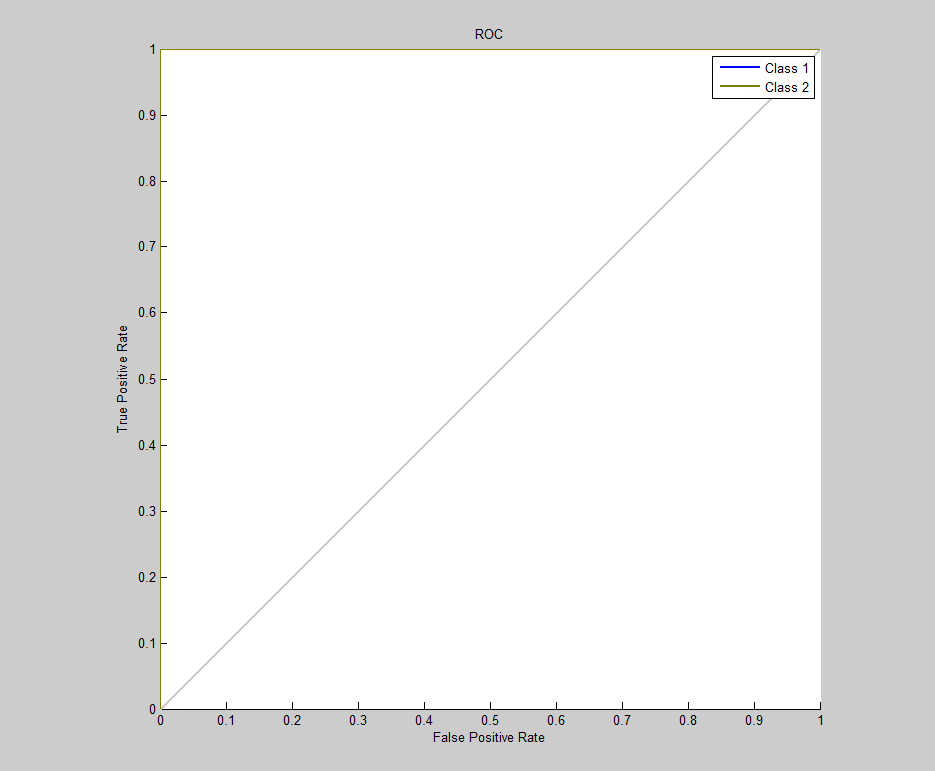
* + 1. **Proposed (PCA+LDA+ALO)**



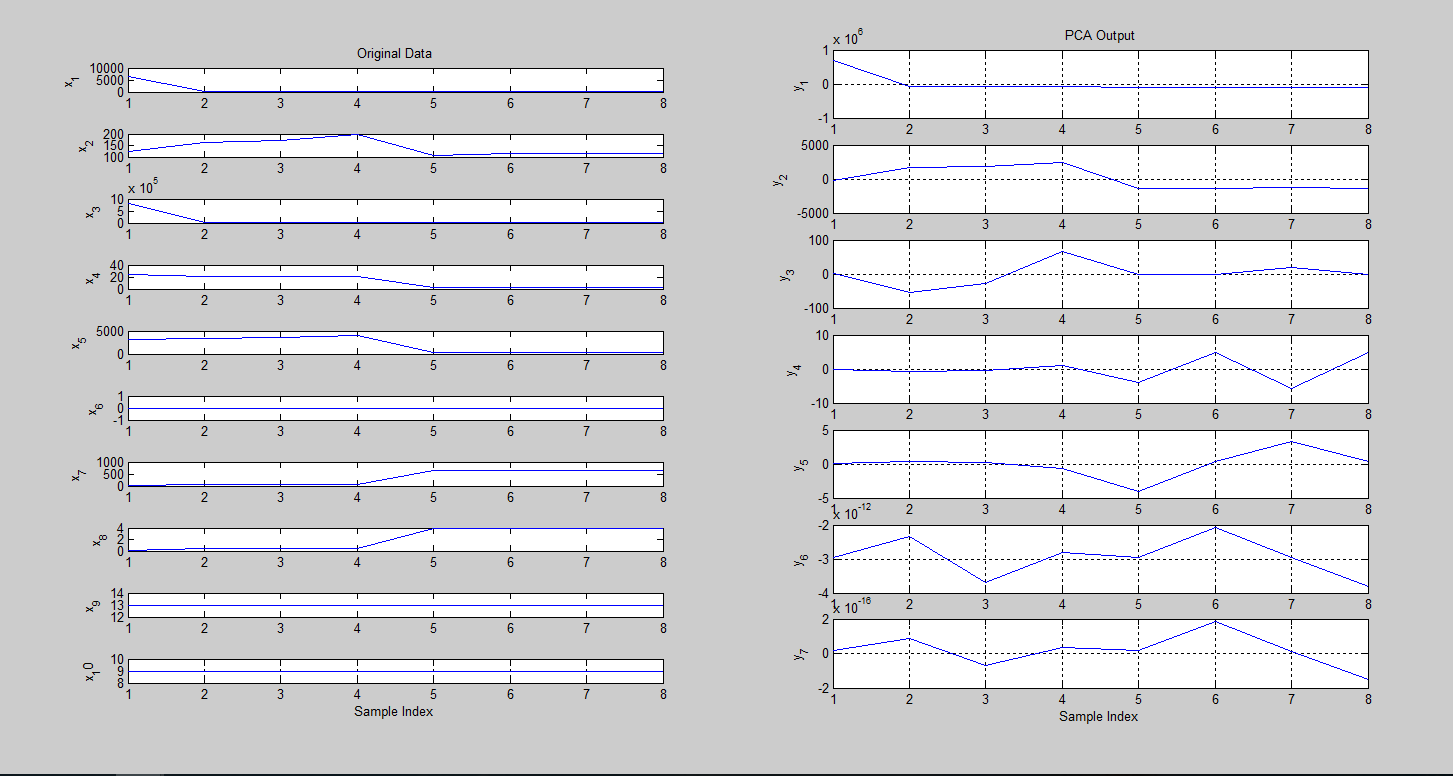
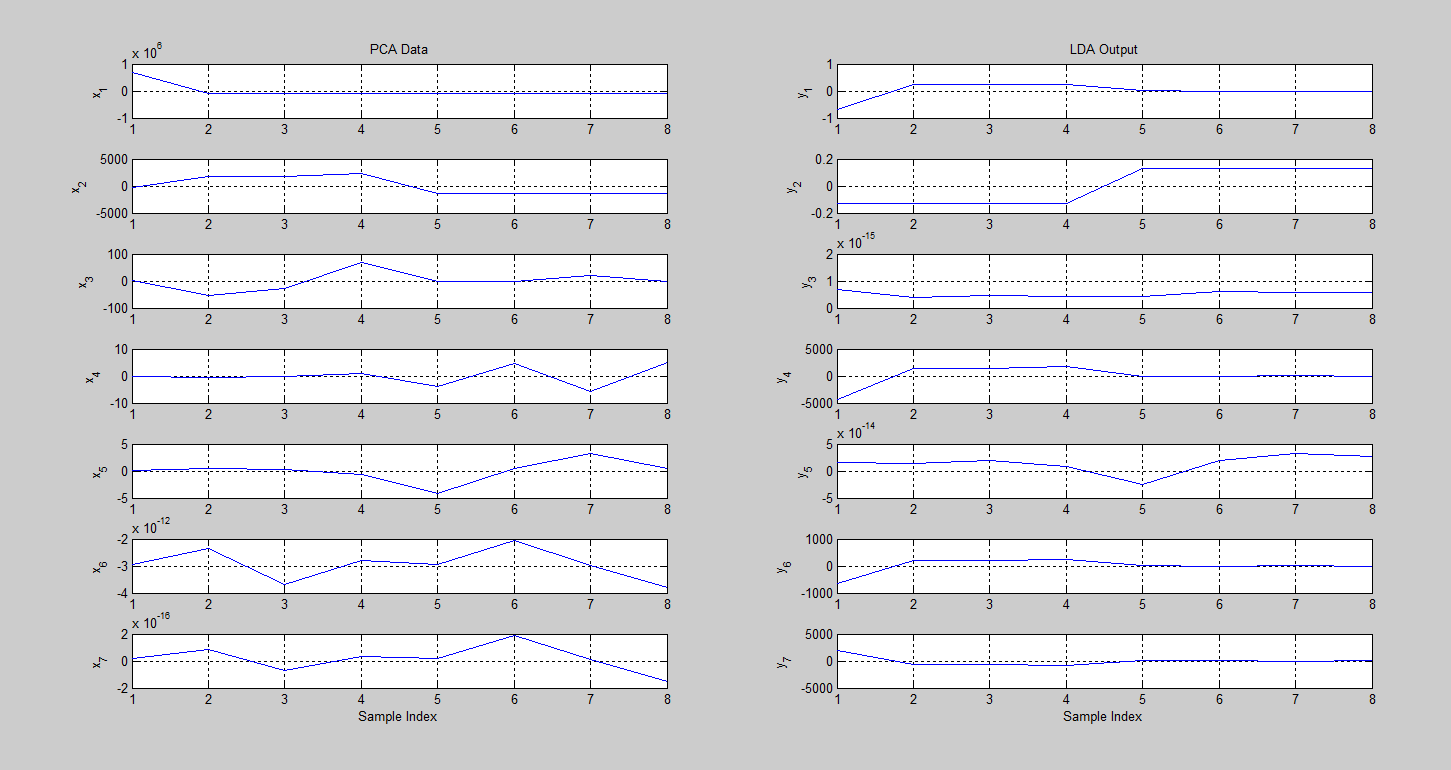
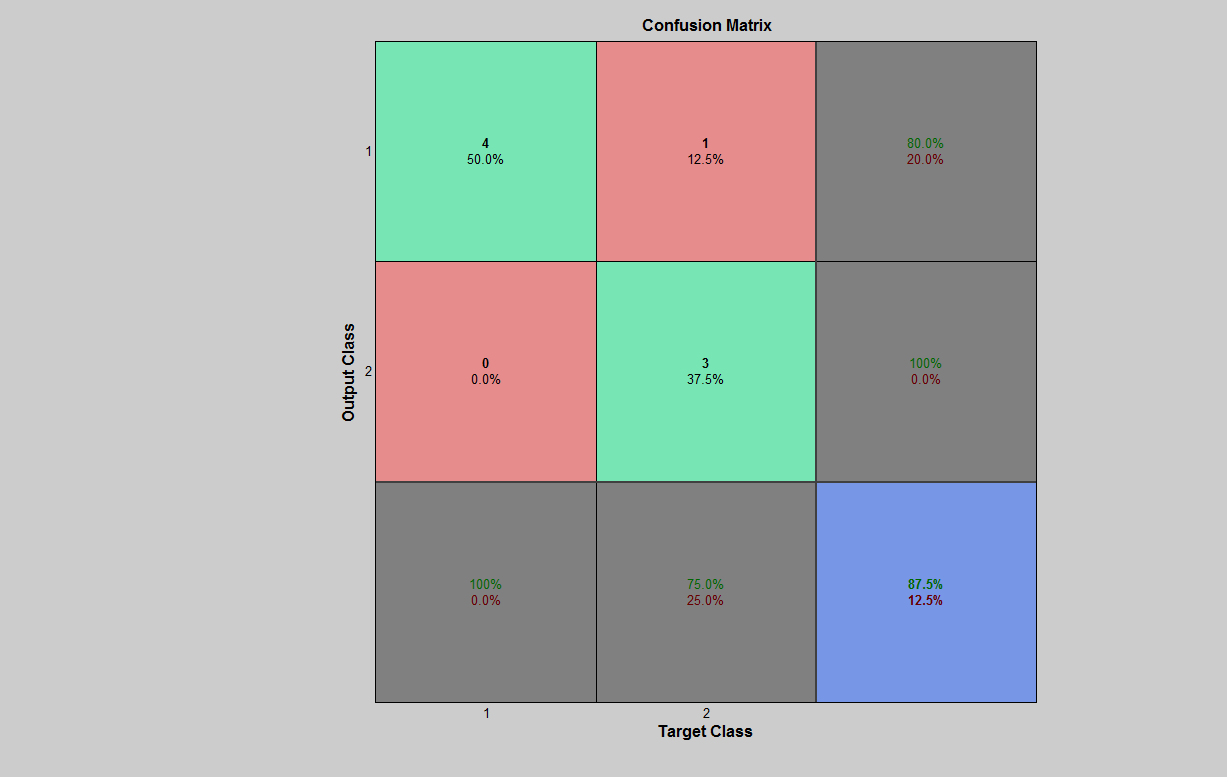
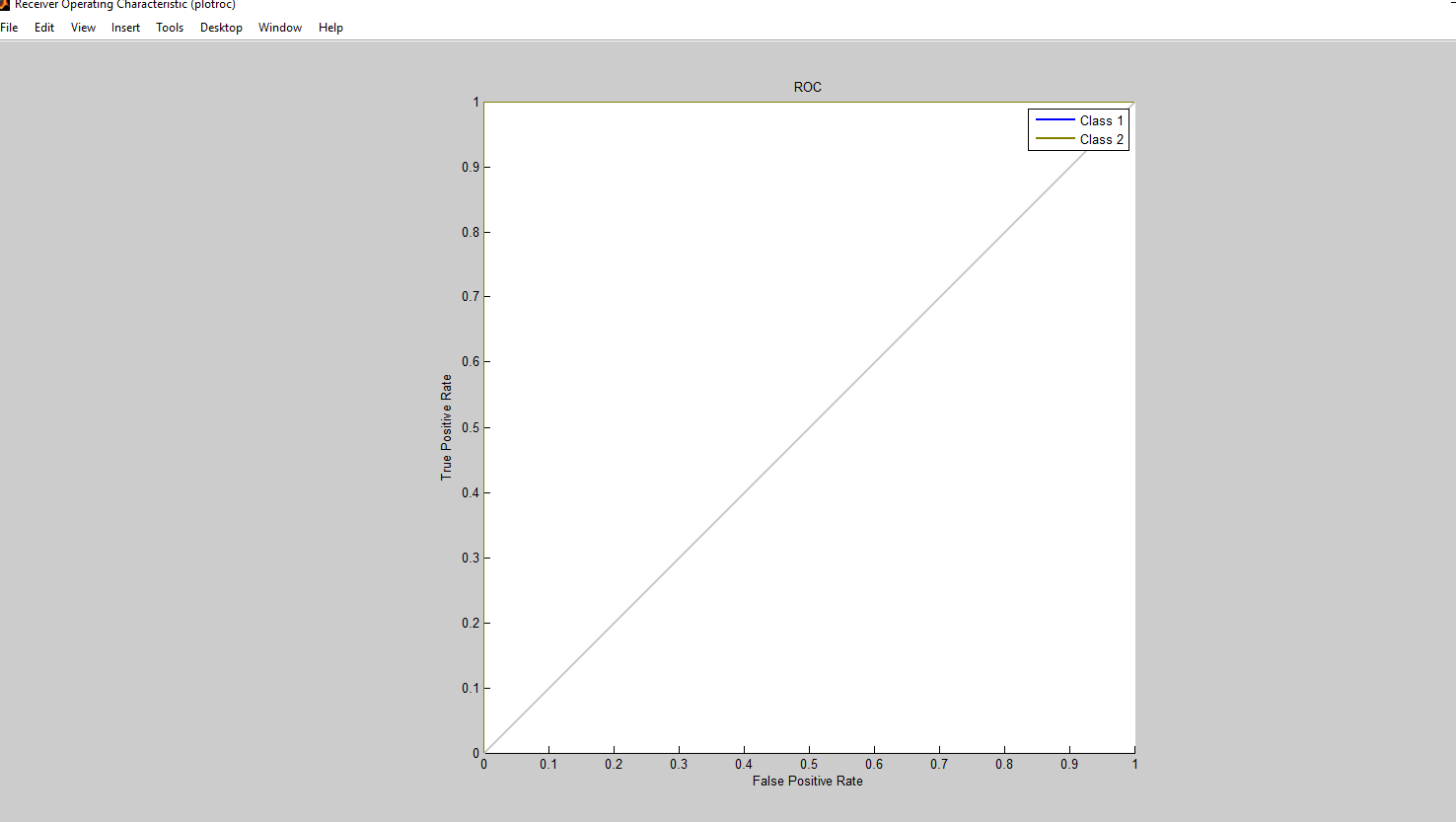
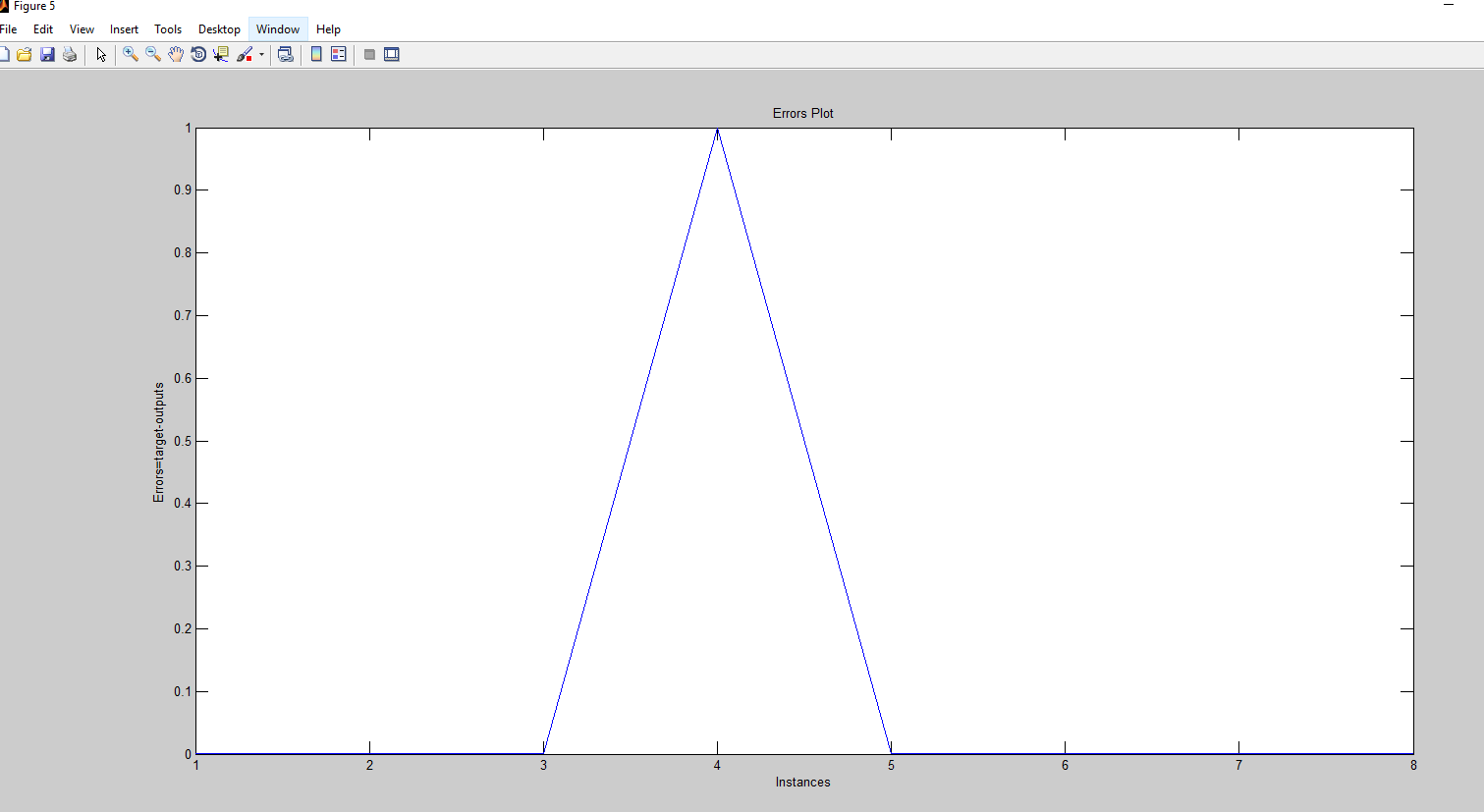
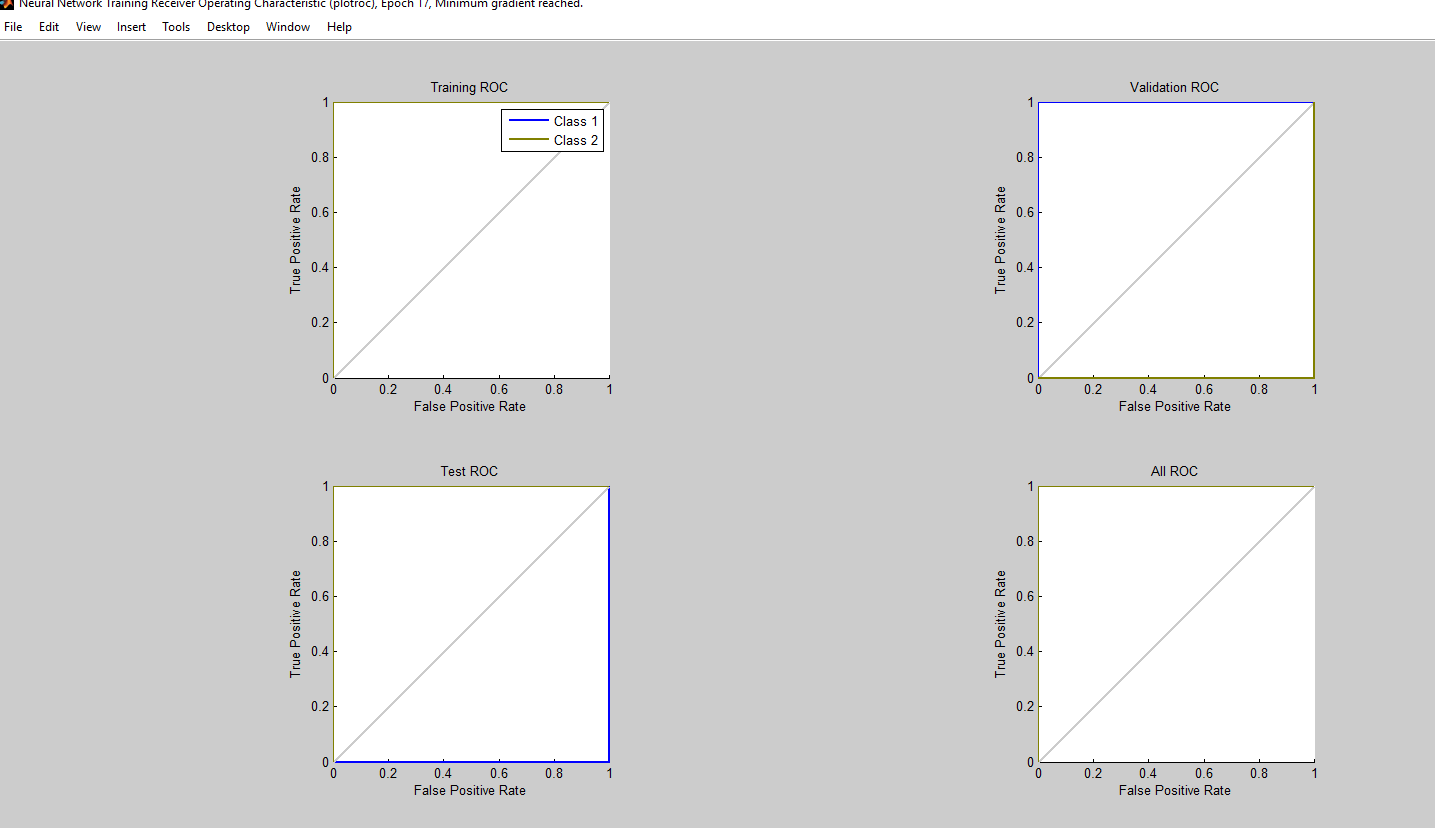
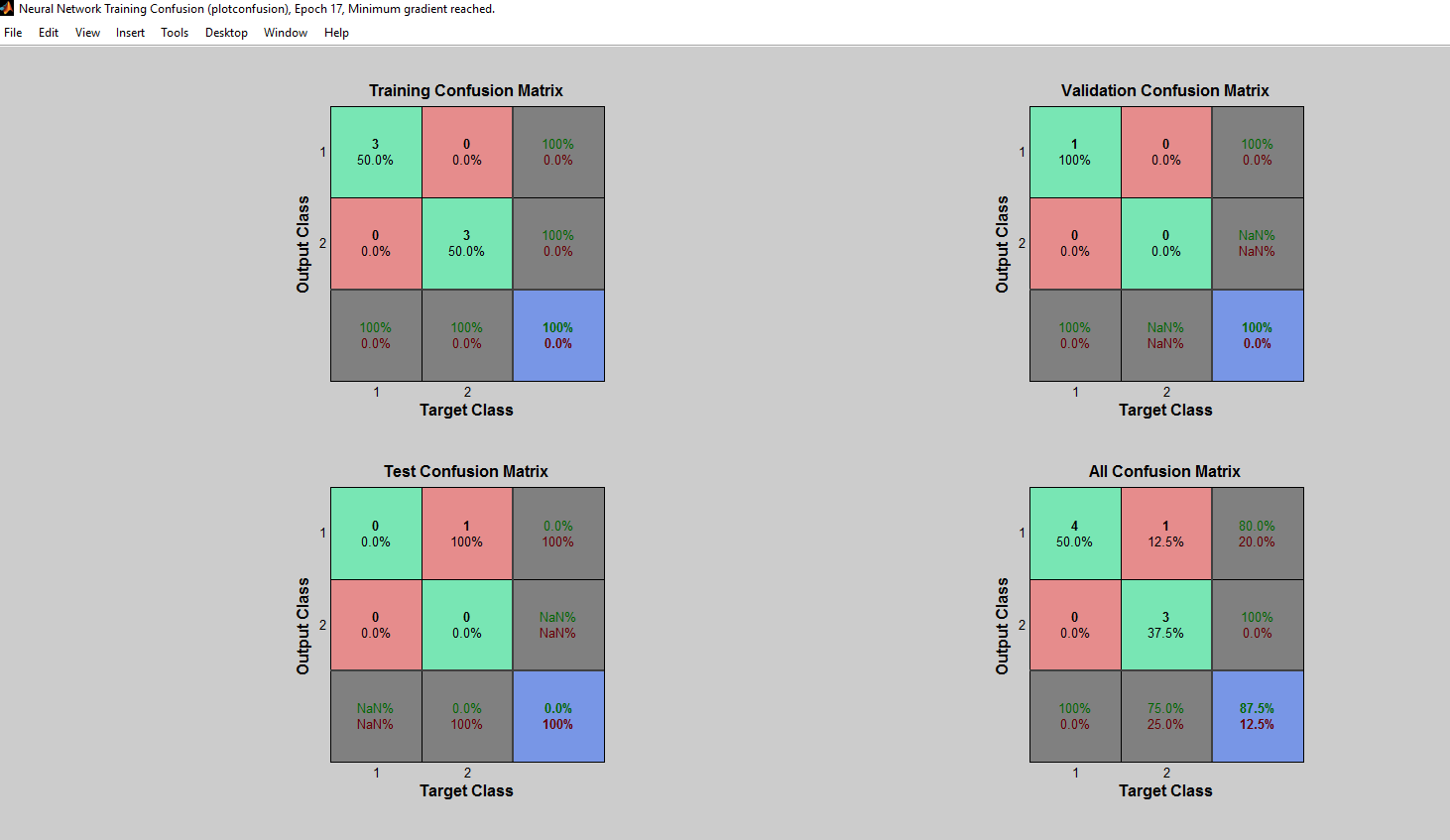
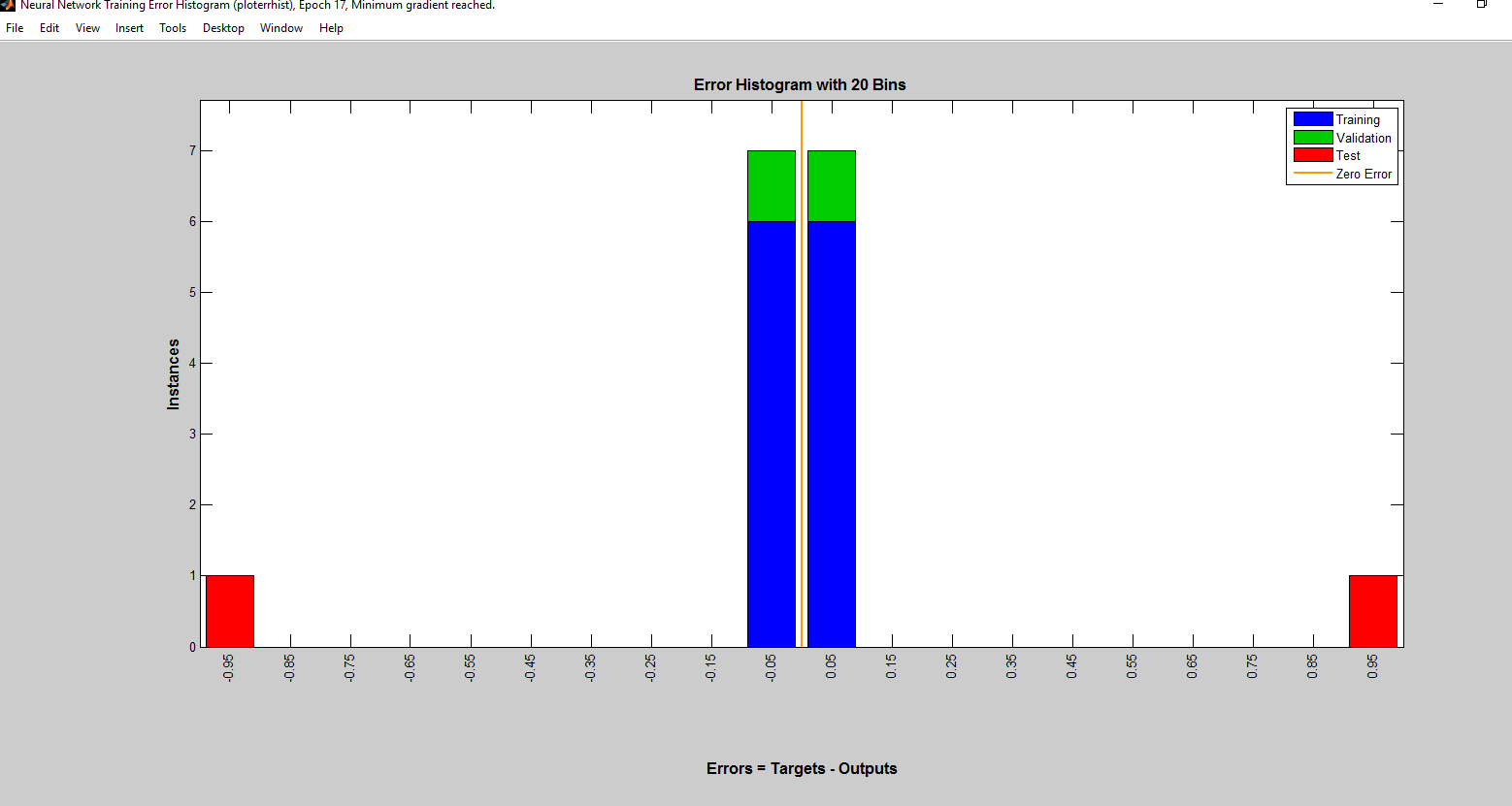
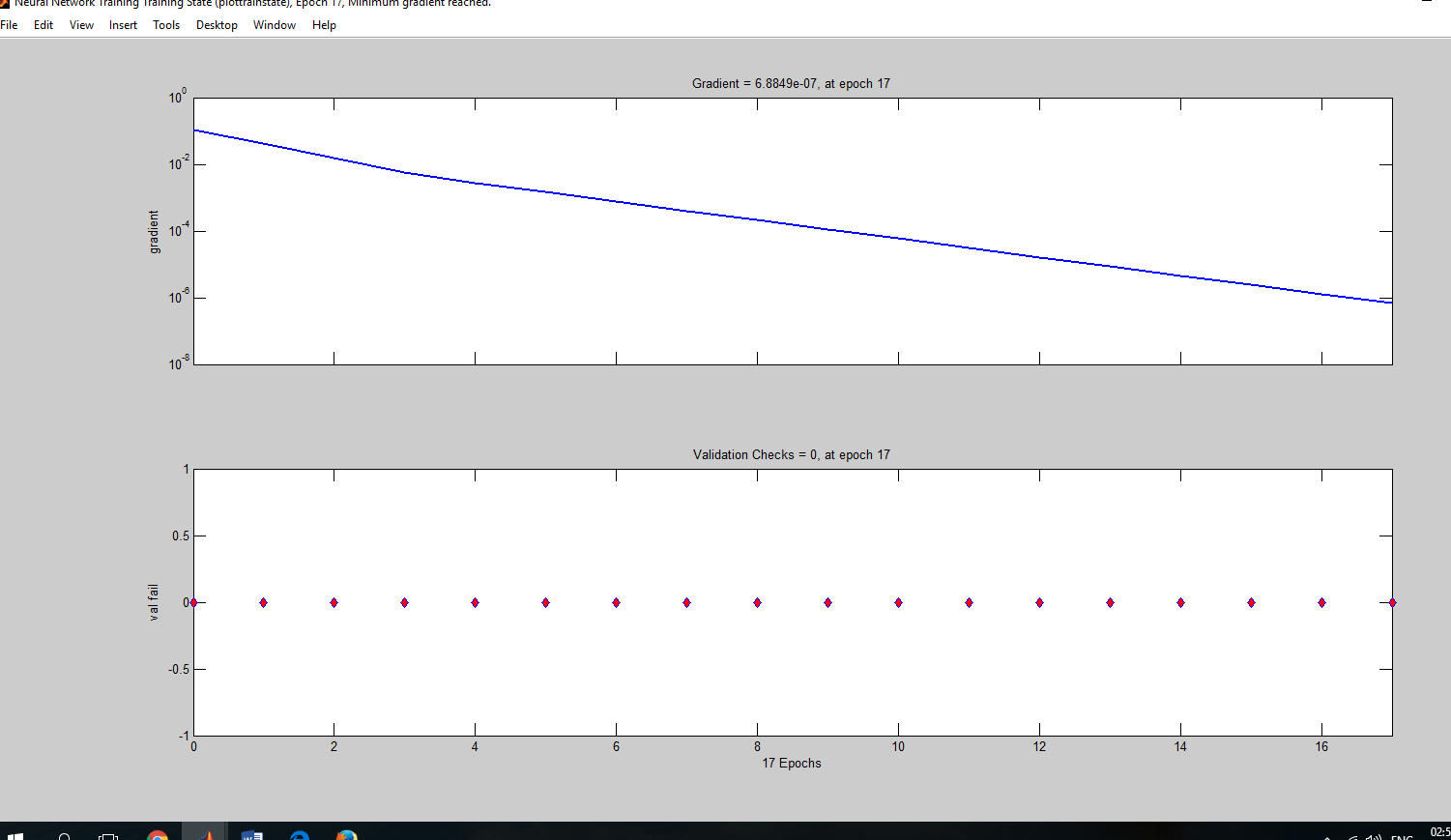
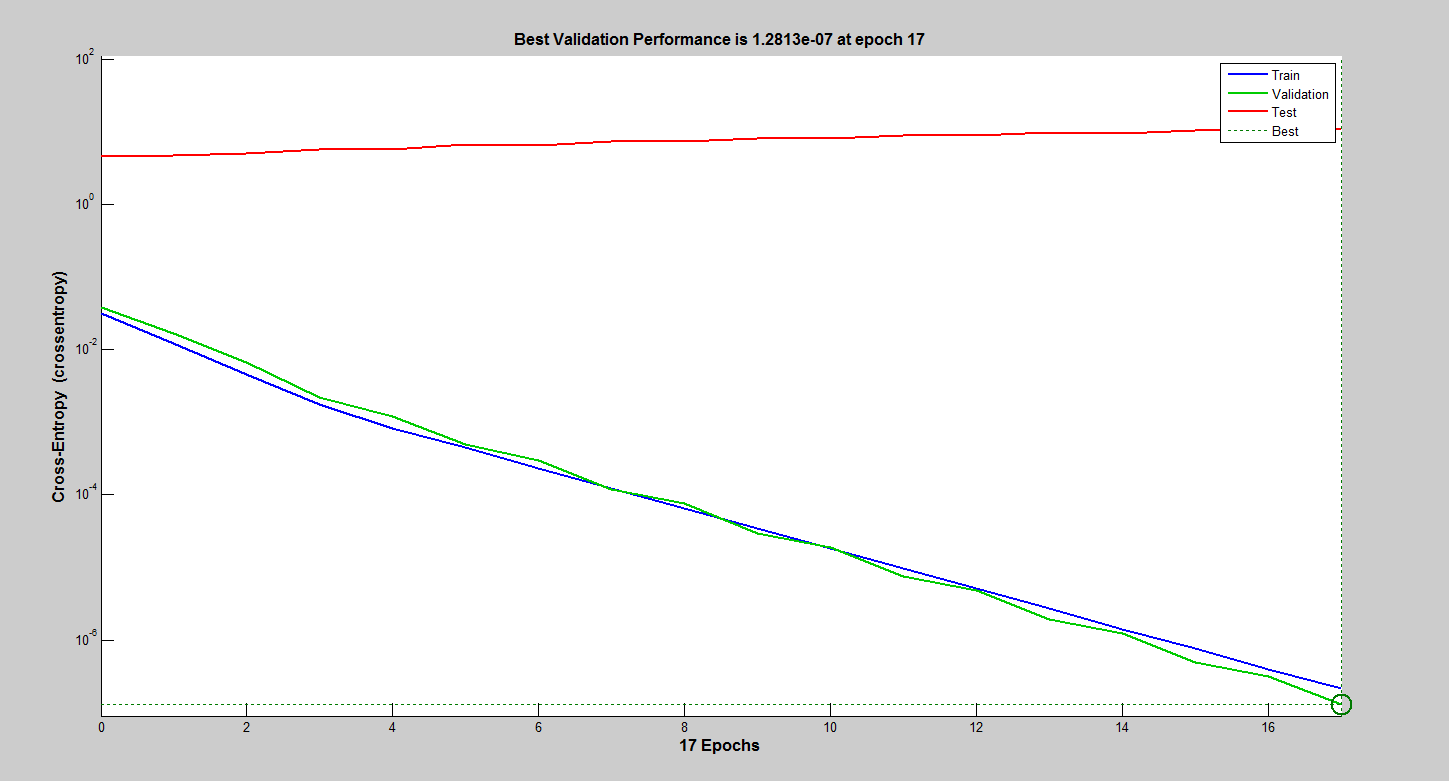
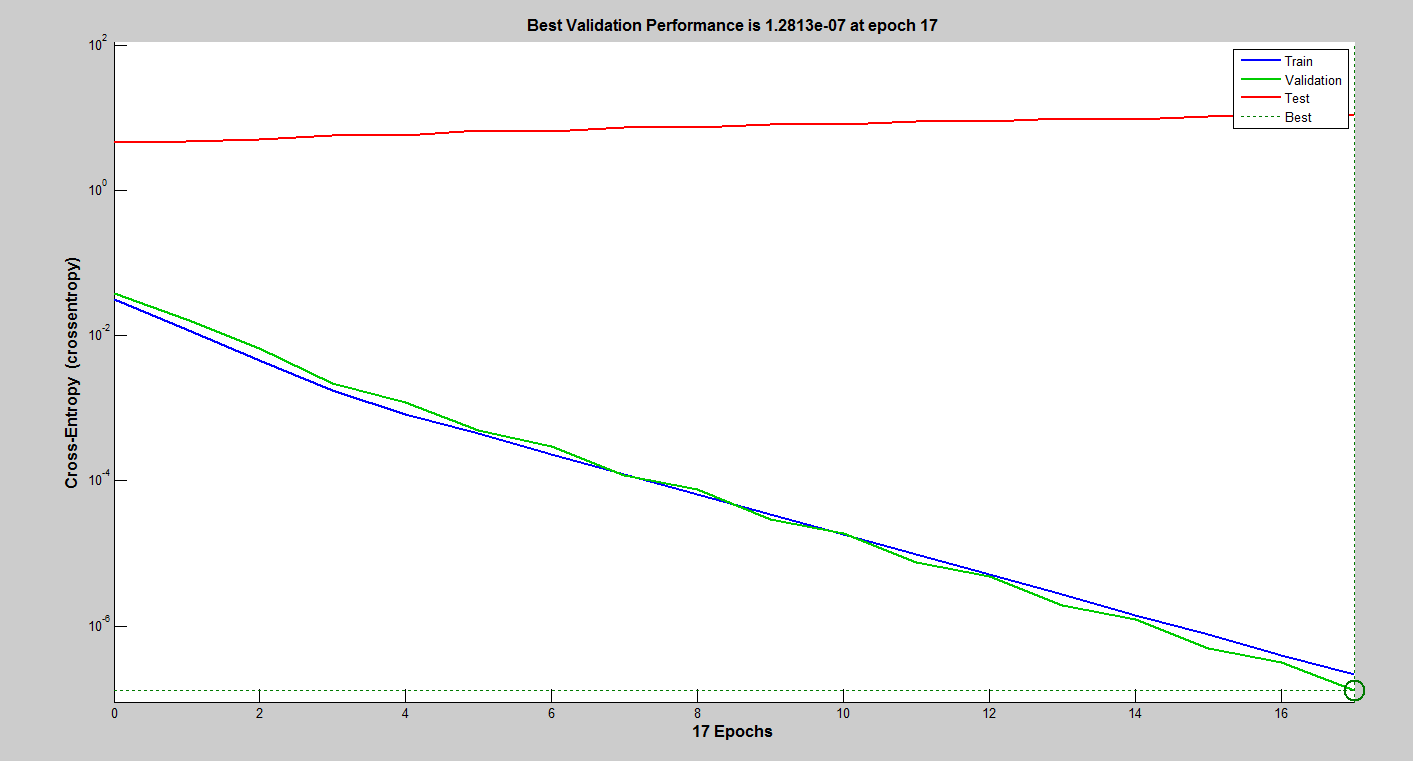
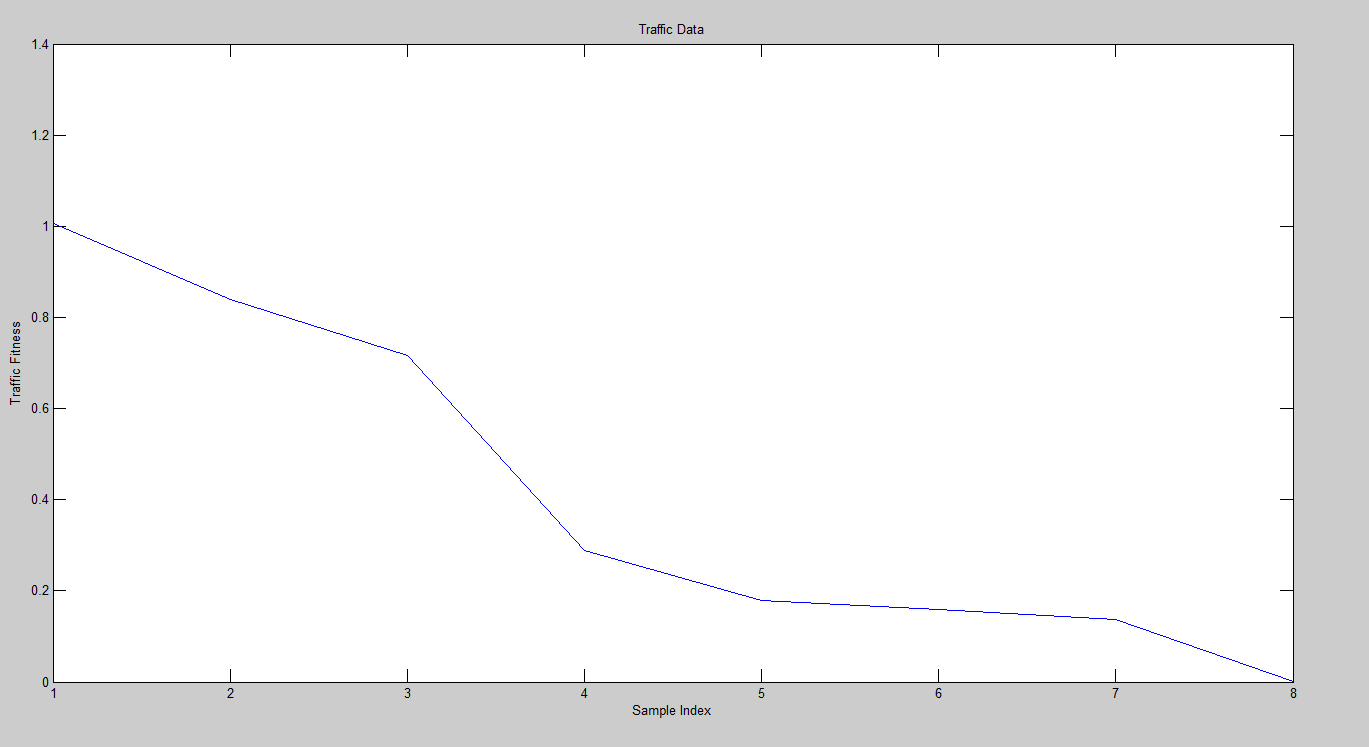
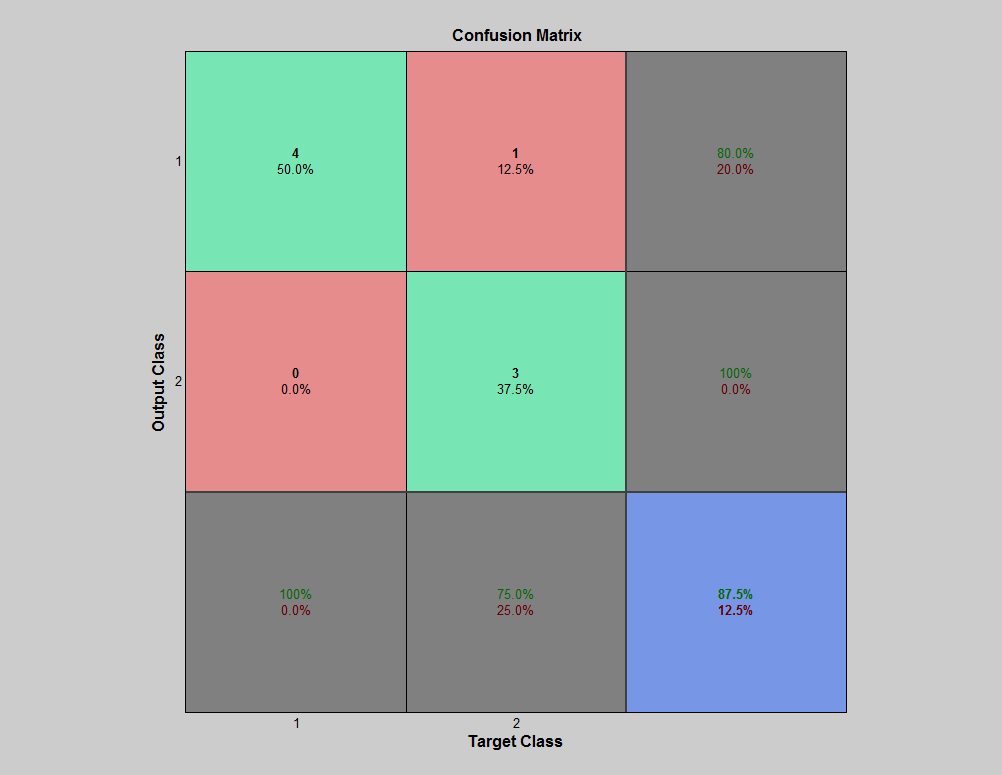
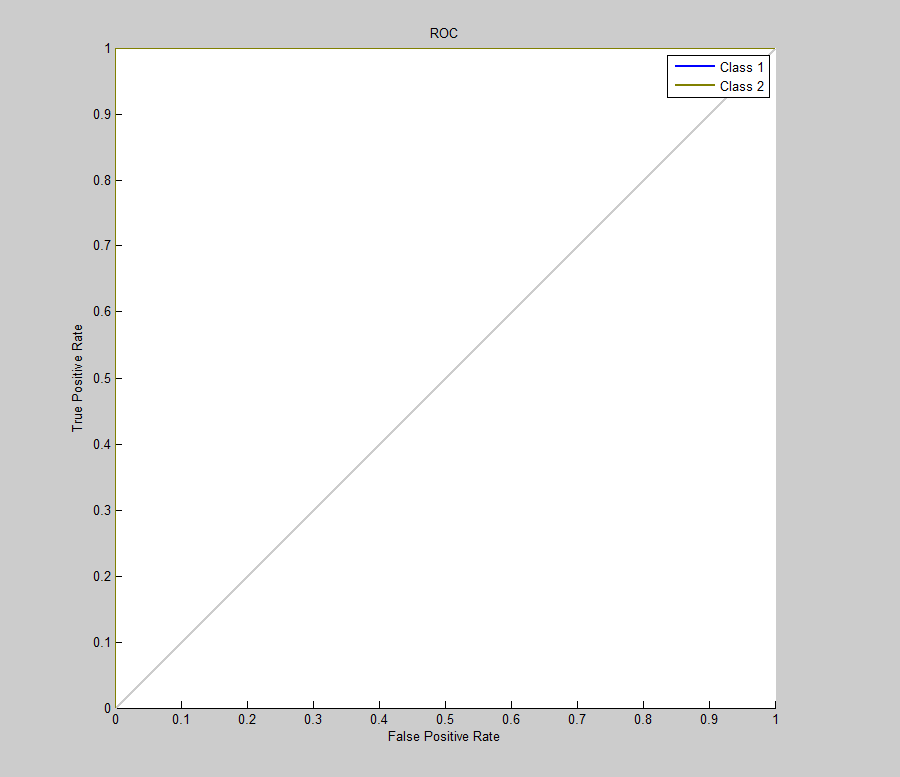
* 1. **TCP PSH ACK Attack cloud computing**
     1. **PCA**



* + 1. **LDA**



* + 1. **Proposed (PCA+LDA+ALO)**



**Result phase**

The incoming traffic profile (packets per second) of the CAIDA dataset has one-way traffic profile clearly shows two rather distinct traffic rates, as experienced by the target victim. The first one is the packet rate during the first-half of the trace i.e. from the start of the trace until around the 1500second mark. For simulation purposes, the packet rate during this period is referred to as the normal packet rate. The average packet rate for this period was approximately 385 packets per second in the original dataset. The second distinct traffic rate is seen during the second half of the trace i.e., starting from approximately 1500 seconds until the end of the trace. During this period, the packet rate is referred to as the attack packet rate, and contains approximately125,705 packets per second on average. In order to simulate network traffic similar to the CAIDA dataset, the normal and the attack packet rates, calculated from the original dataset, are provided to the attack configuration file of the MATLAB to generate a 5 minute sample network trace.

IDPS DDoS attack preventing model

**Evaluation and comparison phase**