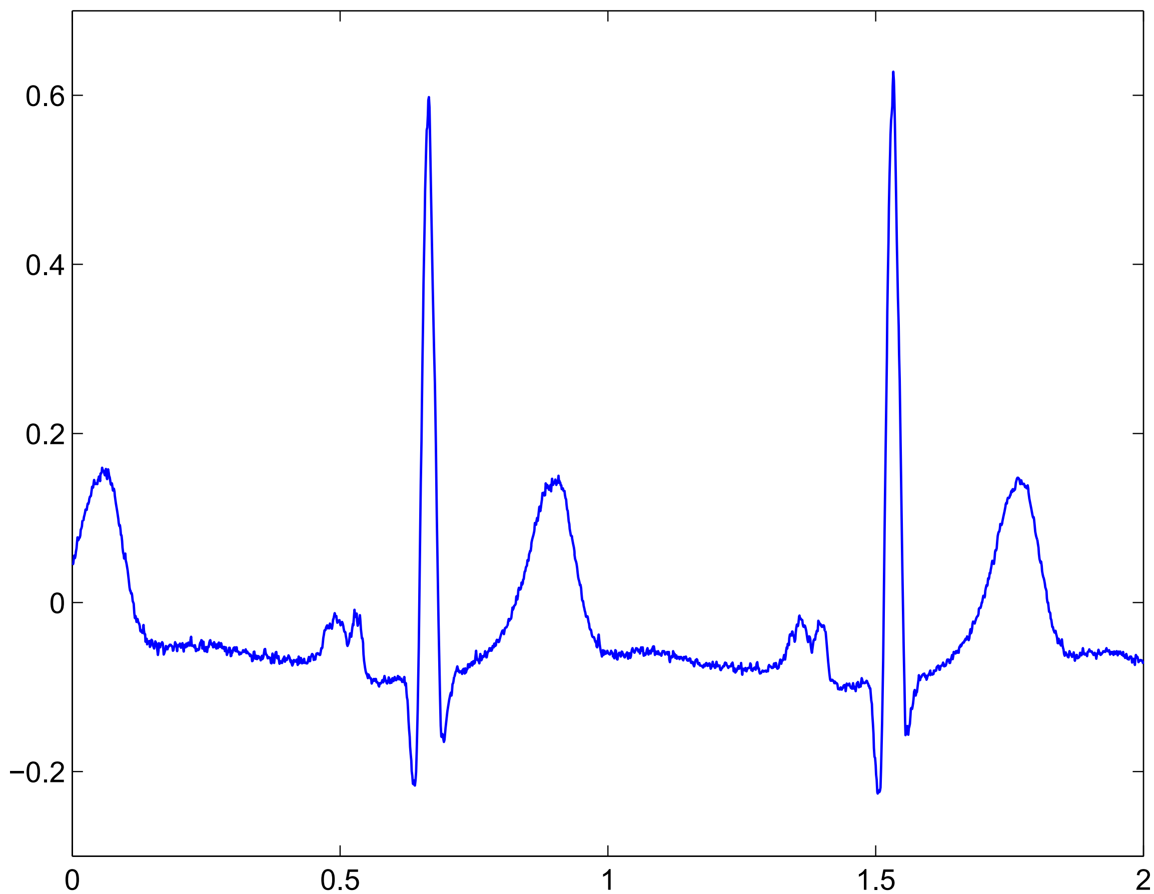
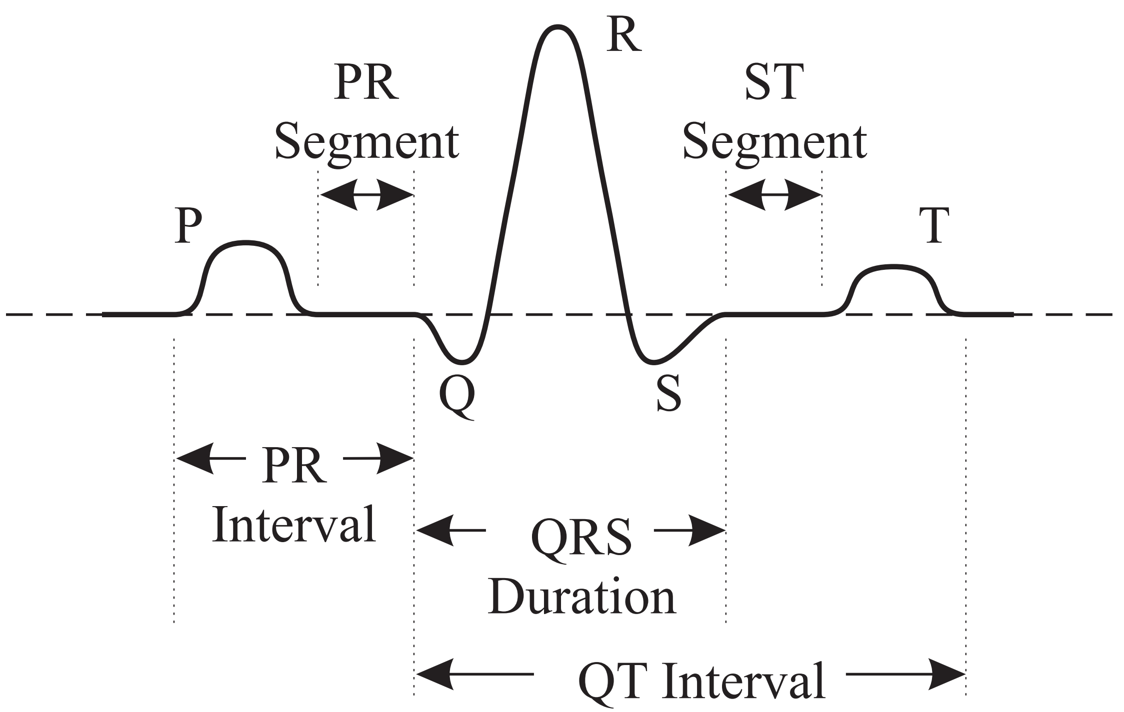
Pattern Recognition
Case Study:
EKG Analysis

EKG Signal



Raw Data set: For each patient, 15 leads measured at a rate of 1000 Hz for several seconds.

Feature Extraction



Total Features (1)

- age in years
- gender, -1=female, 1=male
- maximum heart rate in beats/min
- minimum heart rate in beats/min
- average time between heart beats in sec
- rms deviation of the mean heart rate in beats/sec
- full width at half maximum for the heart rate distribution
- average qt interval for lead with max t wave
- average qt interval for all leads
- average corrected qt interval for lead with max t wave
- average corrected qt interval for all leads
- average qrs interval for all leads
- average pr interval for lead with maximum p wave
- rms deviation of pr intervals from average-max p lead
- average pr interval for all leads

Total Features (2)

- rms deviation for pr interval from average-all leads
- percentage of negative p waves-max p lead
- average percentage of negative p waves for all leads
- maximum amplitude of any t wave
- rms deviation of qt intervals
- rms deviation of corrected qt intervals
- average st segment length
- rms deviation of st segment lengths
- average heart rate in beats/min
- rms deviation of heart rate distribution in beats/min
- average rt angle averaged over all amplitude beats
- number of missed r waves (beats)
- % total qt intervals not analyzed or missing
- % total pr intervals not analyzed or missing
- % total st intervals not analyzed or missing
- average number of maxima between t wave end and q

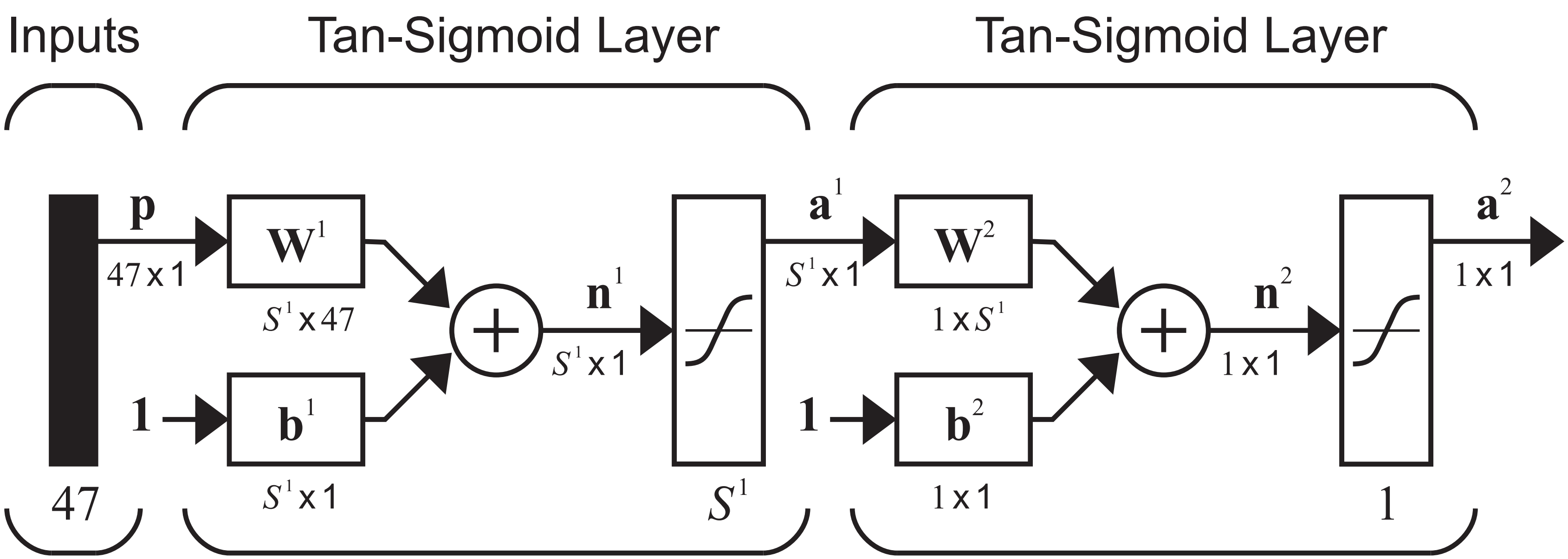
Total Features (3)

- rms deviation of rt angle for all beats
- ave qrs from amplitude lead
- rms deviation of qrs from amplitude lead
- ave st segment from amplitude lead
- rms deviation of st segment from amplitude lead
- ave qt interval from amplitude lead
- rms deviation of qt interval from amplitude lead
- ave bazetts corrected qt interval from amplitude lead
- rms deviation of corrected qt interval from amplitude lead
- ave r-r interval from amplitude lead
- rms deviation of r-r interval from amplitude lead
- average area under qrs complexes
- average area under s-t wave end
- average ratio of qrs area to s-t wave area
- rms deviation of rt angle within each beat averaged over all beats in amplitude signal
- st elevation at the start of the st interval for amplitude signal

Data Set

- The data set contains 447 records. Each record has 47 input variables, and one target value. The target is 1 for a healthy diagnosis and -1 for an MI diagnosis.
- There are only 79 records for the healthy diagnosis, while there are 368 records for the MI diagnosis.
- Repeat the healthy records in the data set, so that the total number of healthy records is equal to the number of MI records. (Could also use weighted least squares.)
- Randomly set aside 15% of the data for validation and 15% for testing. The remaining 70% is used for testing.
- The input data were scaled to the range [-1,1].
- The targets were set to values of -0.76 and +0.76.

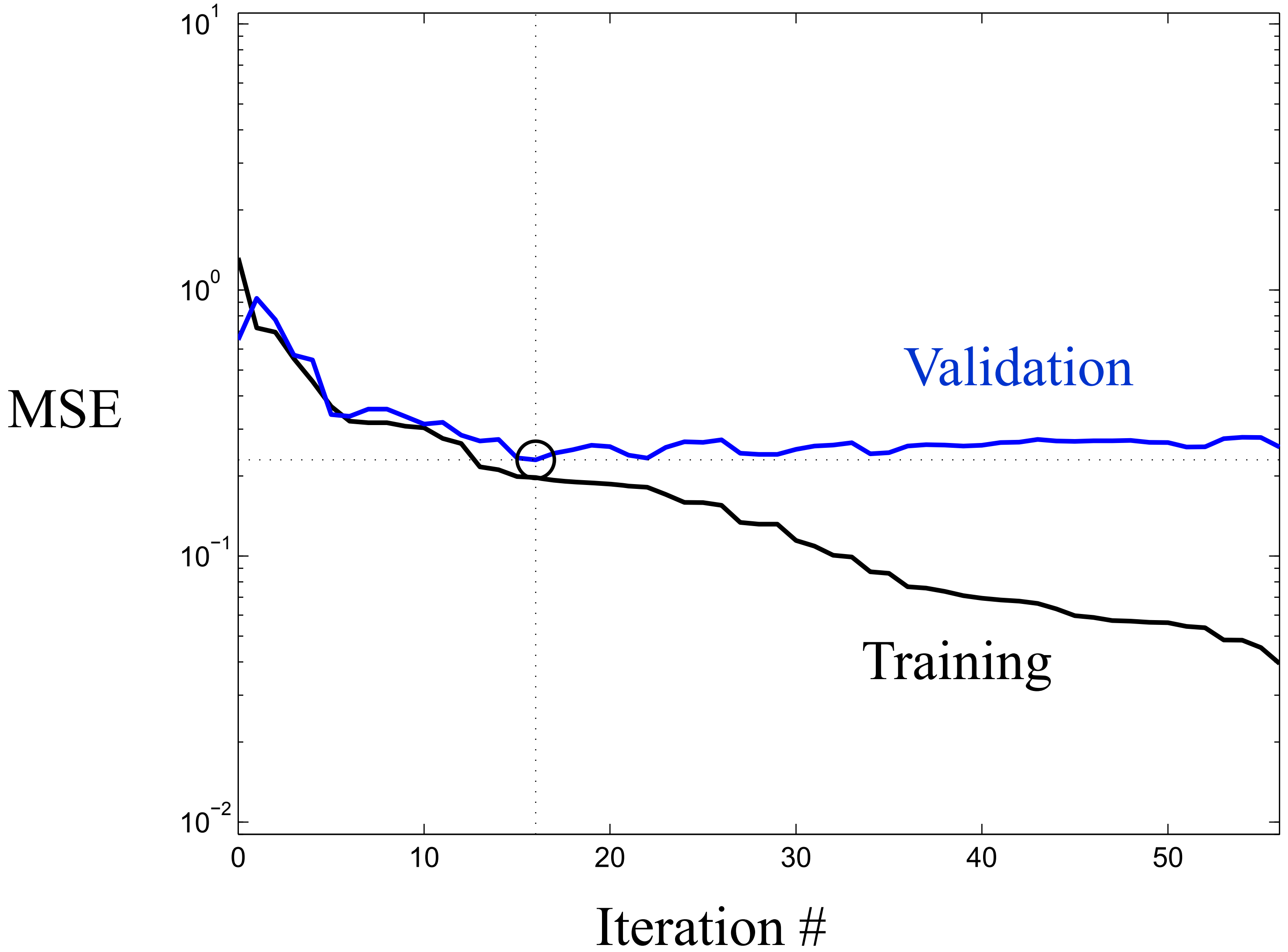
Network Architecture



$$\mathbf{a}^1 = \mathbf{tansig}(\mathbf{W}^1\mathbf{p} + \mathbf{b}^1)$$

$$\mathbf{a}^2 = \mathbf{tansig}(\mathbf{W}^2\mathbf{a}^1 + \mathbf{b}^2)$$

Training Performance (S^1=10)



Confusion Matrix

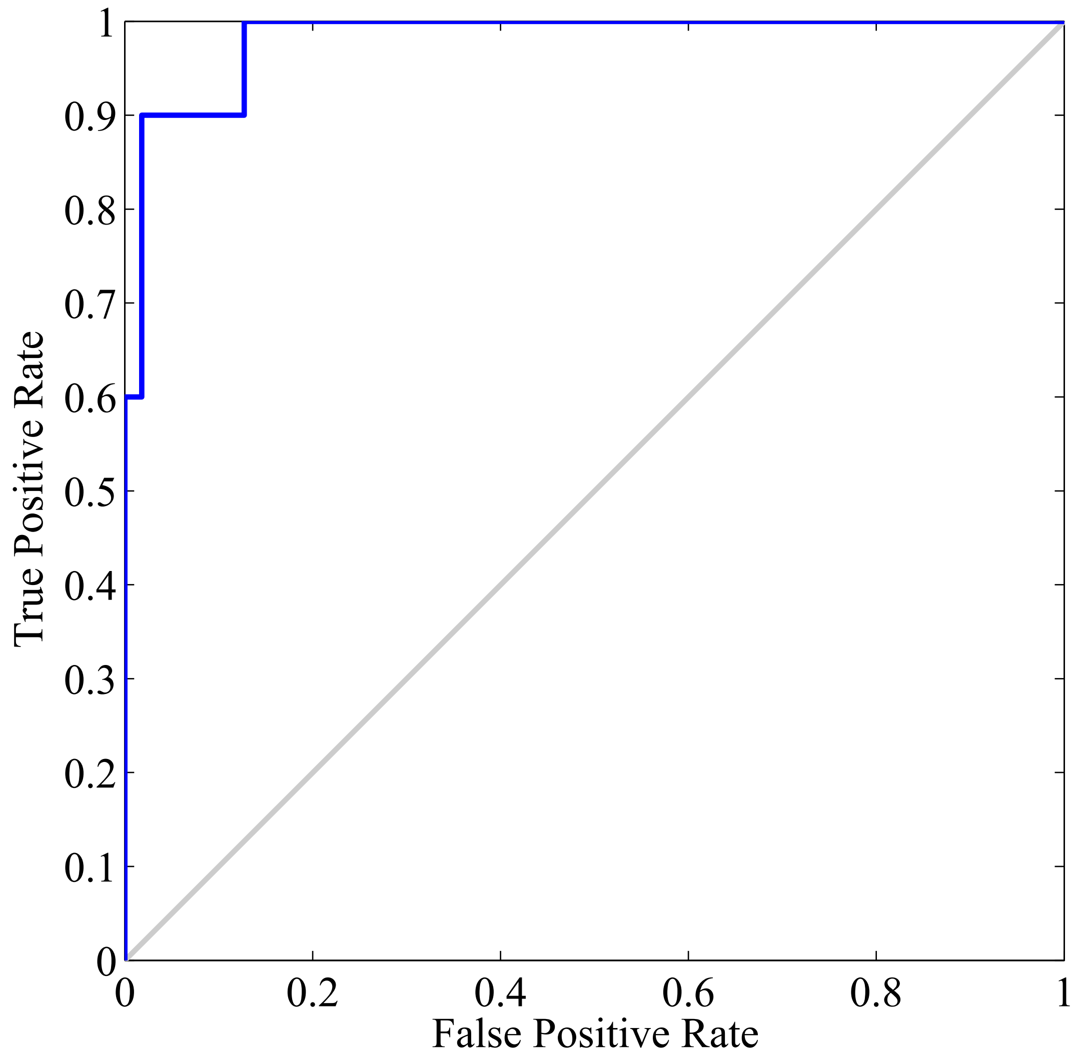
Confusion Matrix

Output Class \ Target Class	1	2	
1	**13** 15.3%	**5** 5.9%	72.2% 27.8%
2	**1** 1.2%	**66** 77.6%	98.5% 1.5%
	92.9% 7.1%	93.0% 7.0%	**92.9%** **7.1%**

Target Class

25

ROC Curve



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
True Positive Rate
0
0.2
0.4
0.6
0.8
1
False Positive Rate

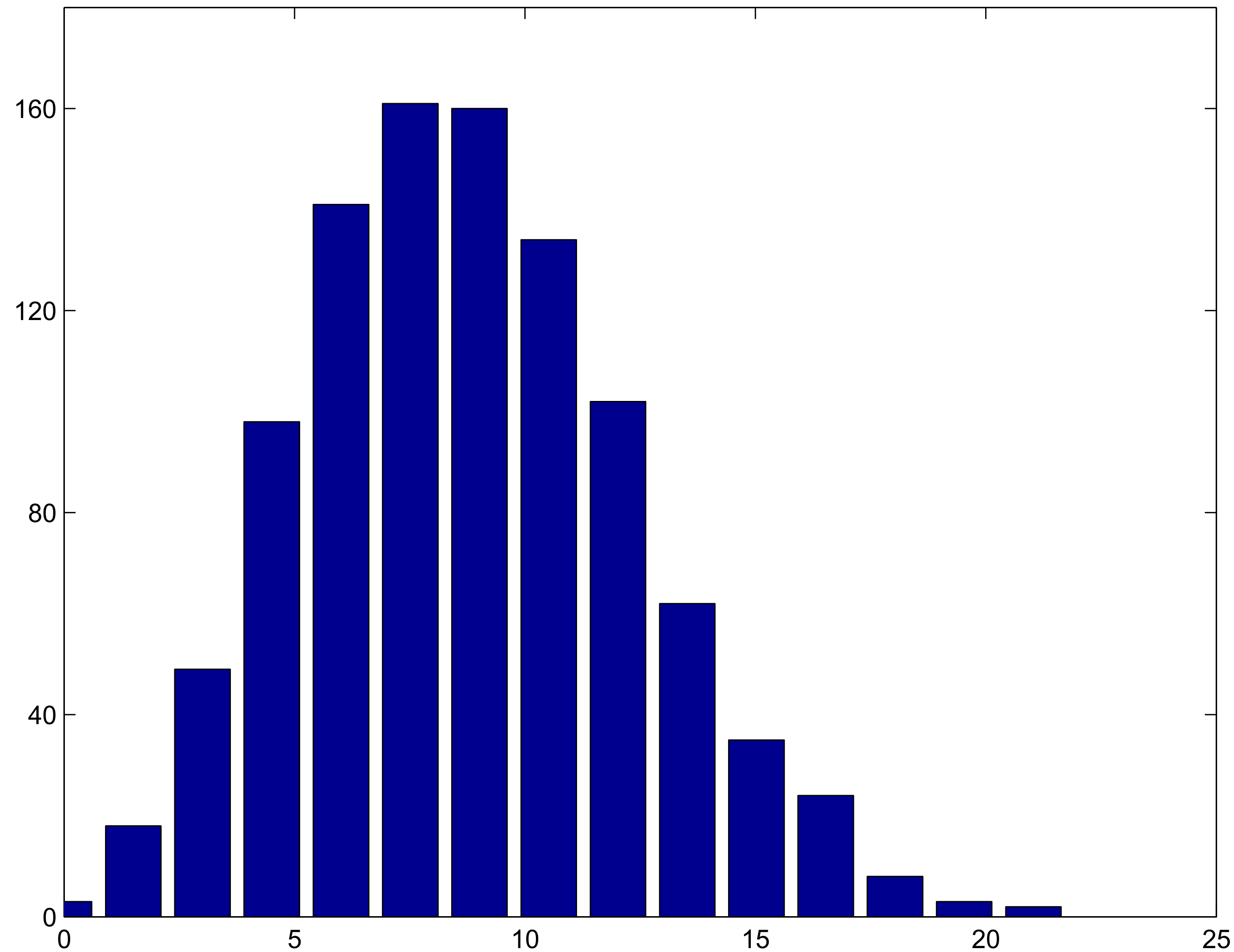
25

Average Confusion Matrix (1,000 Trials)

Confusion Matrix

Output Class \ Target Class	1	2	
1	**9.11** 14.0%	**3.69** 5.7%	71.2% 28.8%
2	**2.51** 3.9%	**49.83** 76.5%	95.2% 4.8%
	78.4% 21.6%	93.1% 6.9%	**90.5%** **9.5%**

25 Percent Error Histogram (1,000 Trials)



Percent Error

25

Uses of Committees of Networks

- The Monte Carlo process can help validate the data and the training process.
- Identify which patients are misclassified in each Monte Carlo run. Those patients can be carefully investigated.
- These cases can be helpful in two areas. First, they may identify mislabeled data.
- Second, if the data point was correctly labeled, these points may help identify new features which can capture the key characteristics of the EKG. It may also indicate that we need more data.
- The committee of network outputs can be combined to produce improved classification. A majority vote of the committee can produce a single classification.