SEISMIC DISCRIMINATION USING NEURAL NETWORKS:
A STUDY OF SHALLOW EVENTS NEAR KNOWN BLASTING SITES

by

Sean F. Walden

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

University of Washington

1992

Approved by ________________________________
(Chairperson of Supervisory Committee)

Program Authorized to Offer Degree ________________________________

Date ________________________________
SEISMIC DISCRIMINATION USING NEURAL NETWORKS:
A STUDY OF SHALLOW EVENTS NEAR KNOWN BLASTING SITES

by

Sean F. Walden

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

University of Washington

1992

Approved by ____________________________

(Chairperson of Supervisory Committee)

Program Authorized
to Offer Degree ____________________________

Date ____________________________
Master's Thesis

In presenting this thesis in partial fulfillment for a Master's degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this thesis is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U. S. Copyright Law. Any other reproduction for any purpose or by any means shall not be allowed without my written permission.

Signature ____________________________

Date ____________________________

University of Washington

Abstract

SEISMIC EVENT DISCRIMINATION USING NEURAL NETWORKS: A STUDY OF SHALLOW EVENTS NEAR KNOWN BLASTING SITES
by Sean F. Walden
Chairperson of the Supervisory Committee: Professor Stephen D. Malone
Geophysics Program

Seismic event discrimination between natural earthquakes and explosions has always been a difficult problem. Seismic signals are complicated, and several variables influence signal propagation. With new pattern recognition tools, classifying events based on signal characteristics can be easily automated. In this study a neural network is used to classify 88 events in Northeastern Washington. Of these 88 events, 44 are earthquakes and 44 are explosions. The events have depths of 0-10 km and coda duration magnitudes of $M_c=1.5-2.7$. The signals from three seismic stations are analyzed for spectral content in both the P and S waves. For all three stations the neural network converges on a solution for the spectral pattern recognition problem. Testing the classification scheme of the neural network by the leave-one-out method gave accuracy rates from 75% to 87%. When the network was trained with half the available dataset then tested with the untrained events, accuracy rates ranged from 75% to 90%. Varying the number of hidden units showed negligible difference in classification accuracy. Reasons for the success of the classification scheme are not clear but I argue there are two possibilities based on observations of spectral content. First, the explosions have dominant low frequency spectral energy in relation to the earthquakes. Second, the explosions signals are more monochromatic when compared to the earthquakes. A thorough sensitivity analysis should detect the spectral characteristics responsible for the discrimination scheme's success.
# TABLE OF CONTENTS

List of Figures iii
List of Diagrams iv
List of Tables v

Chapter 1: Introduction 1
  Seismic Discrimination in Washington State 3

Chapter 2: Data 5
  Problems Choosing Data 10
  Using Spectral Analysis 12
  Spectral Analysis of Data 13

Chapter 3: Neural Networks 23
  Using Neural Networks as A Pattern Recognition Tool 23
  The Back-propagation Method 23
  Neural Network Configuration 29

Chapter 4: Results 32
  Accuracy of Classification 32
  Accuracy of Cross-Classification 35
  Optimizing Network Performance 36

Chapter 5: Discussion 38
  Comparing Performance with Linear Network 38
  Comparing Inputs 38
  Aiding Solution Convergence 39
  Understanding Event Classification 40
  Sensitivity Analysis 44

Understanding Misclassified Events 46
Chapter 6: Conclusions 50
List of References 51
Appendix A: Derivation of the Delta Rule 54
Appendix B: Derivation of the Sensitivity Gradient 56
LIST OF FIGURES

Number                                                                                     Page
2a. Epicentral Location of Earthquakes Detected at Station ETT                         6
2b. Epicentral Location of Explosions Detected at Station ETT                           6
3a. Epicentral Location of Earthquakes Detected at Station WAT                         7
3b. Epicentral Location of Explosions Detected at Station WAT                           7
4a. Epicentral Location of Earthquakes Detected at Station WEN                          8
4b. Epicentral Location of Explosions Detected at Station WEN                           8
5. Geographic Distribution of Washington Regional Seismic Network                     11
6. Power Spectrum of Single Earthquake at Stations ETT,WAT,WEN                          16
7. Power Spectrum of Single Explosion at Stations ETT,WAT,WEN                           17
8. Stacked Power Spectra for Events Detected at Station ETT                            18
9. Stacked Power Spectra for Events Detected at Station WAT                             19
10. Stacked Power Spectra for Events Detected at Station WEN                             20
11. Comparison of Explosion Spectra at Different Blasting Sites                         21
12. Comparison of Earthquake Spectrum with Explosion Spectrum                          22
13a. Sigmoidal Transfer Function                                                        25
13b. Derivative of Sigmoidal Transfer Function                                          25
14. Neural Network Configuration                                                         30
15. Event Classification After Learning Process                                         33
16. Event Classification After Leave-One-Out Method                                     34
17. Event Magnitude Distribution                                                        42
18. Event Depth Distribution                                                            43
19. Distribution of Misclassified Events                                                47
20. Power Spectra of Misclassified Events Recorded at WAT                               48

LIST OF DIAGRAMS

Number                                                                                     Page
1. Mapping Phase for Single Neuron                                                        26
2. Mapping Phase for Single Output Unit                                                   26
3. Complete Mapping Process by Neural Network                                             27
4. Back-propagation Phase for Output Unit k                                               27
5. Illustration of Error Minimization with Respect to Weight w                            28
6. Representation of Input Vector Presented to the Neural Network                        29
7. Example of Classifying Data Vector After Converging on Solution                        31
The support I received while hectically finishing this thesis was overwhelming. Steve Malone offered several helpful suggestions over the course of my research. My committee members, Steve Malone, Ken Creager, Alan Rohay, and Jenq-Neng Hwang, gave useful critiques on the thesis and provided many ideas. Special thanks to Alan Rohay who introduced me to seismic discrimination methods and arranged funding for me during a NORCUS summer research program at Battelle. Thanks to Stewart Smith for his review of the thesis and Jenq-Neng Hwang, professor in electrical engineering, for his help with neural network theory. The MITRE Corporation provided the neural network software. Financial support was provided by the Westinghouse Hanford Company under contract numbers PMM-RJU-505 and MLR-SVV-666685. I would like to thank fellow graduate student, Shawn Dewberry for his continual encouragement and assistance.

Lesenia Arnett provided much needed comic relief and some damn good banana bread as well. Tony Beasley, Jeff Musiak, and Chris Walter gave useful reviews of the thesis in its early stages and did their best to keep my priorities straight by taking me out for beer. Officemate Seth Moran provided helpful tips on formatting the thesis and kept me physically fit with office juggling sessions. Thanks also to Margaret Rozendaal for morning coffee and putting up with the juggling. I would like to thank my family for their complete support of my work. Most of all I would like to thank Keiko Maeda who offered emotional support and kept me alive by bringing in lunches and dinners while I typed feverishly at the computer. Her compassion and generosity enabled me to finish this thesis.
Chapter 1
INTRODUCTION

History

With the introduction of the limited Nuclear Test Ban Treaty in 1965, considerable research efforts have been made in the field of seismic source identification. To ascertain whether countries were acting in accordance with the treaty, methods were developed to discriminate nuclear explosions from earthquakes. These methods were based on observed differences in seismic signals due to explosions versus seismic signals due to earthquakes. These observed differences were summarized by Ives (1976): 1) earthquakes produce approximately equal amounts of compressional wave (P) and shear wave (S) energy, while explosions produce more P wave energy, 2) earthquakes give positive and negative first motions (double-couple source), and explosions give only positive first motions, 3) earthquakes are located deeper than explosions, and 4) wave train duration is longer in earthquakes than explosions. Many ideas have been proposed to quantitatively measure the above signal qualities, however, no single method offers a confident global discrimination scheme. As Ives mentions, discrimination based on these differences is not always available, especially for small magnitude events. While explosion signals do exhibit more P wave energy than S wave energy, many earthquake signals fall under this category as well. Also, the discrepancy between P and S wave energy is difficult to analyze for events of small magnitude due to the difficulty in identifying particular phases. Determining polarity of first motion is often obscured by noise in the signal. Concerning source location; events confidently located at a depth of 5 to 10 km can be identified as an earthquake (Douglas, 1981). However, depth estimates of shallow focus events have a large uncertainty (Pomeroy et. al., 1982). Many researchers have suggested using a combination of discrimination methods and this prompted a comprehensive review of potential discriminators by Pomeroy et. al. (1982). Some of the indicators Pomeroy suggests using are: 1) P wave polarity, 2) event depth, 3) amplitude ratios between different phases, 4) third moment of frequency, 5) spectral ratios in several phases and 6) frequencies of spectral peaks. Each of these discriminators have been used with some success, however, most tested only small data sets from single geographic regions.

Recently, research efforts in discrimination have involved pattern recognition; results of many studies have found high success rates. In studying several seismic datasets, Chen (1978), classified each set with an accuracy of 86-89% correct identification. Chen also tested one dataset using older discrimination methods (spectral ratios and third moment of frequency) and achieved only 72% correct classification. The pattern recognition techniques improved the accuracy by 15%. Other studies have found similar classification successes using pattern recognition routines. In an analysis of 66 events recorded by the NORESS array in Norway, Pulli (1990) used a neural network as a pattern recognition routine. After training the network with 33 of the events, he was able to correctly classify 80% of the 33 untrained events. A similar study of 170 events in North America by Dowla et. al. (1990), achieved a correct recognition rate of 93%. While the above classification tests were done in different geographic areas, all of the events were of magnitude \( m_b > 4.0 \). There has been some success in discriminating large explosions from earthquakes, as in the studies mentioned above, but the problem becomes much more difficult with smaller events, and "hidden" explosions. Hidden explosions may include events masked in underground salt domes, or an explosion initiated during an earthquake (Douglas, 1981). Tests still remain to understand how well these methods work for smaller events. The effectiveness of such methods on "small" (coda duration magnitude \( M_s = 1.5-2.7 \)) events will be examined in this study.
Seismic Discrimination In Washington State

To fully understand the tectonics of any region it is important to know which seismic events are natural earthquakes. Washington state is a region of considerable shallow tectonic activity mingled with explosions from active mining and construction operations. Over the last twelve years there has been a considerable amount of blasting in Washington State which has been recorded by the Washington Regional Seismic Network (WRSN). In general, the shallow earthquakes are either located near the Cascade mountain range, or near clustered regions in the Columbia Basin. However, their overall epicentral distribution is widespread. The majority of explosions are located in clusters around blasting areas, but often times the epicentral location of the explosions overlap the locations of naturally occurring earthquakes. A look at earthquake activity versus explosion activity can be seen in Figure 1. These plots represent shallow events (depth = 0-10 km) of coda duration magnitude \( M_c > 1.0 \) from 1980-1992. The overlapping of these two seismic event types can misguide tectonic interpretation of the Pacific Northwest. In order to separate these events we need to find distinguishing characteristics between their associated signals. Many techniques have been proposed, and some involve time consuming or statistically weak methods. In the past, certain seismic events were suspected of being blasts due to their location, time of occurrence and dominance of low frequency in the wavetrain. The suspected blasts could then be confirmed by corresponding with local mining operations to find out precisely when blasting had occurred. The effectiveness of this method relies on the accuracy and the experience of the processing analyst. A more desirable technique of discrimination would involve automation of robust quantitative methods. Neural networks can provide the means for accurate and quick decision making. In this study I investigate the potential of using neural networks for discriminating explosions from nearby shallow earthquakes.
Chapter 2
DATA

This study focused on a region of shallow seismicity. Deep events which are confidently located are obviously naturally occurring earthquakes, so the real problems with discrimination are found at shallow depths. An area near Lake Chelan in Northeastern Washington was chosen because it is near known blasting locations which also have shallow earthquakes. The majority of blasts were recorded from two projects. One group of explosions is the result of a state highway project from Chelan to Hugo. The second group of explosions is related to a project at Cannon Mine. A distribution of the blasts and earthquakes can be found in Figures 2-4. Information on each event can be found in Table 6. It should be noted that the events originally cataloged as blasts could have been misclassified. While the processing seismic analyst tries to confirm suspected blasts with the project officials, blasts are sometimes cataloged based on strong blast-like characteristics. The characteristics may include calculated event depth, event location, time of occurrence (most blasts occur during the daytime), and a monochromatic nature of the recorded signal. If a blast was initiated during the nighttime hours, the analyst may have assumed the event to be an earthquake and cataloged the event incorrectly. If an earthquake located very shallow it may have been cataloged as a blast. This potential for error in the catalog should be recognized when judging the accuracy of neural network classification based on agreement with previous analyst classifications.

Observing the distribution of events, there are two distinct clusters of blasting locations which are separated by a looser cluster of earthquakes. With this distribution I hope to avoid discriminating differences in the seismic signals due to path effects from source to receiver. Also, any common characteristics in local source geology at the two blasting sites would be expected in the earthquake dataset as well, so this would not be a factor in separating the blasts from the earthquakes. The earthquake cluster separates the
two regions, so the gross distinctions in seismic signals should be primarily based on source mechanism and source depth.

The data consist of 44 known explosions and 44 earthquakes. The calculated magnitudes of the events range from $M_e = 1.5-2.8$ and their calculated depths range from 0-10 km. Most of the selected events had previous picks for the P (compressional) phase. Only events with strong P arrivals were used. Detectable S (shear) phases were not a consideration in choosing the dataset because predicted S (shear wave) arrivals were used to analyze the waveforms. The primary goal in choosing the events was to use data recorded at a single seismic station to distinguish earthquake signals from explosion signals. Another motivating factor in selecting data was based on the coverage by the seismic network. Ideally, one event would have data recorded at several stations which could then be classified as an earthquake or a blast by each station. The classification by a single station could then be cross referenced with classifications by other stations recording the event. The number of stations used in this study was limited to three stations.

Due to the sparse number of seismic stations in the studied region, data for the current study were originally limited to using eight stations in the WRSN. However, poor signal to noise ratios restricted the number of potential stations used in the area. There were eight potential stations to be used but the lack of several well recorded events at five stations discouraged their use in the discrimination analysis. This left only three possible stations to classify a single event (ETT, WAT, WEN). Not all selected events were recorded at each of these three stations. In some cases only one station was available due to poor signal quality at the other stations. Whenever possible, I tried to use at least two of the three stations, allowing for cross referencing of event discrimination. The number of events used at each station are given in Table 1.

### Table 1: Number of Events Used at Seismic Stations

<table>
<thead>
<tr>
<th>Station</th>
<th>Station Latitude (°, ″)</th>
<th>Station Longitude (°, ″)</th>
<th>Number of Earthquakes</th>
<th>Number of Known Explosions</th>
<th>Total Number of Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETT</td>
<td>47,39,18</td>
<td>120,17,36</td>
<td>39</td>
<td>23</td>
<td>62</td>
</tr>
<tr>
<td>WAT</td>
<td>47,41,55</td>
<td>119,57,15</td>
<td>28</td>
<td>27</td>
<td>55</td>
</tr>
<tr>
<td>WEN</td>
<td>47,31,46.2</td>
<td>120,11,39</td>
<td>29</td>
<td>25</td>
<td>54</td>
</tr>
</tbody>
</table>

All data used in this research were provided by the Washington Regional Seismic Network (WRSN) operated by the Geophysics Program at the University of Washington (Station distribution can be seen in Figure 5). This network consists of approximately 120 short-period, vertical-component seismic stations. As of 1980, the signals from these stations are telemetered to the seismic laboratory at the University of Washington. A triggering algorithm is used to determine when an event has occurred and this activates the computer to digitally record data at 100 samples/second at each station. The triggering algorithm uses a short-time-average to long-time-average (STA/LTA) ratio. The recorded waveforms are then analyzed for phase arrival times. The arrival times are then used to determine event location and focal mechanism.

### Problems choosing data

Many of the blasts were not well recorded by the seismic network. Due to their limited size ($M_e = 1.5-2.8$) and the distribution of seismometers, there were usually only a few waveforms that could be used for each event in developing a discrimination scheme. This constrained the possible number of stations to be used in analyzing the waveform.

Also station geology often has a dominating effect on the character of the signal. These recording site characteristics are clearly seen in the power spectrum of the detected
signal. However, if single stations are used as discriminators the local station response will affect the blasts and earthquakes similarly. Therefore, expected differences in the waveforms should be primarily due to the source.

Using Spectral Analysis

A detected seismic signal is assumed to be a time domain convolution of the seismic source effects $S(t)$, the path effect $P(t)$, the receiving site effects $R(t)$, and the instrument response $I(t)$.

$$ V_i(t) = S(t) \times P(t) \times R(t) \times I(t) $$

In the frequency domain, this expression becomes:

$$ V_i(\omega) = S(\omega) P_i(\omega) R(\omega) I_i(\omega) $$

In this study only single stations are used for discrimination. Therefore, the site effects and instrument response are the same for all events. Thus, they will not be a factor for differences found in recorded seismic signals. The differences in path effects should be small because the events and stations are tightly located in the same geological region. However, with varying source to receiver distances, signals can experience differences due to frequency dependent attenuation. I expect this factor will be small because the group of events are clustered in a small region south of Lake Chelan. This minimization of path effects, site effects and instrument response, leaves source effects as the primary distinguishing characteristic for the detected seismic signal.

The idea for using spectral analysis involves source properties in the blasts. Due to differences in source characteristics, certain frequencies are expected to be enhanced in explosions that are not necessarily prevalent in earthquakes. Primary source distinctions
are the event depth and the rupture mechanism in explosions. One identifying feature of the source in some blasting procedures is the ripple-firing technique. Ripple-firing produces several dominant frequencies in the spectra at integral periods of the blasting interval (Baumgardt, 1988; Smith, 1989; Day, 1991). While this characteristic is useful for "ripple-fired" explosions there are many single explosions which do not share this feature. In order to distinguish both ripple-fired and single explosions from naturally occurring earthquakes a discrimination routine is needed to generalize common features shared by these two blast types.

Another technique for analyzing source distinctions involves spectral shapes for different phases in the seismic signal. While the shape may depend on the event magnitude and regional geology, the discriminant has given promising results for events in the western United States (Bennett, 1986; Taylor et al., 1988). Analyzing Nevada Test Site explosions and surrounding earthquakes of magnitude \( m_{0} \geq 3.0-4.5 \), Taylor achieved misclassification rates from 4% to 33% using single stations as discriminators. The spectral ratio used by Taylor was \((1-2 \text{ Hz})/(6-8 \text{ Hz})\) for the seismic phases \( P_{n}, P_{g}, \) and \( L_{g} \). The success of spectral ratios in the western U.S. was the motivation for using spectral shapes to quantify characteristics in the recorded seismic signals.

Spectral Analysis of Data

I analyzed the spectral content of the P and the S phases. These are the only phases consistently observed from the small events used in this study. Many of the blasts did not have distinct S waves. This feature is often used to aid in discrimination methods. Therefore this information is useful to send to the neural network. In order to isolate common similarities and differences in the source, I used a 2 second window of the P arrival and a 4 second window of the predicted S arrival. Some events were close to stations, so in order to isolate the P coda I was limited to using only the first 2 seconds. The predicted S arrival

was estimated using a relative velocity ratio of \( V_{p}/V_{s} = 1.73 \). This usually gave an estimated S wave arrival that was slightly before the observed S arrival, but no earlier than 0.5 seconds.

Below is the procedure used for analyzing spectra:

1) Each trace was reviewed to determine the accuracy of any previous first arrival pick. The first 2 seconds of the P phase and 4 seconds of the S phase were then windowed.

2) The data were filtered with a 2 pole butterworth filter with corner frequencies of 0.5 and 25.0 Hz.

3) After filtering, the data was tapered with a 5% Hamming window and an auto-correlation was performed.

4) The P phase auto-correlation was zero padded to 512 points and the S phase auto-correlation was zero padded to 1024 points.

5) The FFT was then computed and divided by the number of data points used to give the power spectra of the P and S phases (Weiner-Khintchine Theorem; Karl, 1989)

6) Spectra were normalized by dividing the mean value from 1-11Hz with an arbitrary value of 100. The normalization factor was calculated as norm = (spectral mean of 1 to 11Hz)/100.

7) The spectra was then scaled by log10, and the average values in 2 Hz bins were fed into the neural network. There were five 2 Hz bins from 1 Hz to 11 Hz for both the P and S spectra, giving a total of 10 inputs to the neural network for each station-event pair. The spectral average is taken for 2 Hz bins to smooth the spectral estimate. This way we reduce the effect of strong anomalous peaks and get a more general representation of the power spectra.
Example spectral plots are displayed in Figures 6 & 7. Figure 6 shows the P and S spectra for all three stations. Notice the similarity between stations ETT and WAT, while the majority of energy at WEN is at much lower frequency (~2 Hz). Stations ETT and WAT do have spectral peaks at about 2 Hz but they are not nearly as monochromatic. I expect this low frequency energy is related to the site effects of the recording station WEN. The peak in energy at 3-4 Hz for stations ETT and WAT is a source effect, but may be masked at WEN due to the enhanced peak at 2 Hz from the site effects. For the earthquakes there is an obvious shift in spectral energy toward higher frequencies at all stations (Figure 7).

To gain some insight on the gross spectral shape, I have plotted a sum of the P and S spectra for earthquakes and explosions for all three stations (Figures 8, 9, 10). The characteristic peaks due to ripple-firing can be seen in both the P and S spectra at stations ETT and WAT (Figures 8(c,d) and 9(c,d)). These are the peaks at approximately 2 Hz and 4 Hz in both the P and the S spectra. Observing these "average" spectral plots it appears there is a definite trend for the explosions to have more low frequency energy. However, this plot can be misleading. When comparing the spectra from both blasting sites, there is clear disparity in the spectral peaks (Figure 11). This plot illustrates the distinctions between two blast sources recorded at station WAT. With so much variation in spectral energy among explosions the possibility of finding a single spectral ratio to discriminate these two explosion types from earthquakes is remote. This disparity in location of the spectral peak for the explosions appears discouraging. However, the explosions are observed to be more monochromatic (composed of a single frequency) than the earthquakes (Figure 12). This characteristic provides valuable information to a neural network.
Figure 7. Comparison of an earthquake detected at stations ETT, WAT, and WEN.
(a) P spectrum at ETT  (b) S spectrum at ETT  (c) P spectrum at WAT
(d) S spectrum at WAT  (e) P spectrum at WEN  (f) S spectrum at WEN.

Figure 8. These plots represent stacked spectral sums or averaged spectra for events detected at station ETT.  
(a) Average P wave spectrum for earthquakes. 
(b) Average S wave spectrum for earthquakes.  
(c) Average P wave spectrum for explosions.  
(d) Average S wave spectrum for explosions.
Figure 9. These plots represent stacked spectral sums or averaged spectra for events detected at station WAT. (a) Average P wave spectrum for earthquakes. (b) Average S wave spectrum for earthquakes. (c) Average P wave spectrum for explosions. (d) Average S wave spectrum for explosions.

Figure 10. These plots represent stacked spectral sums or averaged spectra for events detected at station WEN. (a) Average P wave spectrum for earthquakes. (b) Average S wave spectrum for earthquakes. (c) Average P wave spectrum for explosions. (d) Average S wave spectrum for explosions.
Figure 11. Comparison of earthquake spectra and explosion spectra for events detected at station WAT. (a) P spectrum of an earthquake  (b) S spectrum of an earthquake  
(c) P spectrum of an explosion  (d) S spectrum of an explosion.

Figure 12. Comparison of explosion spectra from the two blasting sites. The signals were recorded at station WAT. (a) P spectrum of road blasting project  (b) S spectrum of road blasting project  (c) P spectrum of mining blast  (d) S spectrum of mining blast.
Using neural networks as a pattern recognition tool

Neural networks have proven to be effective tools in pattern recognition. The motivation is to make several comparisons of the data that could not feasibly be made by regular statistical techniques. The network is trained to develop weights using a set of well understood data. Then these weights can be used to categorize unresolved data. A problem with this technique is that the process is, in many ways, a black box, and the evolution of the weights is not always clear. Statistically, the confidence in the classification schemes are high and this is the primary justification for using them.

The back-propagation method

This is a training procedure that presents the network with an input vector, and a target vector. The input vector contains characteristics of the data, while the target vector determines the classification group of the data. Initially the network randomly weights components of the input vector, maps to an output vector, and compares with the given target vector. This is commonly called the “feed-forward phase”. If there is a difference between the output vector and the target vector, the weights are changed to minimize this difference. The new weights are then “back-propagated” to the beginning of the algorithm and the forward mapping continues. When there is no difference between the output value and the target value, no changes are made to the weights and the network has found a solution for that input vector. The process of adjusting the weights is governed by an algorithm called the delta rule, and this is the backbone of the back-propagation scheme. A more detailed description of the process is given below.

The following is a summary of previous studies (Rumelhart, 1988; Leighton, 1991; Maren, 1990; Lippman, 1987). Each component of the input vector is initially given some random weight, and the net sum is fed to the hidden unit $j$,

$$\text{sum}_j = \sum_i w_{ji} x_i$$

where $\text{sum}_j$ is the sum of the weighted input and $w_{ji}$ are the weights for the $i^{th}$ component of the input vector going to the $j^{th}$ hidden unit. The hidden unit $j$ then maps the sum with a nonlinear transfer function $f$ to produce the following activation value:

$$a_j = f(\text{sum}_j)$$

The typical transfer function used is the sigmoidal or logistic activation function,

$$a_j = \frac{1}{1 + e^{-\eta \text{sum}_j}}$$

which has a derivative,

$$\frac{\partial a_j}{\partial \text{sum}_j} = a_j(1 - a_j)$$

where $\eta$ is a constant factor called the learning rate in equation (3). Various theoretical curves of the sigmoidal function and its derivative can be found for various values of the learning factor $\eta$, as seen in Figure 13. Notice that the above derivative reaches a maximum for $a_j = 0.5$ and a minimum when $a_j = 0$ or 1. Because the weight change is proportional to the derivative, the weights will be changed most for output units whose derivative is near the central peak. This is thought to contribute to the stability of the system (Rumelhart, 1988).
An illustration of this first mapping to an activation value can be seen below:

Diagram 1. Mapping phase for single neuron (hidden unit)

After the each neuron in the hidden layer maps the weighted input to a different activation value, all the activation values are in turn weighted and then sent to each unit in the output layer. The new weighted activation sum is given by the following expression:

$$\sum_{j} w_{kj} a_j$$

Here $\sum_{j}$ represents the weighted sum of the activation values from all hidden units going to output unit $k$. The number $N$ is the number of hidden units in the hidden layer. The output unit then takes this sum and performs a nonlinear mapping of its own.

Diagram 2. Mapping phase for single output unit

The entire process so far can be illustrated in the following diagram.
adjustment of the weights in the back-propagation algorithm is equivalent to a minimization of the error between the output vector and the desired target vector. The error is defined as the least mean squared error and is given by the following expression:

\[ E = \frac{1}{2} \sum \frac{(t_m - o_m)^2}{m} \]

(7)

This is the error for the kth output unit in terms of the target value, \( t_m \) and the output value, \( o_m \). As is discussed in Appendix A, weights are adjusted based on the idea that change in a particular weight should be proportional to the contribution of that weight on the total error (E).

\[ \Delta w_{kj} = \frac{\partial E}{\partial w_{kj}} \]

(8)

The weight adjustment process can be illustrated hypothetically through a diagram given by Maren (1990).

Diagram 5. Illustration of error minimization with respect to weight w
The above diagram illustrates the possible problems with minimizing the error in the neural network. If the error is not near a local minimum, it may converge to the global minimum. Unfortunately there is a possibility that the error will converge on a local minimum. However, if the local minimum still provides an adequate set of solution weights for the pattern recognition problem, the fact that it is not a global minimum in the error is not a practical concern.

Neural Network Configuration

The network used in this study is multi-layered with a back-propagation learning algorithm. There are three layers in the network; the input layer, the hidden layer, and the output layer. The input layer has 10 units, the hidden layer has 20 units, and the output layer has two units. See Figure 14 for a display of the network configuration. The input layer consists of 10 numbers which represent 2 Hz wide frequency bins from 1-11 Hz for the P and S spectra. The spectral averages are then scaled by the logarithm to constrain the variation of the spectral values. Values an order of magnitude or more from the mean value can drastically affect the weighting scheme of the neural network and can prevent convergence toward a stable solution. A diagram of the inputs and the network is shown below.

![Diagram of inputs and network configuration](image)

**Diagram 6.** Representation of input vector presented to the Neural Network

**Figure 14.** This is a neural network with 3 layers. There is an input layer with 10 units, a hidden layer with 20 units, and an output layer with 2 units. Lines represent inputs being weighted, summed, and sent to a single neuron or hidden unit. The neuron then maps the weighted sum to two output values.
A weighted sum of these inputs is then fed into the hidden layer where they are mapped by a nonlinear transfer function. The output is compared to the desired target vector and the weights are adjusted. The process continues until the output and the target value agree to within some specified tolerance.

When the network learns to distinguish patterns in the training dataset, it has found a set of weights which map each input vector to its desired output within a given tolerance. Thus, the network has converged on a solution to the recognition problem. These solution weights can then be used to classify events not involved in the learning process. This process is illustrated below in Diagram 7.

![Diagram 7. Example of classifying data vector after converging on solution](image)

Chapter 4
RESULTS

Accuracy of Classification by Single Station

In order to establish some degree of confidence in the solution weights we need some way to test the variability of convergence. One way to do this is to use a method called the leave-one-out method (Lachenbruch and Mickey, 1968). In this process the training dataset is deleted by one event and learning occurs with one fewer data vector. After the network converges the deleted event is tested with the solution weights to see how it maps without being involved in the learning scheme. This method avoids forcing the network to map the event as an earthquake or blast during learning, allowing the event to freely map afterward with the group it shares the most similarities.

After mapping each event with solution weights derived by the leave-one-out method the events were classified as either an earthquake or blast. Earthquakes map to (1,0) and blasts map to (0,1). Events which were thought to be blasts and then mapped positively as an earthquake were considered incorrect classifications. Events which mapped exactly between the two groups were classified as unknown. For example, a mapping of (.55, .45) would be classified as unknown. The two previous output values are dependent so the value of the first output distinctly classifies the data vector. Therefore if output1 was higher than .55, or lower than .45, the event would be considered classified or misclassified, depending on its original target.

In a test of three stations (ETT, WEN, and WAT) by the leave-one-out method, the results were drastically different. ETT had a correct classification rate of 81%, while WAT and WEN had accuracies of 72% and 87% respectively. The main reason for the high performance at WEN is due to the relative ease in distinguishing a pattern for the data. Most of the blasts used for station WEN occurred in the same location, so their spectra were extremely similar. The results of testing by station can be seen in
Figure 15. Above are the results of event classification after training the neural network. There are three different datasets represented: (a) events recorded at ETT, (b) events recorded at WAT, and (c) events recorded at WEN. The earthquakes should map to an output 1 value of 1.0, and the explosions should map to an output 1 value of 0.0. As seen above the two events types are well separated with only a few ambiguous classifications between 0 and 1.

Figure 16. Above are the results of testing event classification with the leave-one-out method. After training the neural network with one less event, the solution weights were then used to classify the left out event. Each plot represents several tests, one for each event recorded at the particular station. For example, there were 62 events recorded at station ETT, so (a) represents 62 tests. Similarly, (b) represents 55 tests for events recorded at WAT, and (c) represents 54 tests of events recorded at WEN.
classificaiton is the categorization by the processing seismic analyst, so there is some potential for misclassification. To address this potential error in the previous categorizations, I also test the "self-consistency" of the neural network classification. The "self-consistency" test allows for classifications which agree in a majority rules sense, but disagree with the previous categorization, to be considered correct. Testing the 65 events by the "majority rules" scheme gave a correct classification rate of 70%. Testing by a "self-consistent" majority rules scheme had a 75% correct classification. Comparing these two rates suggests that there are not many misclassified events in the original catalog. The complete neural network event classification catalog can be seen in Table 6. The explosion classification is indicated by an X and the earthquake classification is indicated by a Q. Original classifications are indicated by the first two spaces in the first column. Explosions are indicated with the letters AX and the earthquakes are indicated by the letter A followed by a space.

Another interesting point is raised for events recorded at all three stations. While there are only 16 of these events, 15 agree with the "known" catalog in a "majority rules" sense. This is a correct classification of 94%. The high classification rate for these well-covered events suggests two things. First, these events were originally cataloged correctly. Second, the discrimination routine is more accurate when the event is well recorded. To assure that the event is well recorded there should be a high density seismic array near potential blasting areas. This would definitely improve the discrimination accuracy of the neural network.

**Optimizing Network Performance**

In the previous neural network there were 20 units in the hidden layer. There is some question as to whether this is the optimum number of hidden layers and the optimum number of hidden units. Researchers in the past have investigated the optimum
number of hidden layers and the optimum number of units in the hidden layer for a given
discrimination problem. Cybenko(1989) shows that one hidden layer is sufficient to com-
pute arbitrary decision boundaries for the classification problem. Choosing the right
number of hidden units in the layer is a more difficult problem. If we choose too few hid-
den units, the network may have problem converging on a solution. The introduction of
more hidden units allows for more versatility in the initial random weighting scheme. In-
creasing the versatility increases our chances of converging on a stable solution. Howev-
er, reducing the number of hidden units reduces the computational time needed for
training. While there have been formulas put forth to determine the correct number of
hidden units, choosing the optimum number is still based on empirical results. Past re-
searchers have simply compared the classification problem on several configurations of
networks to test the change in accuracy. I have chosen to follow this approach for the
seismic discrimination problem. In Tables 2,3,4 are results of training and testing the
data with the leave-one-out method. In Table 2 the results are obtained from a network
with 20 hidden units. In Table 3, 10 hidden units are used and in Table 4, 1 hidden unit
is used. For events detected at station ETT the neural network with 20 hidden units is no-
ticeably more accurate. Between the networks with 1 and 10 hidden units there is no ap-
preciable difference in the classification accuracy. Other researchers have usually found
that increasing the number of hidden units can improve the classification accuracy, al-
though sometimes the improvement is only a few percent at best. Because the computa-
tion times can be rather extensive when increasing the number of hidden units I did not
investigate trying more than 20 hidden units. Testing the data by the leave-one-out meth-
od involves about 60 tests for each station, so any marginal increase in accuracy did not
seem worth the computation time.

Chapter 5
DISCUSSION

Comparing performance with Linear Network

While linear networks are less flexible for pattern recognition problems, they do have an advantage over the nonlinear network since they operate in a strict gradient de-
scent algorithm. This ensures the solution will converge to the global minimum in the to-
tal error. One problem with the nonlinear network is that the solution may converge to a
local minimum in the total error. This may not be a consideration if the derived weights
still provide an accurate means of separating blasts and earthquakes. However, it would
be nice to see how the weights developed by a similar linear network compare to those
developed by the nonlinear net.

By using a similar linear network (1 Input layer, 1 Output layer), the network
was only able to converge on a solution for data recorded at station WEN. Obviously the
flexibility of the nonlinear mapping allowed for better event separation at each station.

Comparing Inputs

One obvious question is what distinguishing characteristics were used in the sepa-
ration algorithm. In previous studies, P/S spectral ratios proved to be decent discrimina-
tors (Taylor et al., 1988). This led to a test of these P/S ratios. I took P/S ratios of 1 Hz
frequency bins from 1-11 Hz. In each case the network did not converge, and I conclu-
ded that this was not a distinguishing characteristic. Rather the spectral shape is believed
to be the discriminator for the events used in this study.

The lack of differentiation in P/S wave spectral content for blasts and earthquakes
is characteristic for Washington. In 1990 Baumguardt et al, found that spectral shapes
for individual phases are a better discriminant than relative P-wave to S-wave energy in
orogenic regions such as the Western United States.
Aiding Solution Convergence

To increase the likelihood of finding meaningful solutions, the network should ideally be presented with as much training data as possible. Large learning sets will also provide more consistent discrimination when there is some variability in the inputs.

However, some training data were not easily mapped to their desired output and this prevented the network from converging on a solution. In reaching the solution weights, the learning process had to be "helped" in one case. When presented with learning data from station ETT, the network did not converge upon a solution in a reasonable amount of time, so the test data were decreased. The reason for this is that some of the "outliers" were not enabling the network to converge within the specified tolerance. To remove the outliers there are a few methods at our disposal. One can either look at the raw data fed to the neural net and delete the obvious outliers or go back to the power spectra itself to find differences. At first one can be liberal with throwing out the suspected "outliers".

The goal is to get the network to converge on a solution with a smaller set of data. Once a solution is achieved, the network can be propagated in the forward mode on the entire set of data. The output mapping of each input vector should be a good indicator as to how well the solution weights work for that given input vector. If the output is close to the desired output, the input shares similarities with the training set. If the output for a given input vector is far from a desired output, it would appear to have little in common with the training set and would be considered an outlier. Using this method we can maximize the number of events with similarities to achieve the solution weights.

The outliers could exist for a variety of reasons. For example, outliers could have anomalous spectra due to a strong high frequency spectral peaks away from the 1-11 Hz interval. This could put the characteristic spectral peaks from the source well into the side-bands. The strong high-frequency peaks may have been influenced by the depth of the event or by the close proximity of the event and the station. If the station is near

the event origin, the high frequency portion of the seismic signal is not significantly attenuated and can lead to strong high frequency peaks in the spectral estimate. Also the effect of noise may have been particularly strong on the signal. Another reason for outliers has to do with single event location versus the location of other data used in the learning process. If the event lies in an area where there are not many nearby events, the spectra may be different due to path effects as well as source effects. Also there is the possibility that the event was originally misclassified by the seismic analyst. There can be other reasons for outliers and I will investigate general properties of the misclassified events later on in this chapter.

Understanding Event Classification

In the past researchers have found explosions (Mb < 4.5) in the western U.S. to be rich in low spectral energy relative to earthquakes. Using spectral ratios (1-2 Hz/6-8 Hz) Talyor et. al. (1988) found that explosions have a higher ratio than do earthquakes. He concludes that, for events in this magnitude range, the explosions have more energy at low frequencies than do the earthquakes. Taylor explains one possibility for this disparity is a depth dependent effect due to variations in the depth-dependent attenuation factor, Q(z). He hypothesizes that if earthquake sources were located in a high Q medium and explosion sources in a low Q medium, the earthquakes would have more high frequency energy. Modeling the observed spectral differences, he explained the experimental results with a two layer model of different Q values. The model consists of a top layer of 2 km with a Q of 10 and a bottom layer of 8 km with a Q value of 100. Taylor also recognized potential differences due to source mechanism but he did not explain how the mechanism exhibits the observed spectral differences.

In this study, a concern for understanding the reasons for event discrimination leads to analysis of the distribution of event magnitude and event depth. To consider the
possibility of event magnitude or event depth as a factor in the discrimination routine, one can plot a histogram of magnitudes/depths for earthquakes and blasts. A display of magnitude distribution and depth distribution for the events recorded at stations ETT, WAT, and WEN can be seen in Figures (17a,b,c). In each case there is little evidence for magnitude being an important factor in the discrimination routine. Plots of depth distribution for events recorded at each station can be seen in the Figures (18a,b,c). Initially a case might be made for distinctions in the depth distribution between blasts and earthquakes. However with a closer look at blast depth there are some events which are obviously too deep to be accurate depth locations. These errors in depth locations can be explained by the close proximity of the source and the recording station or by the relatively few stations used in the location algorithm. For events of magnitude $M_e=1.5\text{--}2.8$ this can often be a problem because they do not have enough energy to be strongly detected at many stations. Due to questionable depth location in these cases, it is obvious that depth alone should not be a distinguishing characteristic for discriminating events. This lack of disparity between earthquakes and blasts is the reason to look for distinct indicators in the signal.

I chose to quantify the signal in terms of the frequency content. The frequency content is illustrated by estimating the power spectrum of the time series. By stacking the spectra for explosions and earthquakes we can get a feel for the gross overall characteristics of the explosions versus the earthquakes (Figures 8-10). In looking at the spectra for all three stations there is an obvious trend for the explosions to have more energy in the low frequencies relative to the earthquakes. So from the stacked spectra it is reasonable to expect the neural network probably uses this characteristic to make its discrimination scheme. However, when we look at explosions from each blasting site at one station, there is an apparent disagreement with the overall observations (Figure 11). The mining blast spectrum (Figure 11b) follows the trend of the stacked spectra, with a con-
centration low frequency energy, while the road blast spectrum has a concentration of high frequency energy. At first this may seem discouraging but on closer examination with the other explosions and earthquakes, there seems to be another consistent difference between the two event types. The explosion's signal appears to be monochromatic (Figure 12b), while the earthquake's signal is comprised of several frequencies (Figure 12a). Thus, I believe this must also be a factor in the weighting scheme.

My argument for how the events are classified should be tested. In order to gain any quantitative insight about the spectral components responsible for the difference between earthquakes and explosions, a sensitivity analysis should be done on the dataset with the neural network. In this study I have not done the sensitivity analysis problem but I outline the procedure in the following section.

Sensitivity Analysis

In reaching a solution to the discrimination problem, the neural network finds patterns in the explosion dataset that differ from the earthquake dataset. The relatively sharp spectral peaks of the explosions, (Figure 12), is expected to be the basis for classification. To test this theory, a sensitivity analysis should be done. Sensitivity analysis with a neural network involves finding the gradient of the network output with respect to a given input in the data vector. That is to say we need to find how much the output mapping of the neural network is influenced by changing a particular input of the data vector. An illustration can be seen below in Diagram 7.
For a small change in the first input of the data vector, $x_1$, there will be an associated change to the output by the neural network. After testing the sensitivity of each input in the data vector, the gradient can be estimated for each neuron (i.e., there are 20 neurons in the hidden layer). If the gradient for a single input is similar at each neuron, the solution is thought to be stable with respect to that input. If the gradient of the output varies wildly from neuron to neuron, the input is not stable with respect to the solution (J. Hwang, personal communication). The inputs with stable gradients should be the essential factors in the discrimination scheme.

The gradient is derived in Appendix B, and is given by the following expression

$$\frac{\partial o_m}{\partial x_i} = f'(sum_l) \sum_{j=1}^{N} w_{kj} \frac{\partial o_j}{\partial x_i}$$  \hspace{1cm} (9)$$

Here $o_m$ is the output of the $k^{th}$ hidden unit or neuron, $x_j$ is the $j^{th}$ input of the data vector, $f'$ is the derivative of the sigmoidal transfer function, $sum_l$ is the $k^{th}$ weighted sum of the data vector, $w_{kj}$ is the weight of the $j^{th}$ input going to the $k^{th}$ hidden unit, and $a_j$ is the activation value of the $j^{th}$ hidden unit. The above expression can then be used to quantitatively determine the inputs most important in developing the discrimination routine between earthquakes and explosions.

Understanding Misclassified Events

To understand why some events were possibly misclassified by the neural network, I have plotted the locations of these earthquakes and explosions (Figure 19). Events misclassified by any of the three stations are included in the previous figure. In Figure 19a there are three earthquakes which clearly lie outside the main earthquake cluster. There is one just north of the cluster and two northeast of the cluster. It is not surprising that these three earthquakes may have been improperly classified because they are so isolated. The signals from these events could be influenced by path effects which are quite different from path effects for earthquakes occurring near the central cluster. In order confidently classify an earthquake not involved in the learning process, it should be located near the majority of earthquakes used in the training dataset.

While misclassification of the three outlying earthquakes is easily explained due to their isolation, there are several other misclassified earthquakes which occur in the main data cluster. If we look at Figure 19b, there are also several misclassified explosions in tight clusters around the blasting sites. In order to understand why these events may have been misclassified, I first tried to find some correlation of event magnitude or event depth (Table 6). The misclassified events have varying magnitudes and depths, so I looked for indicators in the power spectra.

By looking at the P and S spectra for misclassified waveforms at station WAT they consistently deviate from the expected general spectral properties. The misclassified earthquake signals are monochromatic and the misclassified explosion signals are comprised of several frequencies (Figure 20). In short this means the signals of misclassified earthquakes are more explosion-like and the signals of misclassified explosions are more
Figure 19. (a) Locations of the misclassified earthquakes. (b) Locations of the misclassified explosions.

Figure 20. Comparison of misclassified earthquake and explosion. The signals were recorded at (a) P spectrum of misclassified earthquake (b) S spectrum of misclassified earthquake (c) P spectrum of misclassified explosion (d) S spectrum of misclassified explosion.
earthquake-like. It is very possible that these events were actually misclassified when cataloged by the seismic analyst. The neural network may not have misclassified these events at all. Unfortunately with the imperfect catalog, the exact accuracy of the neural network is not well determined. However, the accuracy rates calculated in this study do determine a lower limit for correct classification by the neural network.

In this study neural networks provide the ability to discriminate smaller seismic events ($M_e = 1.5-2.8$) with a reasonable degree of confidence. The ability to consistently discriminate earthquakes from explosions relies on spectral differences in the two source types. Searching for a global discrimination scheme based on the location of spectral peaks of the signal is a difficult problem. As shown by the spectral plots at different blasting locations (Figure 12), the explosion's spectral source characteristics can vary drastically. However, the explosions do exhibit a concentration of low frequency energy in their spectra and a consistent monochromatic nature in their signals. I believe these two distinctions are the primary reasons for event classification. The lower limit of classification accuracy varied from 75% to 89%. This result was based on a previous catalog with potential classification errors. Even this lower limit in accuracy is good when considering the difficulties in classifying small events. The accuracy is especially impressive when considering that the classification is based on a single seismic trace.
LIST OF REFERENCES


APPENDIX A: Derivation of the Delta Rule (Rumelhart, 1988)

Weights are adjusted based on the idea that change in a particular weight should be proportional to the contribution of that weight on the total error ($E$). For the output layer,

$$\Delta w_j = \frac{\partial E}{\partial w_j} \tag{10}$$

Where the total error associated with the given input/output pair is

$$E = \frac{1}{2} \sum_w (t_w - o_w)^2 \tag{11}$$

The above derivative, $\frac{\partial E}{\partial w_j}$, can be decomposed into two parts using the chain rule

$$\frac{\partial E}{\partial w_j} = \frac{\partial E}{\partial \text{sum}_k} \cdot \frac{\partial \text{sum}_k}{\partial w_j} \tag{12}$$

Where the first term on the right hand side represents the change in error as a function of the change in the net input, and the second term represents the effect of changing a particular weight on the net input.

From equation (5) the second partial derivative can be written for the output layer as

$$\frac{\partial}{\partial \text{sum}_k} \frac{\partial \text{sum}_k}{\partial w_j} = o_j \tag{13}$$

To evaluate the first partial derivative we again decompose with the chain rule

$$\frac{\partial E}{\partial \text{sum}_k} = \frac{\partial E}{\partial o_k} \cdot \frac{\partial o_k}{\partial \text{sum}_k} \tag{14}$$

Now using the definition of the error, equation (10), the first factor can be written as

$$\frac{\partial E}{\partial o_k} = -(t_k - o_k) \tag{15}$$

Using equation (2) the second term on the right hand side is just

$$\frac{\partial o_k}{\partial \text{sum}_k} = f'(\text{sum}_k) \tag{16}$$

which is simply the derivative of the transfer function for the $k^{th}$ unit.
We have now developed the weight changing scheme

$$\Delta w_{ij} = \eta (t_n - o_n) f'(sum_n) a_j$$

(17)

Or as is often written in shorthand notation

$$\Delta w_{ij} = \eta \delta_j a_j$$

(18)

Where $\eta$ is the learning rate, which dictates the amount of change in the weights. The above expression is only for the output layer. The error for the previous hidden layers is determined by back-propagating the error in the output layer and is given by the general expression

$$\delta_j = f'(sum_j) \sum_k \delta_k w_{kj}$$

(19)

After recursively determining the error we eventually get back to the input layer and the weights for the input data are adjusted. (See Rumelhart, 1988 for a more thorough discussion).

If the learning rate is too high (Equation 17), high frequency perturbations in the training set can keep the network from converging. However, if the learning rate is too low, the time for convergence will be very long. To avoid this problem one can introduce a momentum term that allows for high learning rates and effectively filters out the high frequency oscillations in the data set. The momentum factor is a way of smoothing the learning by using previous changes in the weights. This concept is seen in the following expression for the new weight change

$$\Delta w_{ij}(n+1) = \eta \delta_j a_j + \kappa \Delta w_{ij}(n)$$

(20)

Here the $\kappa$ is a constant called the momentum factor.

APPENDIX B: Derivation of the Sensitivity Gradient

The derivation of the sensitivity gradient follows from the derivation of the delta rule. The gradient of the network output, $o_n$ with respect to a given input, $x_i$, must be calculated recursively in a layer by layer process. There are three layers in the neural network used in this study: the input layer (layer 0), the hidden layer (layer 1) and the output layer (layer 2). I will derive the gradient of the activation of the hidden layer with respect to a given change in the input layer and will then quote the general result for all layers.

First, the change in the activation of the hidden layer, $a_j$, with respect to a change in a single component from the input layer, $x_i$, can be written as the product of two new derivatives.

$$\frac{\partial a_j}{\partial x_i} = \frac{\partial a_j}{\partial sum_j} \cdot \frac{\partial sum_j}{\partial x_i}$$

Now the activation $a_j$ is given by the following expression.

$$a_j = f(sum_j)$$

So the derivative of $a_j$ with respect to $sum_j$ is simply

$$\frac{\partial a_j}{\partial sum_j} = f'(sum_j)$$

Similarly, the weighted sum, $sum_j$ is given by

$$sum_j = \sum_{i=1}^{n} w_{ji} x_i$$
so the derivative of $\text{sum}_j$ with respect to $x_i$ is just

$$\frac{\partial \text{sum}_j}{\partial x_i} = w_{ji}$$

Multiplying the two partial derivatives yields a simple expression for the sensitivity gradient for the hidden layer

$$\frac{\partial z_j}{\partial x_i} = f'(\text{sum}_j) w_{ji}$$

The above result is the sensitivity for the activation of hidden unit $j$ with respect to input $x_i$. In order to estimate the change in the final output of the neural network the above results must propagate through the next layer (i.e., the output layer). This gives the following expression for the gradient at the next layer

$$\frac{\partial o_k}{\partial x_i} = f'(\text{sum}_j) \sum_{j=1}^{N} w_{kj} \frac{\partial z_j}{\partial x_i}$$

This is the change in the network output with a change in a given input. Here $\text{sum}_j$ is the weighted sum of the activations from the hidden layer to the $k$th unit in the output layer. The number $N$ is the number of hidden units in the hidden layer. See Hwang (1991) or Kuhn (1992) for a more general expression and detailed applications.
Table 5: Results of Testing Groups of Unlearned Data

<table>
<thead>
<tr>
<th>Station</th>
<th>Number of Events in Training (X/Q)</th>
<th>Number of Test Events (X/Q)</th>
<th>Correctly Classified</th>
<th>Incorrectly Classified and Undecided</th>
<th>Accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETT</td>
<td>19/23</td>
<td>20/10</td>
<td>24</td>
<td>6</td>
<td>80</td>
</tr>
<tr>
<td>WAT</td>
<td>17/18</td>
<td>10/10</td>
<td>15</td>
<td>5</td>
<td>75</td>
</tr>
<tr>
<td>WEN</td>
<td>15/19</td>
<td>10/10</td>
<td>18</td>
<td>2</td>
<td>90</td>
</tr>
</tbody>
</table>

Table 6: Neural Network Classification Catalog

<table>
<thead>
<tr>
<th>yr,mo,day,hr,min</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Depth (km)</th>
<th>Mag.</th>
<th>ETT</th>
<th>WAT</th>
<th>WEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>AX8105081901</td>
<td>47N3738</td>
<td>119W5404</td>
<td>1.70</td>
<td>2.2</td>
<td></td>
<td></td>
<td>Q</td>
</tr>
<tr>
<td>AX8402092328</td>
<td>47N5012</td>
<td>119W5874</td>
<td>0.02</td>
<td>1.9</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8403071745</td>
<td>47N5006</td>
<td>119W5880</td>
<td>0.27</td>
<td>1.9</td>
<td>Q</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8403212018</td>
<td>47N5096</td>
<td>119W5846</td>
<td>0.22</td>
<td>2.1</td>
<td>X</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>AX8405161608</td>
<td>47N5091</td>
<td>119W5818</td>
<td>0.33</td>
<td>2.4</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AX8406061748</td>
<td>47N5031</td>
<td>119W5905</td>
<td>0.36</td>
<td>2.1</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8406151454</td>
<td>47N5083</td>
<td>119W5830</td>
<td>0.50</td>
<td>2.0</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>AX8407101517</td>
<td>47N5087</td>
<td>119W5840</td>
<td>0.38</td>
<td>2.1</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>AX8410031837</td>
<td>47N2403</td>
<td>120W2026</td>
<td>0.50</td>
<td>2.2</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AX8410060030</td>
<td>47N2466</td>
<td>120W2022</td>
<td>0.81</td>
<td>2.0</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AX8410070025</td>
<td>47N2411</td>
<td>120W2000</td>
<td>0.39</td>
<td>2.0</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AX8410182359</td>
<td>47N2364</td>
<td>120W2094</td>
<td>0.71</td>
<td>2.3</td>
<td>Q</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AX8410240020</td>
<td>47N2342</td>
<td>120W2099</td>
<td>0.42</td>
<td>2.2</td>
<td>Q</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AX8410262209</td>
<td>47N2395</td>
<td>120W2163</td>
<td>0.40</td>
<td>1.7</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AX8410300105</td>
<td>47N2355</td>
<td>120W2072</td>
<td>0.55</td>
<td>2.3</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8411142157</td>
<td>47N5062</td>
<td>119W5795</td>
<td>0.54</td>
<td>1.5</td>
<td></td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>AX8411162007</td>
<td>47N5086</td>
<td>119W5835</td>
<td>0.04</td>
<td>1.8</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8411191652</td>
<td>47N4910</td>
<td>119W5943</td>
<td>0.03</td>
<td>1.8</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8411202316</td>
<td>47N2377</td>
<td>120W2078</td>
<td>0.84</td>
<td>2.0</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AX8411262324</td>
<td>47N2354</td>
<td>120W2069</td>
<td>3.56</td>
<td>1.7</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8412012113</td>
<td>47N2387</td>
<td>120W2122</td>
<td>0.99</td>
<td>1.9</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8412121647</td>
<td>47N4899</td>
<td>119W5922</td>
<td>0.32</td>
<td>1.6</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AX8412280013</td>
<td>47N2383</td>
<td>120W2097</td>
<td>0.45</td>
<td>1.5</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AX8501032211</td>
<td>47N2370</td>
<td>120W2098</td>
<td>0.59</td>
<td>1.5</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8505182258</td>
<td>47N2348</td>
<td>120W2028</td>
<td>0.39</td>
<td>2.1</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>yr,mo,day,hr.min</td>
<td>Latitude</td>
<td>Longitude</td>
<td>Depth (km)</td>
<td>Mag.</td>
<td>ETT</td>
<td>WAT</td>
<td>WEN</td>
</tr>
<tr>
<td>------------------</td>
<td>----------</td>
<td>-----------</td>
<td>------------</td>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>AX8505202350</td>
<td>47N2379</td>
<td>120W2084</td>
<td>0.52</td>
<td>2.3</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8505221825</td>
<td>47N2361</td>
<td>120W2075</td>
<td>0.48</td>
<td>2.2</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8505251821</td>
<td>47N2381</td>
<td>120W2028</td>
<td>0.45</td>
<td>2.1</td>
<td>Q</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8505281810</td>
<td>47N2401</td>
<td>120W2099</td>
<td>0.67</td>
<td>2.3</td>
<td>Q</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8505291812</td>
<td>47N2407</td>
<td>120W2071</td>
<td>3.53</td>
<td>2.1</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8505301811</td>
<td>47N2446</td>
<td>120W2097</td>
<td>6.64</td>
<td>1.6</td>
<td>Q</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8505311817</td>
<td>47N2396</td>
<td>120W2136</td>
<td>0.64</td>
<td>1.6</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AX8506041819</td>
<td>47N2429</td>
<td>120W2152</td>
<td>5.56</td>
<td>1.5</td>
<td>?</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>AX8506022300</td>
<td>47N2411</td>
<td>120W2158</td>
<td>0.65</td>
<td>1.9</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8506024105</td>
<td>47N2417</td>
<td>120W2074</td>
<td>0.39</td>
<td>1.6</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8506022523</td>
<td>47N2436</td>
<td>120W2126</td>
<td>0.36</td>
<td>1.8</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8506022625</td>
<td>47N2457</td>
<td>120W2181</td>
<td>2.51</td>
<td>1.8</td>
<td>Q</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AX8510052352</td>
<td>47N4914</td>
<td>119W5951</td>
<td>0.02</td>
<td>1.8</td>
<td>X</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>AX8510100010</td>
<td>47N4914</td>
<td>119W5920</td>
<td>0.52</td>
<td>1.8</td>
<td>X</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>AX8510242345</td>
<td>47N4935</td>
<td>119W5914</td>
<td>0.02</td>
<td>1.6</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8511002027</td>
<td>47N4933</td>
<td>119W5951</td>
<td>0.02</td>
<td>1.5</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AX8511050052</td>
<td>47N4961</td>
<td>119W5928</td>
<td>0.04</td>
<td>1.7</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>AX8511090243</td>
<td>47N4920</td>
<td>119W5923</td>
<td>0.55</td>
<td>1.5</td>
<td>X</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>AX8511200002</td>
<td>47N4911</td>
<td>119W5937</td>
<td>0.03</td>
<td>2.3</td>
<td>X</td>
<td>?</td>
<td>Q</td>
</tr>
<tr>
<td>A 8112102046</td>
<td>47N4941</td>
<td>119W3719</td>
<td>3.61</td>
<td>2.2</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>A 8202180327</td>
<td>47N3984</td>
<td>119W4470</td>
<td>0.80</td>
<td>2.8</td>
<td>Q</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 8202181354</td>
<td>47N3947</td>
<td>119W4454</td>
<td>0.78</td>
<td>1.5</td>
<td>Q</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>A 8205080706</td>
<td>47N3219</td>
<td>119W4350</td>
<td>11.01</td>
<td>1.5</td>
<td>Q</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 8210140853</td>
<td>47N4288</td>
<td>120W1151</td>
<td>3.87</td>
<td>2.4</td>
<td>Q</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 8210182043</td>
<td>47N5480</td>
<td>120W0298</td>
<td>0.02</td>
<td>2.3</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A 8211052039</td>
<td>47N4579</td>
<td>120W0508</td>
<td>0.52</td>
<td>1.9</td>
<td>Q</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>yr,mo,day.hr.min</td>
<td>Latitude</td>
<td>Longitude</td>
<td>Depth (km)</td>
<td>Mag.</td>
<td>ETT</td>
<td>WAT</td>
<td>WEN</td>
</tr>
<tr>
<td>------------------</td>
<td>----------</td>
<td>-----------</td>
<td>------------</td>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>A 8509200914</td>
<td>47N4058</td>
<td>120W1190</td>
<td>0.63</td>
<td>2.1</td>
<td>X</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>A 8509262244</td>
<td>47N4910</td>
<td>119W5961</td>
<td>0.51</td>
<td>1.6</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>A 8509301645</td>
<td>47N4332</td>
<td>120W0816</td>
<td>5.09</td>
<td>1.8</td>
<td>Q</td>
<td>X</td>
<td>Q</td>
</tr>
<tr>
<td>A 8510070237</td>
<td>47N4254</td>
<td>120W1923</td>
<td>4.59</td>
<td>1.5</td>
<td>Q</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>A 8510101006</td>
<td>47N4495</td>
<td>120W1593</td>
<td>7.04</td>
<td>3.2</td>
<td></td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>A 8510111127</td>
<td>47N5515</td>
<td>119W3416</td>
<td>0.54</td>
<td>2.1</td>
<td>Q</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>A 8510300338</td>
<td>47N4293</td>
<td>120W0324</td>
<td>5.57</td>
<td>1.8</td>
<td>Q</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>A 851120117</td>
<td>47N4297</td>
<td>120W1619</td>
<td>7.66</td>
<td>2.2</td>
<td>Q</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>A 8512021333</td>
<td>47N4023</td>
<td>120W1958</td>
<td>4.55</td>
<td>1.7</td>
<td></td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>A 8708081332</td>
<td>47N3894</td>
<td>120W1484</td>
<td>0.51</td>
<td>2.2</td>
<td>Q</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>A 8709271418</td>
<td>47N4379</td>
<td>120W270</td>
<td>2.68</td>
<td>2.0</td>
<td>Q</td>
<td>Q</td>
<td></td>
</tr>
<tr>
<td>A 8711290452</td>
<td>47N4128</td>
<td>120W1223</td>
<td>0.71</td>
<td>2.0</td>
<td>Q</td>
<td>Q</td>
<td></td>
</tr>
</tbody>
</table>