VARUN PRAVEEN
CUID: pravee2

ECE – 8720

Takehome #2

1. Balancing the Data
    a. Input Data: The data provided comprised of 27-dimensional EEG data from two classes, namely AEP and non AEP. The AEP dataset comprised of 83 samples, whereas the non-AEP samples comprised of 2482 such samples.
    The artificial neural network based classifier designed for this example uses tanh(net) activation function which swings between a range of [-1,1]. Also, the output layer of the ANN comprises of 2 neurons.
    Thus, the target output for each class was defined as follows:
    AEP = [1, -1]
    Non AEP = [-1, 1]
    b. Balancing Datasets: The data provided does not have equal number of data for both classes, thus, implying the non-AEP class is the majority class, while the AEP is the minority class. Inorder to remove this difference in the datasets which would be used to train the ANN, balancing of the dataset was performed.
    There are different methods of balancing, of which the two balancing techniques compared here were:
        - K – medoids : Here, the majority class was clustered to 83 different clusters and the centroid of each cluster was used to generalize the majority class
        - Near miss 2: Here, 83 vectors from the non-AEP class were chosen, such that they had the least average distances from its respective 3 farthest AEP points. This helps in choosing non AEP class points which are closest to the boundary between two classes.

    For this assignment, the Near-miss 2 balancing methodology was chosen.

    Thus, the cardinality of the total sample space post balancing was 166 sample points with 83 in AEP and 83 in non AEP.

    The training set H comprised of the first 63 AEP points (75% of the number of AEP samples) and first 63 non AEP points.

    The test set S_T comprised of the remaining 20 points in the AEP and non AEP points.

    Also, finally the remaining non AEP data samples were also tested to check if the balancing of the data, affected the generalization of the non AEP class, and it was seen that while the ANN achieved an average classification error of 10% in terms of S_T , but the error for the remaining 2399 samples yielded approximately 88% classification error. Thus, near miss 2 while balancing the dataset, fails at generalizing the data from the majority class.

    Note: The error plotted in all the case for every 200 iterations.

2. Case 1: Generalized Delta Rule (no bias)
   a. Design Parameters:
      - d = 27
      - H = number of neurons in the hidden layer : 2d+1 = 55
      - Squasher function: tanh()
      - Hidden layer weight initialization: random vector in [-0.00005 , 0.00005]
      - Output layer weight initialization : random vector in [-0.05 , 0.05]
      - Learning rate: 10^{-8}
      - Maximum Number of iterations: 100,000
      - Error measure : TSS = (1/2)sum((error).^2)
   b. Results:

The ANN trained using this network resulted with a **TSS : 21.0542**

The confusion matrix for the training and test set is as shown below

H

Ground Truth\Output	AEP	Non AEP
AEP	58	5
Non AEP	0	63

Sensitivity: 92%   Specificity: 100%

S_T

Ground Truth\Output	AEP	Non AEP
AEP	16	4
Non AEP	2	18

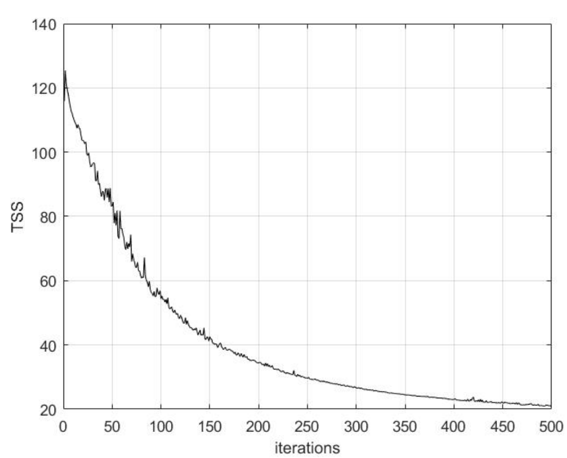
Sensitivity: 80%   Specificity: 90%

Total Dataset:

Ground Truth\Output	AEP	Non AEP
AEP	74	9
Non AEP	2222	260

Sensitivity: 89%   Specificitiy : 10%

TSS Error Plot:



3. Case 2: Generalized Delta Rule (with bias)
    a. Design Parameters:
        - d = 27
        - H = number of neurons in the hidden layer : 2d+1 = 55
        - Squasher function: tanh()
        - Hidden layer weight initialization: random vector in [-0.00005 , 0.00005]
        - Output layer weight initialization : random vector in [-0.05 , 0.05]
        - Learning rate: 10^{-8}
        - Maximum Number of iterations: 100,000
        - Error measure : TSS = (1/2)sum((error).^2)
    b. Results:
    The ANN trained using this network resulted with a **TSS : 30.9142.** When compared with case 1 the specificity has been reduced and the error has increased. The confusion matrix for the training and test set is as shown below

H

Ground Truth\Output	AEP	Non AEP
AEP	57	6
Non AEP	0	63

Sensitivity: 90.4% Specificity: 100%

S_T

Ground Truth\Output	AEP	Non AEP
AEP	16	4
Non AEP	3	17

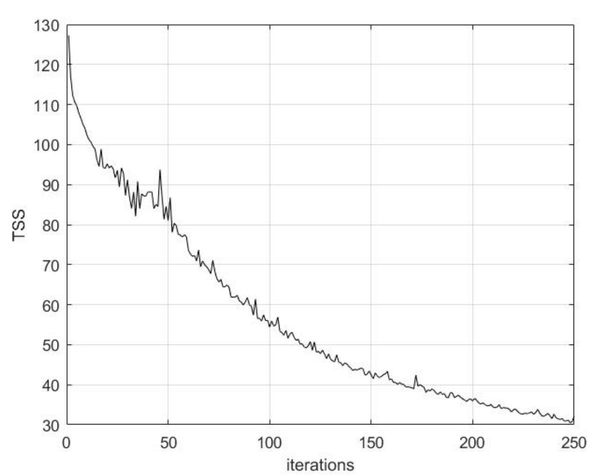
Sensitivity: 80% Specificity: 85%

Total Dataset:

Ground Truth\Output	AEP	Non AEP
AEP	73	10
Non AEP	2220	262

Sensitivity: 87.9% Specificitiy : 10.5%

TSS Error Plot:



4. Case 3: Generalized Delta Rule with bias and weight initialization using ROM
   a. Design Parameters:
      - As in case 2 with Random optimization method for weight initialization
      - ROM iterations: 200
      - ROM Gaussian variables
        1. Hidden layer in [-0.000005 , 0.000005]
        2. Output layer in [-0.005 , 0.005]
      - LR : 10^{-8}
      - Maximum number of iterations : 100,000
   b. Results:
      The ANN trained using this network resulted with a **TSS : 22.5927.** It can be noted that the starting TSS for this case when entering GDR is lesser than that from case 1 and 2, thereby facilitating in converging faster and bouncing off of a local minima if necessary. The confusion matrix for the training and test set is as shown below

H

Ground Truth\Output	AEP	Non AEP
AEP	58	5
Non AEP	0	63

Sensitivity: 92% Specificity: 100%

S_T

Ground Truth\Output	AEP	Non AEP
AEP	16	4
Non AEP	3	17

Sensitivity: 80% Specificity: 85%

Total Dataset:

Ground Truth\Output	AEP	Non AEP
AEP	74	9
Non AEP	2222	260

Sensitivity: 89% Specificity: 10%

TSS Error Plot:

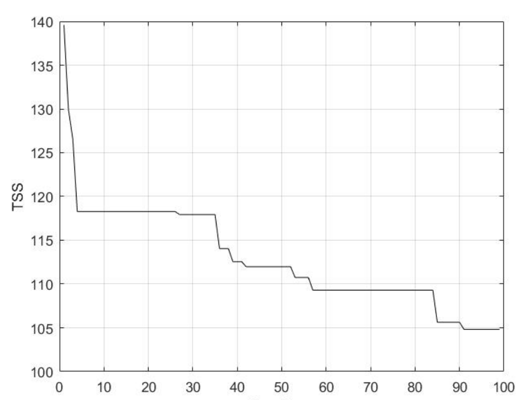


Figure 2: TSS plot for ROM

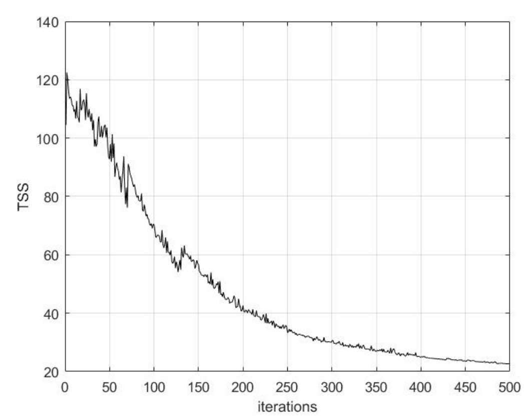


Figure 1: TSS plot after GDR post ROM

5. Case 4: Generalized Delta Rule with bias and weight initialization using ROM and momentum
    a. Design Parameters:
        - As in case 3
        - Maximum number of iterations : 30,000
    b. Results:

The ANN trained using this network resulted with a **TSS : 43.7285.** Thus, the total TSS has converged to a comparable value within 1/3rd of the iterations as in case 3, thereby implying the momentum causes the weight to converge faster

Ground Truth\Output	AEP	Non AEP
AEP	51	12
Non AEP	0	63

Sensitivity: 81% Specificity: 100%

S_T

Ground Truth\Output	AEP	Non AEP
AEP	16	4
Non AEP	1	19

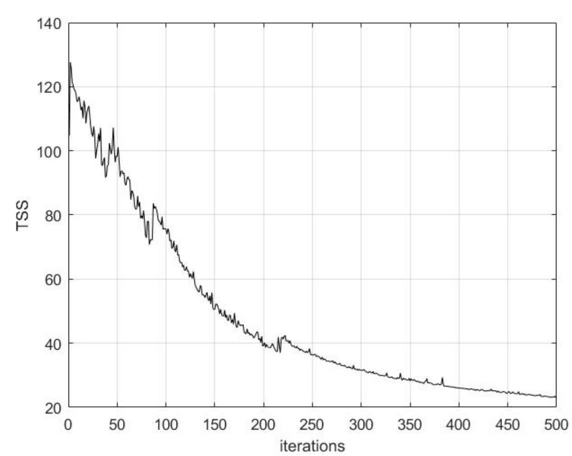
Sensitivity: 80% Specificity: 95%

Total Dataset:

Ground Truth\Output	AEP	Non AEP
AEP	67	16
Non AEP	2189	293

Sensitivity: 80.7% Specificity: 11.8%

TSS Error Plot:



6. Case 5: Generalized Delta Rule with bias and weight initialization using ROM and weight decay
   a. Design Parameters:
      - Weight decay factor : beta: 0.7
      - Maximum number of iterations : 100,000
      - All other parameters remain the same as in case 3
   b. Results:
      The ANN trained using this network resulted with a **TSS : 22.2405.** Thus, the total TSS has to a value lesser than that of case 4. Thereby implying that the weight decay improves the performance of the ANN.

H

Ground Truth\Output	AEP	Non AEP
AEP	59	4
Non AEP	1	62

Sensitivity: 93% Specificity: 98.4%

$\mathbf{S_T}$

Ground Truth\Output	AEP	Non AEP
AEP	16	4
Non AEP	3	17

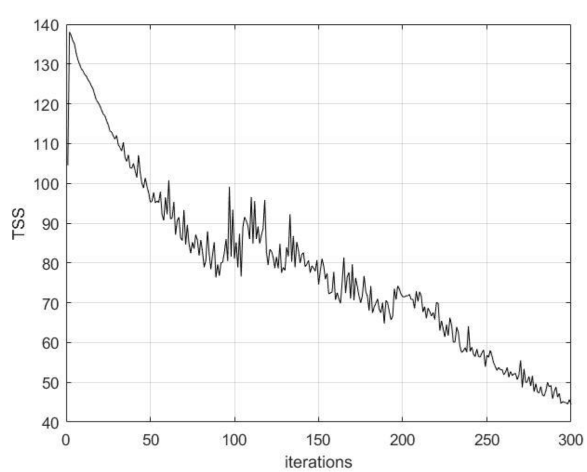
Sensitivity: 80% Specificity: 85%

Total Dataset:

Ground Truth\Output	AEP	Non AEP
AEP	75	8
Non AEP	2247	235

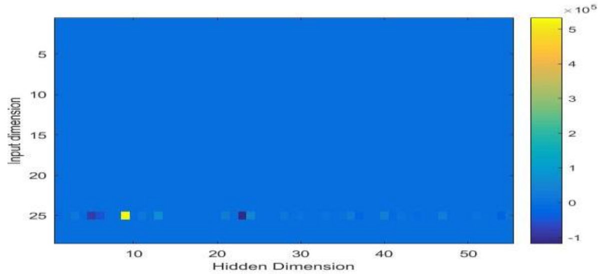
Sensitivity: 90.36% Specificity: 9.47%

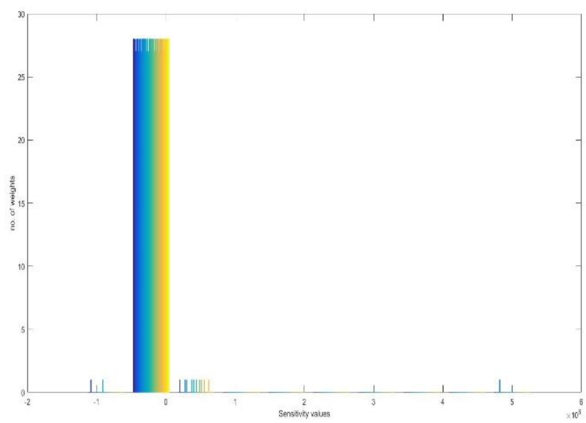
TSS Error Plot:



7. Case 5: Generalized Delta Rule with bias and weight initialization using ROM and weight decay

After ROM, the ANN generated a TSS of 22.5927 after a 100,000 iterations. The sensitivity matrix for the weights in the hidden layer of this case was calculated and the sensitivity map was plotted as below and the histogram of the weight distribution was plotted following that. The final TSS : 22.56





From the histogram of weight sensitivities we can see that there is a significant amount of weights whose sensitivities is close to 0. Therefore, there are a lot of weights which may be pruned. Thus, on checking the histogram and sensitivity map a threshold of 0.001 was chosen. All weights with sensitivity less than this threshold was pruned out and the performance of this updated network was tabulated as given later in this page and is comparable to performance from case 3.The number of weights pruned was noted to be 694. Most of the weights pruned are associated to features 1-4 and the bias implying that the bias does not have a very pronounced effect on the network.

H

Ground Truth\Output	AEP	Non AEP
AEP	59	4
Non AEP	1	62

Sensitivity: 93% Specificity: 98.4%

S_T

Ground Truth\Output	AEP	Non AEP
AEP	16	4
Non AEP	3	17

Sensitivity: 80% Specificity: 85%

Total Dataset:

Ground Truth\Output	AEP	Non AEP
AEP	75	8
Non AEP	2222	260

Sensitivity: 90.36% Specificity: 10.47%