Heaven's Light is Our Guide
Computer Science & Engineering
Rajshahi University of Engineering & Technology

Course No: CSE 4204
Course Name: Sessional based on CSE 4203

Experiment No: 4

Name of the Experiment:

a. Design and implementation of Kohonen Self-organizing Neural Networks algorithm.
b. Design and implementation of Hopfield Neural Networks algorithm.

Submitted to

Rizoan Toufiq
Assistant Professor
Dept. of Computer Science & Engineering
Rajshahi University of Engineering & Technology

Submitted by

Mondol Mridul Provakar
Roll: 1803062
Dept. of Computer Science & Engineering
Rajshahi University of Engineering & Technology

Date of Submission: 19/12/2023

Contents

1 Breast Cancer Dataset

Breast cancer is the most common cancer amongst women in the world. It accounts for 25% of all cancer cases, and affected over 2.1 Million people in 2015 alone. It starts when cells in the breast begin to grow out of control. These cells usually form tumors that can be seen via X-ray or felt as lumps in the breast area. The key challenges against it's detection is how to classify tumors into malignant (cancerous) or benign(non cancerous). The dataset has following characteristics:

- **Size:** The dataset consists of a total of 569 rows and 6 columns.
- **Data Types:** All values are numerical.
- **Features:** Mean_radius, Mean_texture, Mean_perimeter, Mean_area, Mean_smoothness.
- **Target Variable:** Diagnosis.

1.1 Exploratory Data Analysis

In a correlation heatmap, the values displayed within the heatmap represent the correlation coefficients between pairs of variables. These correlation coefficients quantify the strength and direction of the linear relationship between two variables. The values in a correlation heatmap typically range between -1 and +1:

- +1: A correlation coefficient of 1 indicates a perfect positive linear relationship between the variables. This means that as one variable increases, the other variable also increases proportionally.
- 0: A correlation coefficient of 0 indicates no linear relationship between the variables. There's no systematic linear pattern in their relationship.
- -1: A correlation coefficient of -1 indicates a perfect negative linear relationship between the variables. This means that as one variable increases, the other variable decreases proportionally.

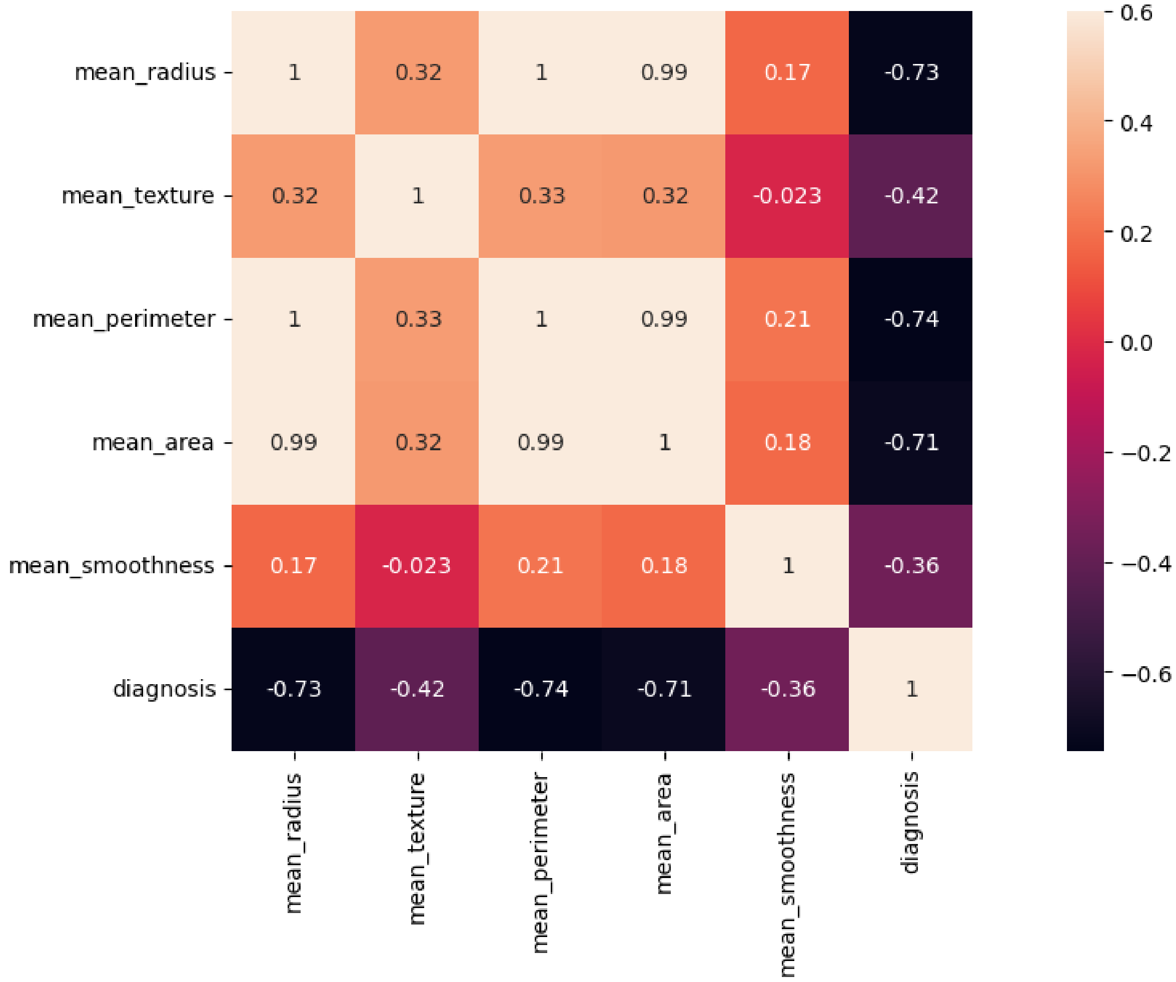

Figure 1: Correlation Heatmap

- **Correlation Heatmap:** The above correlation heatmap was generated to visualize the correlations between continuous attributes. It was found out that there were high correlations among the features: Mean_perimeter and Mean_area. It was seen that Mean_perimeter was correlated with Mean_radius by 1 and Mean_area was correlated with Mean_radius by 0.99.

1.2 Training and Test Dataset Ratio

In this analysis, the dataset is divided into a training set and a test set with an 80/20 split. This means that 80% of the data is used for training the single layer perceptron algorithm, while the remaining 20% is reserved for testing and evaluating the classifier's performance.
After train & test split:

Training dataset size: 455 rows and 6 columns.
Test dataset Size: 114 rows and 6 columns.

2 Kohonen Self-organizing Neural Network

2.1 Kohonen Self-organizing Neural Network

1. Initialise network
   Define $w_{ij}(t)$ (0 i n 1) to be the weight from input i to node j at time t.
   Initialise weights from the n inputs to the nodes to small random values. Set the initial radius of the neighbourhood around node $j, N_j(0),$ to be large.
2. Present input
   Present input $X_p = x_0(t), x_1(t), x_2(t), \ldots, x_{n-1}(t)$ where $x_i(t)$ is the input node i at time t.
3. Calculate Distances
   Compute the distance d_j between the input and each output node $j,$ given by
   $$d_j = \sum_{i=0}^{n-1}(x_i(t) - w_{ij}(t))^2$$
4. Select minimum distance
   Designate the output node with minimum d_j to be j^*.
5. Update weights
   Update weights for node j^* and its neighbours, defined by the neighbourhood size $N_{j^*}(t)$. New weights are
   $w_{ij}(t+1) = w_{ij}(t) + \eta(t)(x_i(t) - w_{ij}(t))$
   For j in $N_{j^*}(t),\ 0 \leq i \leq n-1$
   The term $\eta(t)$ is a gain term $(0 < \eta(t) < 1)$ that decreases in time, so slowing the weight adaption. Notice that the neighbourhood $N_{j^*}(t)$ decreases in size as time goes on, thus localising the area of maximum activity.

2.2 Model Evaluation

The Kohonen Self-organizing Neural Network was implemented and the accuracy score was calculated.

- **Accuracy Score:** 54.39%

3 Hopfield Neural Network

3.1 Hopfield Neural Network

1. Assign connection weights

   $w_{ij} = \begin{cases} \sum_{s=0}^{M-1} x_{si} x_{sj} & i \neq j \\ 0 & i = j, 0 \leq i, j \leq M-1 \end{cases}$

   where w_{ij} is the connection weight between node i and node j,
   and x_{si} is element i of the exemplar pattern for class s,
   and is either +1 or -1. There are M patterns, from 0 to $M-1,$ in total.
   The thresholds of the units are zero.
2. Initialise with unknown pattern

   $\mu_i(0) = x_i, \quad 0 \leq i \leq N-1$

   where $\mu_i(t)$ is the output of node i at time t.
3. Iterate until convergence

   $\mu_i(t+1) = fh\left(\sum_{j=0}^{N-1} w_{ij}\mu_j(t)\right), \quad 0 \leq j \leq N-1$

   The function fh is the hard-limiting non-linearity, the step function. Repeat the iteration until the outputs from the nodes remain unchanged.

3.2 Model Evaluation

The Hopfield Neural Network Neural Network was implemented and the accuracy score was calculated.

- **Accuracy Score:** 98.25%

4 Conclusion

- Explored dataset using Exploratory Data Analysis techniques to understand its structure and relationships.
- Developed Kohonen Self-organizing Neural Network with appropriate architecture.
- Trained the Kohonen Self-organizing Neural Network model using specified parameters (epochs, learning rate) on the dataset. Accuracy was 54.39%.
- Developed Hopfield Neural Network with appropriate architecture.
- Trained the Hopfield Neural Network model using specified parameters (epochs, learning rate) on the dataset. Accuracy was 98.25%.

References

[1] R. Beale and T. Jackson, *Neural Computing: An introduction.*, CRC Press, 1990

[2] Breast Cancer Dataset, [Online]. Available at: https://www.kaggle.com/datasets/yasserh/breast-cancer-dataset

[3] Source Code, [Online]. Available at: https://colab.research.google.com/drive/124Lc4-a1_xR1u_KmKho_Ov0nPl0yr_xY#scrollTo=gGee9Lnlic4o