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: 2

Name of the Experiment: Design and implementation of single layer perceptron learning 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: 18/11/2023

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.

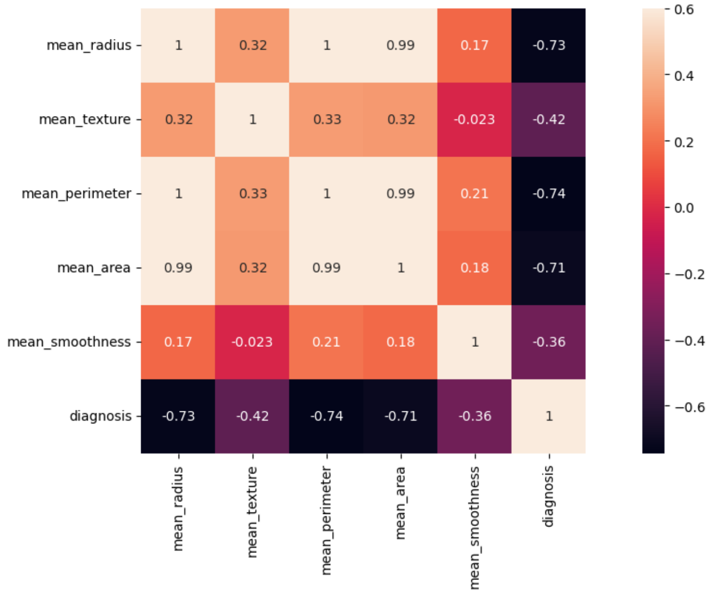

Figure 1: Correlation Heatmap

- **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. So these two columns were removed from the dataset.

  Final Dataset Size: 569 rows and 4 columns.

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 4 columns.
Test dataset Size: 114 rows and 4 columns.

2 Single Layer Perceptron Algorithm

2.1 The perceptron learning algorithm.

1. Initialise weights and threshold:
   Define $w_i(t), (0 \leq i \leq n)$, to be the weight from input i at time t , and θ to be the threshold value in the output node.Set w_0 to be $-\theta$, the bias, and w_0 to be always 1.

   Set $w(0)$ to small random values, thus initialising all the weights and the threshold.

2. Present input and desired output Present input $x_0, x_1, x_2, \ldots, x_n$ and desired output $d(t)$

3. Calculate actual output

$$y(t) = f_h \left[\sum_{i=0}^{n} w_i(t) x_i(t) \right]$$

4. Adapt weights
   if correct

$$w_i(t+1) = w_i(t)$$

   if output 0, should be 1(class A)

$$w_i(t+1) = w_i; (t) - x_i(t)$$

   if output 1, should be 0 (class B)

$$w_i(t+1) = w_i(t) - x_i(t)$$

4. Adapt weights-modified version
   if correct

$$w(t+1) = w(t)$$

   if output 0, should be 1(class A)

$$w_i(t+1) = w_i(t) - \eta x_i(t)$$

   if output 1, should be 0 (class B)

$$w_i(t+1) = w_i(t) - \eta x_i(t)$$

4. Adapt weights-Widrow-Hoff delta rule

$$\Delta = d(t) - y(t)$$
$$w_i(t+1) = w_i(t) - \eta \Delta x_i(t)$$

$$d(t) = \begin{cases} +1, & \text{if input from class A} \\ 0, & \text{if input from class B} \end{cases}$$

where $0 \leq \eta \leq 1$, a positive gain term that controls the adaption rate.

2.2 Model Evaluation

The perceptron learning algorithm was implemented with the given 3 ways to adapt the weights separately. Those are adapt weights, adapt weights - modified version & adapt weights - Widrow-Hoff delta rule. For each way, the accuracy score & the confusion matrix was calculated.

2.2.1 Adapt weights

- **Accuracy Score:** 88.6%
- **Confusion Matrix:**

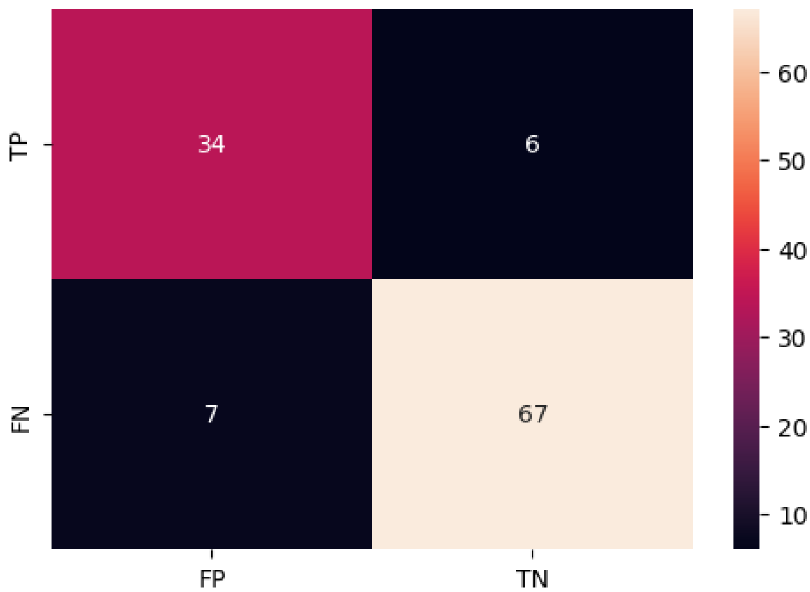

Figure 2: Confusion Matrix

- **Discussion:** The above confusion matrix consists of four different metrics:
- True Positives (TP): The number of samples correctly predicted as positive. Here 34 samples are predicted as true positive. It means they had breast cancer and the model predicted them as positive.
- True Negatives (TN): The number of samples correctly predicted as negative. Here 67 samples are predicted as true negative. It means they did not have breast cancer and the model predicted them as negative.
- False Positives (FP): The number of samples incorrectly predicted as positive. Here 7 samples are predicted as false positive. It means they did not have breast cancer and the model predicted them as positive.
- False Negatives (FN): The number of samples incorrectly predicted as negative. Here 6 samples are predicted as false negative. It means they had breast cancer and the model predicted them as negative.

2.2.2 Adapt weights - modified version

- **Accuracy Score:** 88.6%
- **Confusion Matrix:**

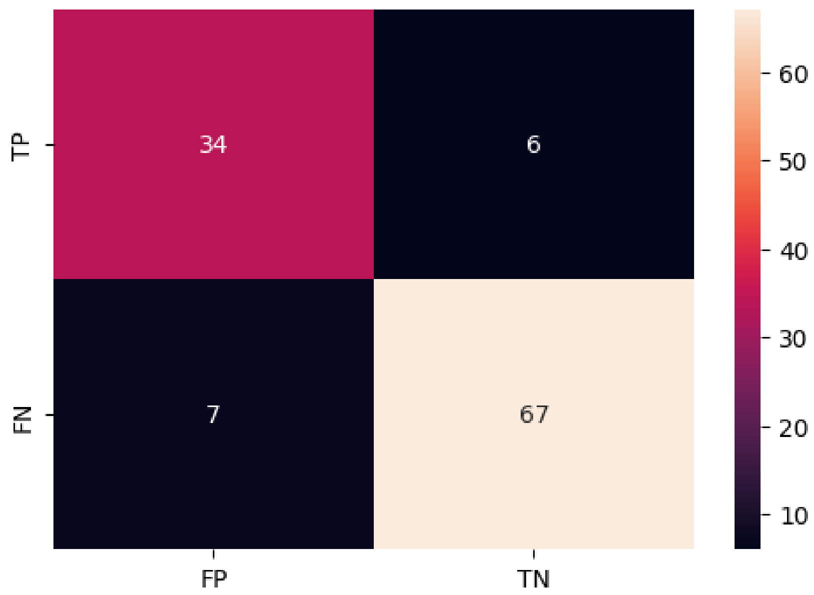

Figure 3: Confusion Matrix

- **Discussion:** The above confusion matrix consists of four different metrics:
- True Positives (TP): The number of samples correctly predicted as positive. Here 34 samples are predicted as true positive. It means they had breast cancer and the model predicted them as positive.
- True Negatives (TN): The number of samples correctly predicted as negative. Here 67 samples are predicted as true negative. It means they did not have breast cancer and the model predicted them as negative.
- False Positives (FP): The number of samples incorrectly predicted as positive. Here 7 samples are predicted as false positive. It means they did not have breast cancer and the model predicted them as positive.
- False Negatives (FN): The number of samples incorrectly predicted as negative. Here 6 samples are predicted as false negative. It means they had breast cancer and the model predicted them as negative.

2.2.3 Adapt weights - Widrow-Hoff delta rule

- **Accuracy Score:** 89.47%
- **Confusion Matrix:**

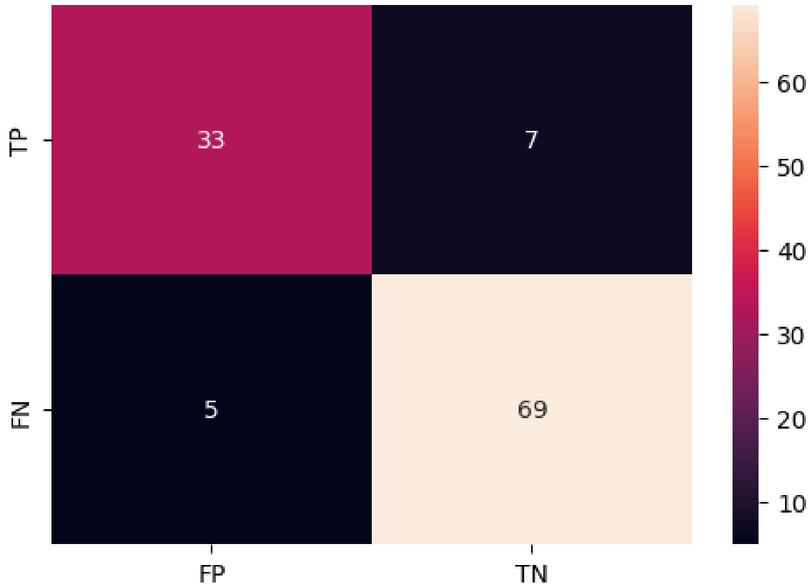

Figure 4: Confusion Matrix

- **Discussion:** The above confusion matrix consists of four different metrics:
- True Positives (TP): The number of samples correctly predicted as positive. Here 33 samples are predicted as true positive. It means they had breast cancer and the model predicted them as positive.
- True Negatives (TN): The number of samples correctly predicted as negative. Here 69 samples are predicted as true negative. It means they did not have breast cancer and the model predicted them as negative.
- False Positives (FP): The number of samples incorrectly predicted as positive. Here 5 samples are predicted as false positive. It means they did not have breast cancer and the model predicted them as positive.
- False Negatives (FN): The number of samples incorrectly predicted as negative. Here 7 samples are predicted as false negative. It means they had breast cancer and the model predicted them as negative.

3 Conclusion

For the chosen dataset, single layer perceptron algorithm was used. The single layer perceptron algorithm works well with those datasets where they are linearly separable. Here, 3 different adapt-weight procedures had been demonstrated. It was found that accuracy was slightly better using Widrow-Hoff delta rule.

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