**Task\_1**

|  |  |
| --- | --- |
| Name | ID |
| عبدالعزيز حسام عبدالعزيز | 20201700457 |
| شهاب مصطفي فهمي | 20201700405 |
| صالح عادل صالح | 20201700415 |
| عبدالرحمن حسني محمد | 20201700434 |
| محمد محروس محمد | 20201700731 |

**Perceptron:**

Different Features:

|  |  |  |
| --- | --- | --- |
| Hyperparameters | | |
| **C1 = 1, C2 = 2, eta = 0.1, m = 1000, b = 0** | | |
| Features | | |
| **F1 = 1, F2 = 2** | **F1 = 2, F2 = 4** | **F1 = 3, F2 = 5** |
| Visualization | | |
|  |  |  |
|  |  |  |
| Analysis | | |
| **Perceptron algorithm managed to discriminate between different features all in one epoch with an accuracy of 100% as all of them are linearly separable as none of them took all the epochs.** | | |

Different Classes:

|  |  |  |
| --- | --- | --- |
| Hyperparameters | | |
| **F1 = 2, F2 = 3, eta = 0.1, m = 1000, b = 0** | | |
| Classes | | |
| **C1 = 1, C2 = 2** | **C1 = 2, C2 = 3** | **C1 = 1, C2 = 3** |
| Visualization | | |
|  |  |  |
|  |  |  |
| Analysis | | |
| **The second case shows that the accuracy is 95% and it took all the epochs which means the model needs more epochs to discriminate well or these features with these classes are not linearly separable.** | | |

Different eta:

|  |  |  |
| --- | --- | --- |
| Hyperparameters | | |
| **F1 = 1, F2 = 4, C1 = 1, C2 = 3, m = 1000, b = 0** | | |
| Eta | | |
| **Eta = 0.1** | **Eta = 0.01** | **Eta = 0.00001** |
| Visualization | | |
|  |  |  |
|  |  |  |
| Analysis | | |
| **Changing the learning rate affects how fast the model reaches the local minimum as shown when the eta is too small it takes more epochs to reach the local minimum.** | | |

Different epochs:

|  |  |  |
| --- | --- | --- |
| Hyperparameters | | |
| **F1 = 4, F2 = 5, C1 = 2, C2 = 3, eta = 0.01, b = 0** | | |
| Epochs | | |
| **M = 10** | **M = 100** | **M = 1000** |
| Visualization | | |
|  |  |  |
|  |  |  |
| Analysis | | |
| **Changing the number of epochs affects the accuracy if it’s too small the model may need more epochs to be well-trained.** | | |

Different bias:

|  |  |
| --- | --- |
| Hyperparameters | |
| **F1 = 1, F2 = 5, C1 = 1, C2 = 2, eta = 0.01, m = 1000** | |
| Bias | |
| **B = 0 B = 1** | |
| Visualization | |
|  |  |
|  |  |
| Analysis | |
| **Adding bias to the model affects the accuracy. Sometime the model train better with bias.** | |

**Adaline:**

Different Features:

|  |  |  |
| --- | --- | --- |
| Hyperparameters | | |
| **C1 = 1, C2 = 2, eta = 0.1, m = 1000, b = 0, MSE = 0.01** | | |
| Features | | |
| **F1 = 1, F2 = 2** | **F1 = 2, F2 = 4** | **F1 = 3, F2 = 5** |
| Visualization | | |
|  |  |  |
|  |  |  |
| Analysis | | |
| **Adaline algorithm managed to discriminate between different features all in one epoch with an accuracy of 100% as all of them are linearly separable.** | | |

Different Classes:

|  |  |  |
| --- | --- | --- |
| Hyperparameters | | |
| **F1 = 2, F2 = 3, eta = 0.1, m = 1000, b = 0, MSE = 0.01** | | |
| Classes | | |
| **C1 = 1, C2 = 2** | **C1 = 2, C2 = 3** | **C1 = 1, C2 = 3** |
| Visualization | | |
|  |  |  |
|  |  |  |
| Analysis | | |
| **The second case shows that the accuracy is 97.5% and it took all the epochs which means the model needs more epochs to learn or these two classes aren’t linearly separable.** | | |

Different eta:

|  |  |  |
| --- | --- | --- |
| Hyperparameters | | |
| **F1 = 1, F2 = 4, C1 = 1, C2 = 3, m = 1000, b = 0, MSE = 0.01** | | |
| Eta | | |
| **Eta = 0.1** | **Eta = 0.01** | **Eta = 0.00001** |
| Visualization | | |
|  |  |  |
|  |  |  |
| Analysis | | |
| **Changing the learning rate affects how fast the model reaches the local minimum as shown when the eta is too small it takes more epochs to reach the local minimum.** | | |

Different epochs:

|  |  |  |
| --- | --- | --- |
| Hyperparameters | | |
| **F1 = 4, F2 = 5, C1 = 2, C2 = 3, eta = 0.01, b = 0, MSE = 0.01** | | |
| Epochs | | |
| **M = 10** | **M = 100** | **M = 1000** |
| Visualization | | |
|  |  |  |
|  |  |  |
| Analysis | | |
| **Every time the model takes all the epochs because of 2 outliers from the given classes or they aren’t linearly separable.** | | |

Different bias:

|  |  |
| --- | --- |
| Hyperparameters | |
| **F1 = 1, F2 = 5, C1 = 1, C2 = 2, eta = 0.01, m = 1000, MSE = 0.01** | |
| Bias | |
| **B = 0 B = 1** | |
| Visualization | |
|  |  |
|  |  |
| Analysis | |
| **Adding bias to the model affects the accuracy. Sometimes the model trains better with bias, in this case it didn’t affect the accuracy but it took less epochs (faster convergence).** | |

Different MSE:

|  |  |  |
| --- | --- | --- |
| Hyperparameters | | |
| **F1 = 1, F2 = 4, C1 = 1, C2 = 3, m = 1000, b = 0, eta = 0.1** | | |
| MSE | | |
| **MSE= 0.1** | **MSE = 0.01** | **MSE = 0.00001** |
| Visualization | | |
|  |  |  |
|  |  |  |
| Analysis | | |
| **Achieving lower MSE requires more training epochs.** | | |

**Conclusion**

In general, Adaline is an improvement as it solves the problem of cost function by using Mean Square Error while Perceptron doesn’t have a cost function to stop and might lead to an infinite loop in case of non-linearly separable data because it stops only in case of zero error.