Report File

Assignment – 1

Name: Shivani Panchiwala

UTA ID: 1001982478

Subject: CSE 6363 – 001

Prof.: Alex Dillhof

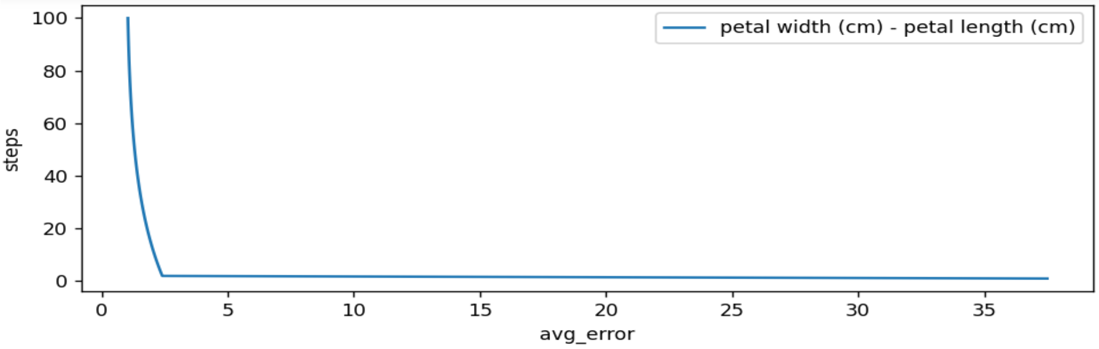
The Iris flower data set (https://en.wikipedia.org/wiki/Iris_flower_data_set)

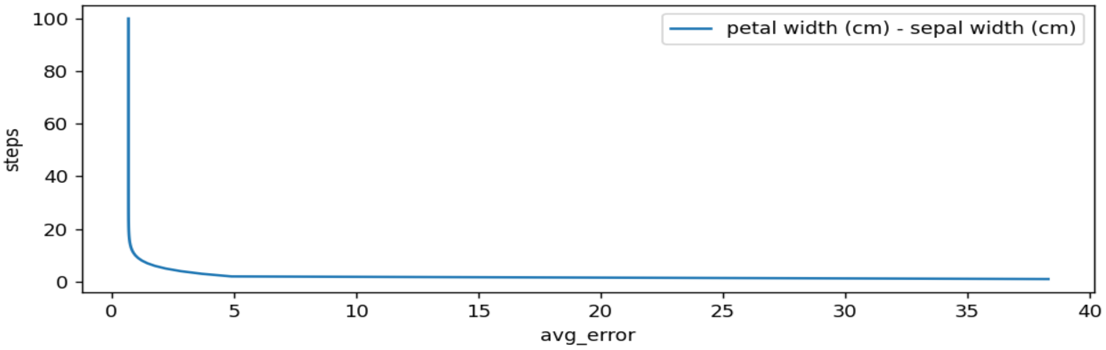
Linear_Regression:

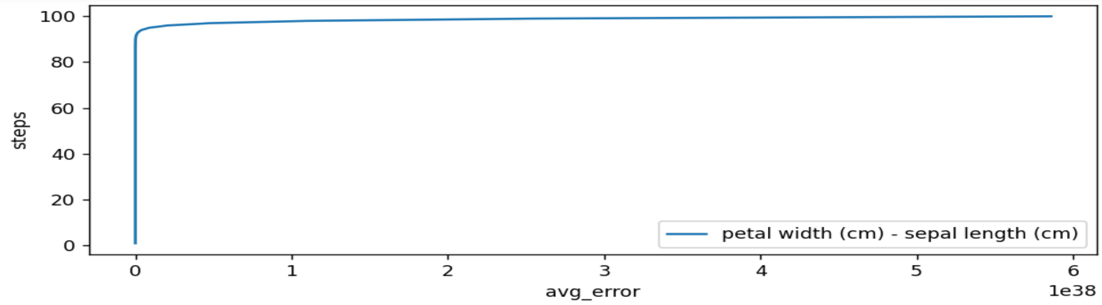
```
Mean Absolute Error: 0.2112698081154823
Mean Squared Error: 0.061480291624700444
Mean Root Squared Error: 0.24795219624899564
```

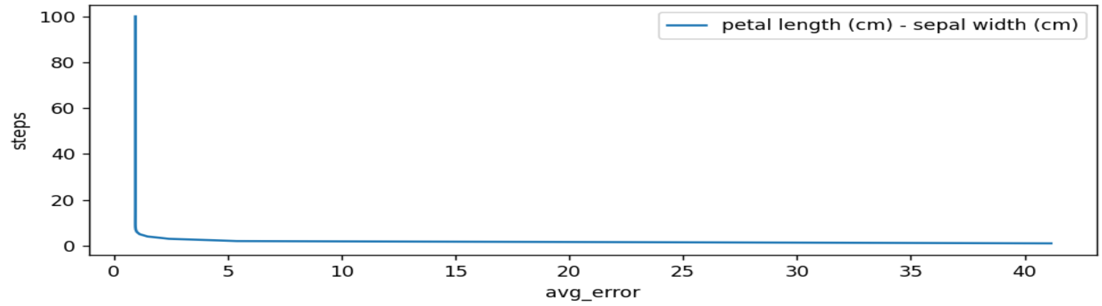
Linear_Regression_Training_Plot:

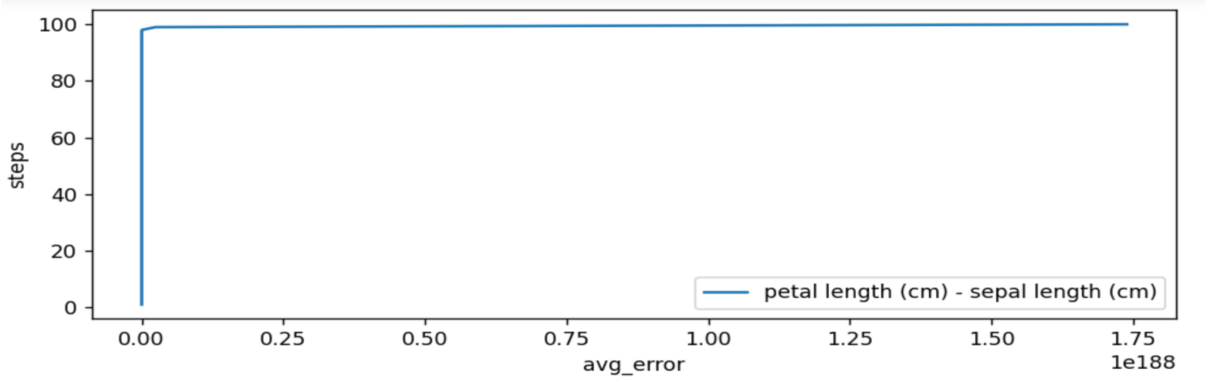
The training data set is split into batches of 32 each. Twelve different linear models are trained through these batches. Below are the plots for all twelve models (errors Vs count).

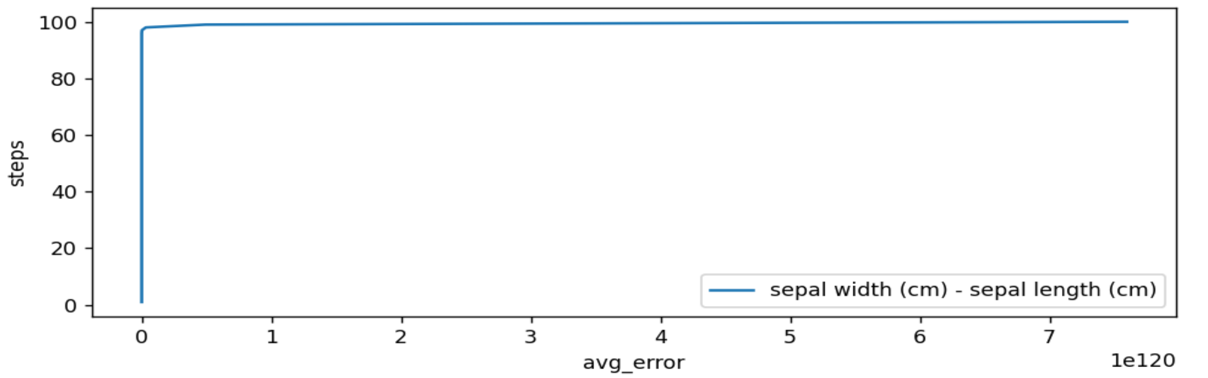


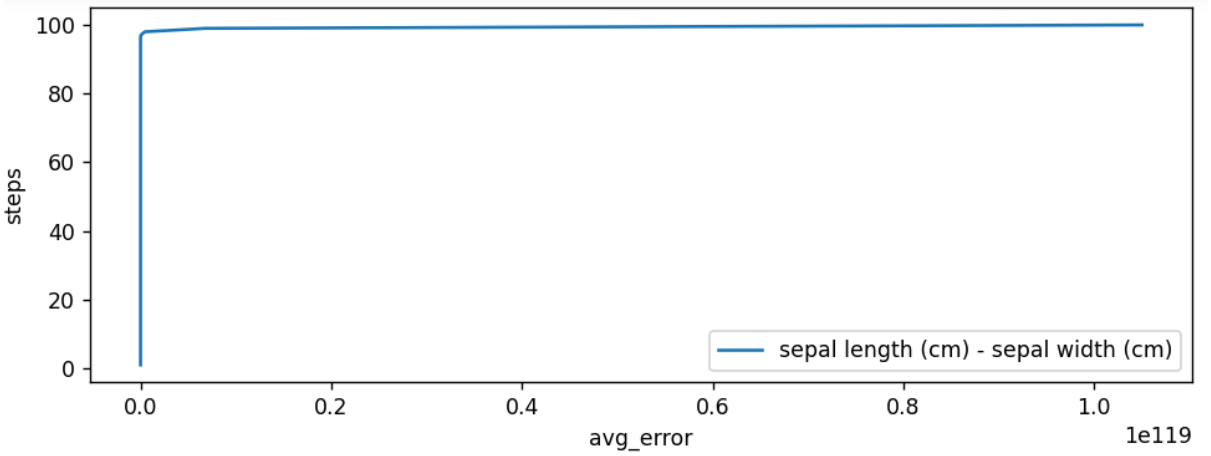


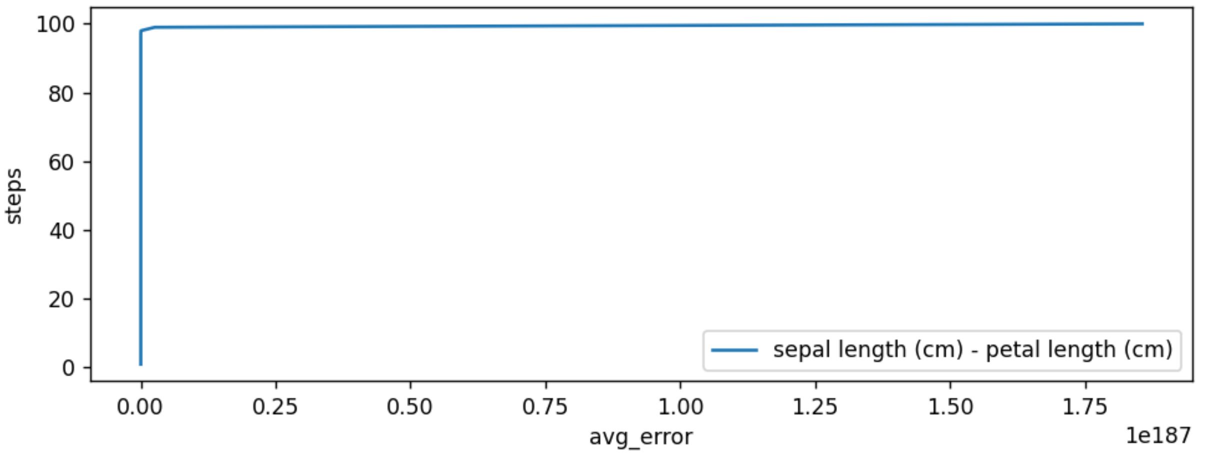


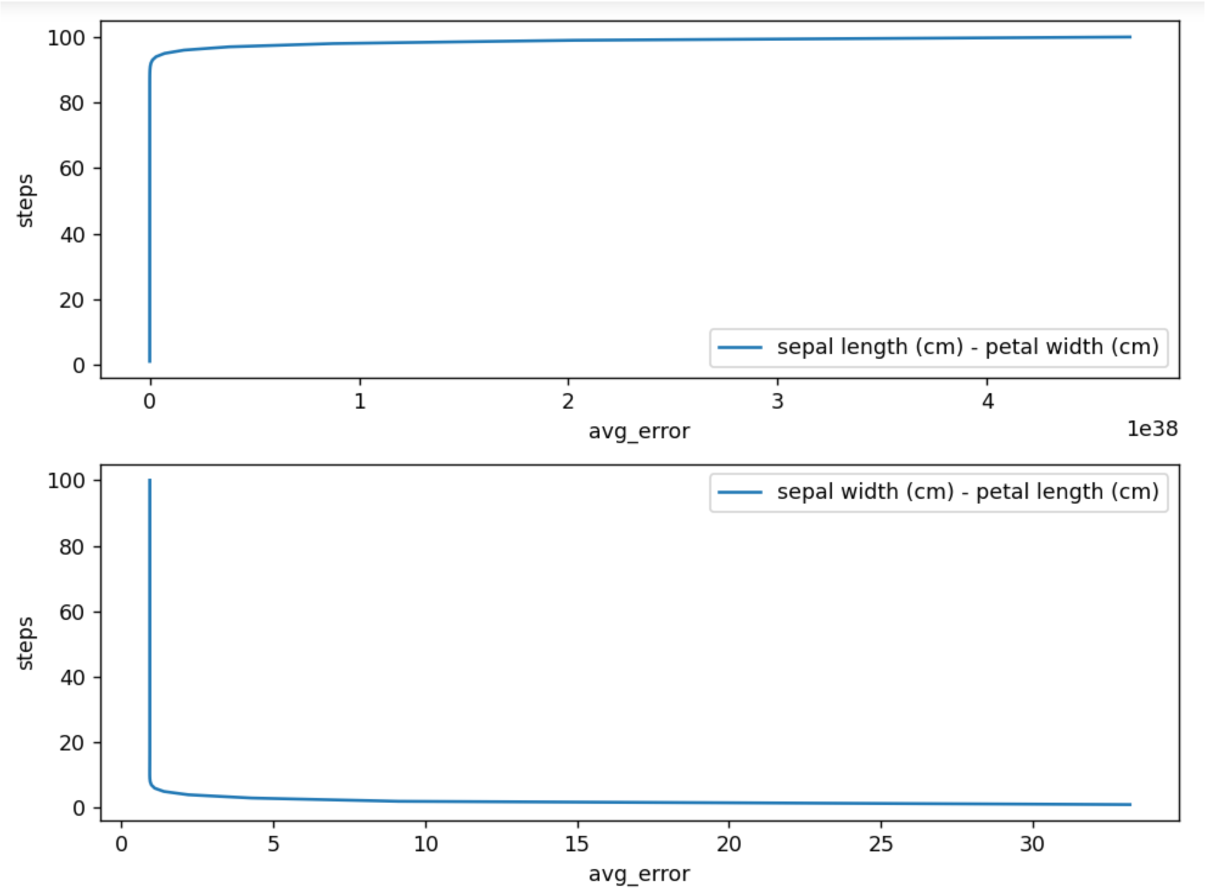


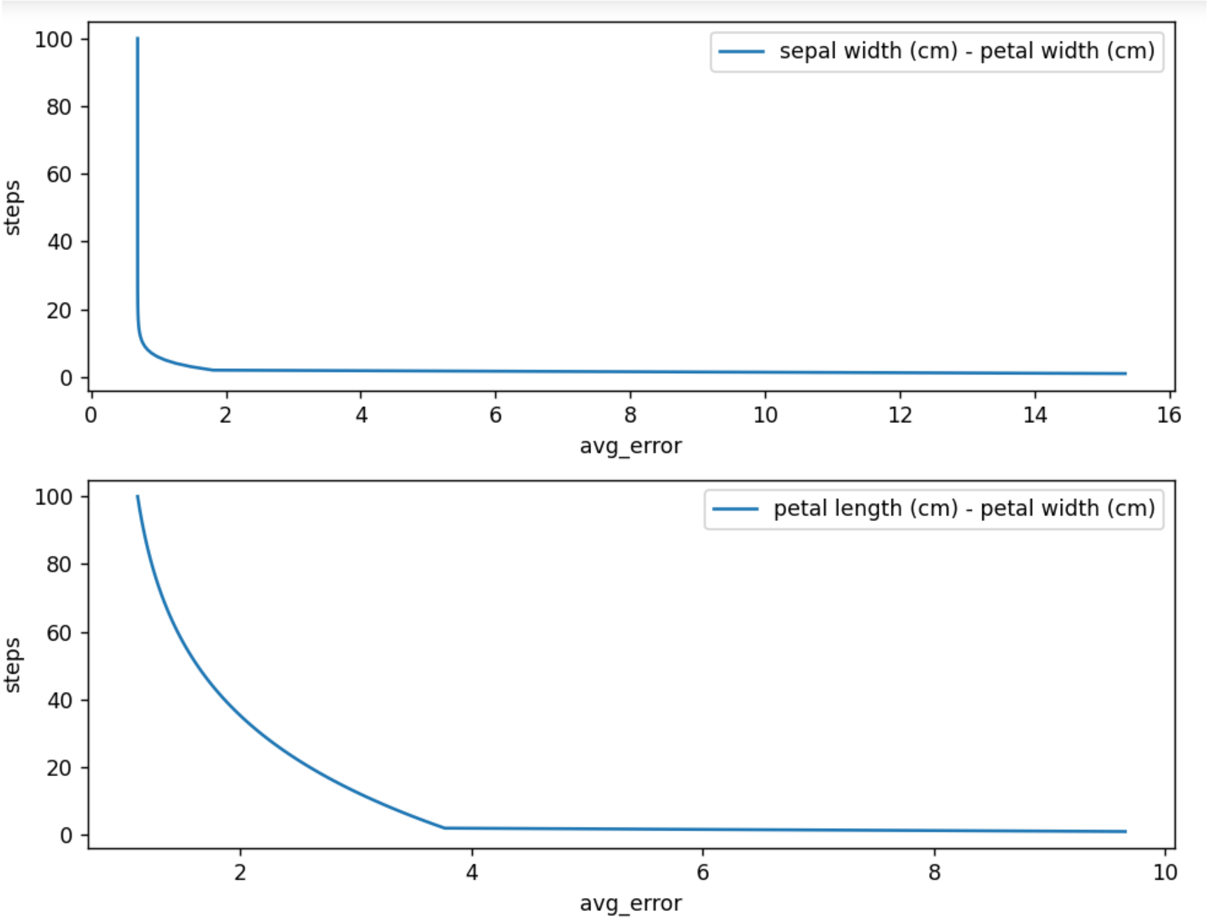

steps
100
80
60
40
20
0
petal length (cm) - sepal length (cm)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
avg_error
1e188


steps
100
80
60
40
20
0
sepal width (cm) - sepal length (cm)
0
1
2
3
4
5
6
7
avg_error
1e120


steps
100
80
60
40
20
0
sepal length (cm) - sepal width (cm)
0.0
0.2
0.4
0.6
0.8
1.0
avg_error
1e119


steps
100
80
60
40
20
0
sepal length (cm) - petal length (cm)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
avg_error
1e187


sepal length (cm) - petal width (cm)
steps
0
20
40
60
80
100
0
1
2
3
4
avg_error
1e38
sepal width (cm) - petal length (cm)
steps
0
20
40
60
80
100
0
5
10
15
20
25
30
avg_error


sepal width (cm) - petal width (cm)
steps
0
20
40
60
80
100
0
2
4
6
8
10
12
14
16
avg_error
petal length (cm) - petal width (cm)
steps
0
20
40
60
80
100
2
4
6
8
10
avg_error

Testing output for linear regression:

Model	Mean Square Error
petal width (cm) - petal length (cm)	0.6598947541848104
petal width (cm) - sepal width (cm)	0.6513834504778802
petal width (cm) - sepal length (cm)	3.445981819678912e+38
petal length (cm) - sepal width (cm)	0.5331846195224962
petal length (cm) - sepal length (cm)	1.430325304763975e+188
sepal width (cm) - sepal length (cm)	4.7133499587031526e+120
sepal length (cm) - sepal width (cm)	6.515512196690499e+118
sepal length (cm) - petal length (cm)	1.5269938906242933e+187
sepal length (cm) - petal width (cm)	2.753666489012849e+38
sepal width (cm) - petal length (cm)	0.5331846195224962
sepal width (cm) - petal width (cm)	0.651383461548725
petal length (cm) - petal width (cm)	0.6307815293491975

L2 regularization in Linear regression training:

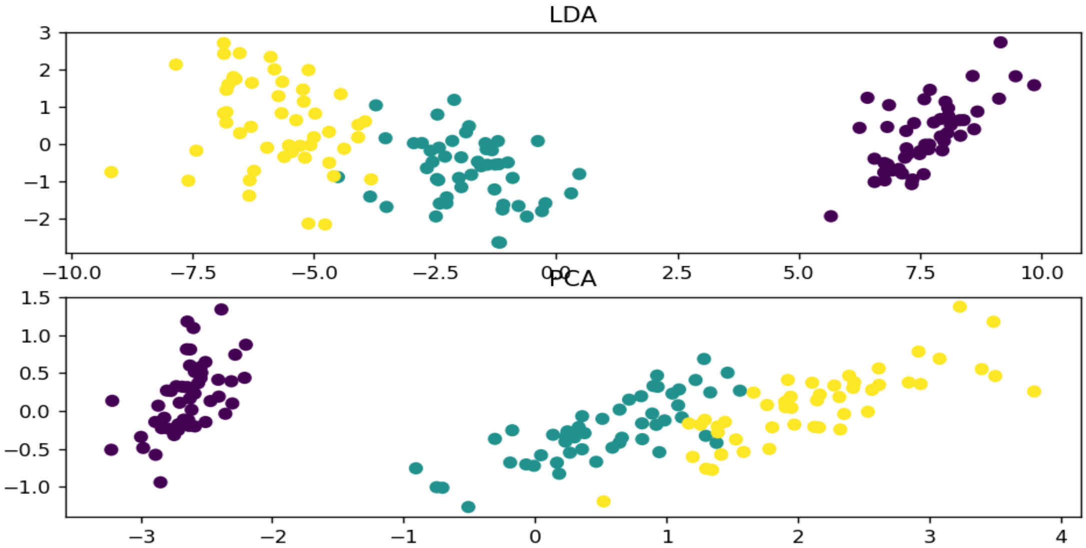
Below are the outputs when a specific model was trained with L2-regularized and non-regularized weights:

```
Weight with regularization : [0.36919318 0.17167277] and error : 54.094
55229689472
Weight without regularization : [ 1.38608699 -0.13016192] and error : 5
4.05265734237033
```

Classification:

1) Linear Discriminant Analysis:

LDA vs PCA Plotting the graph:

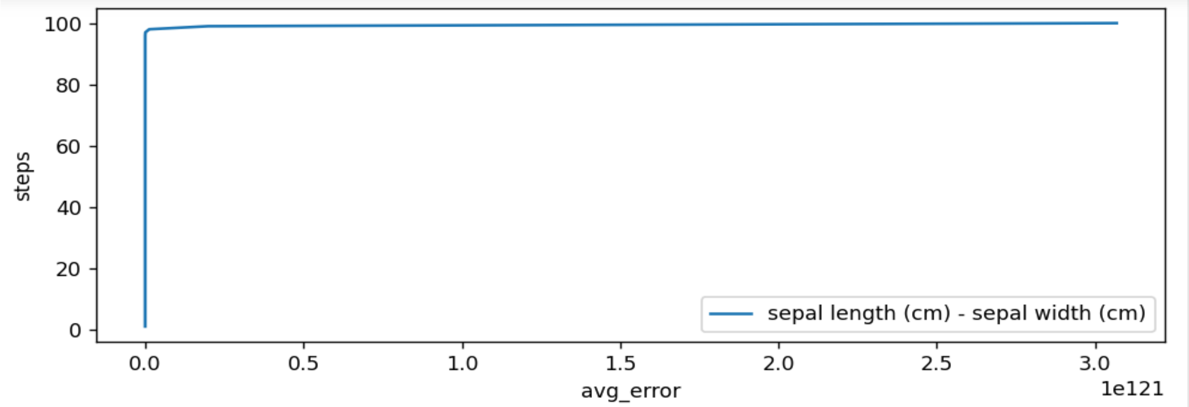


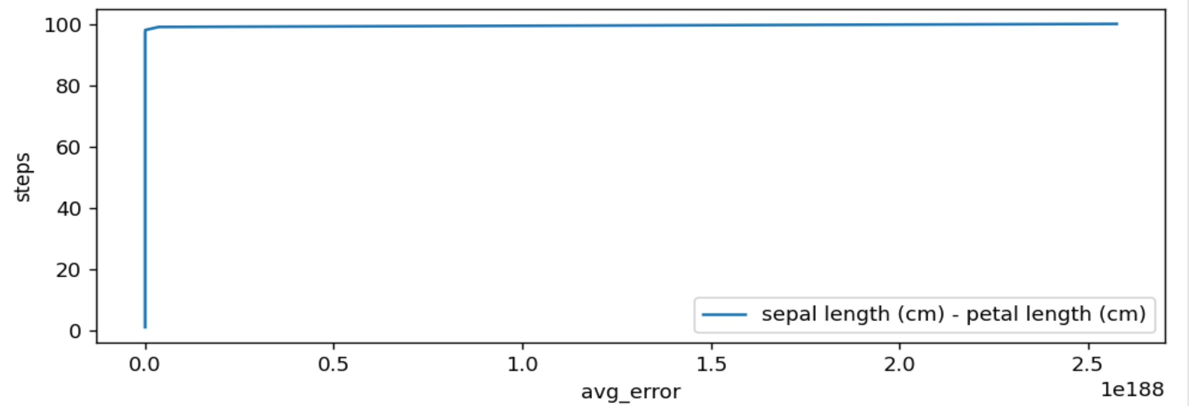
Accuracy:

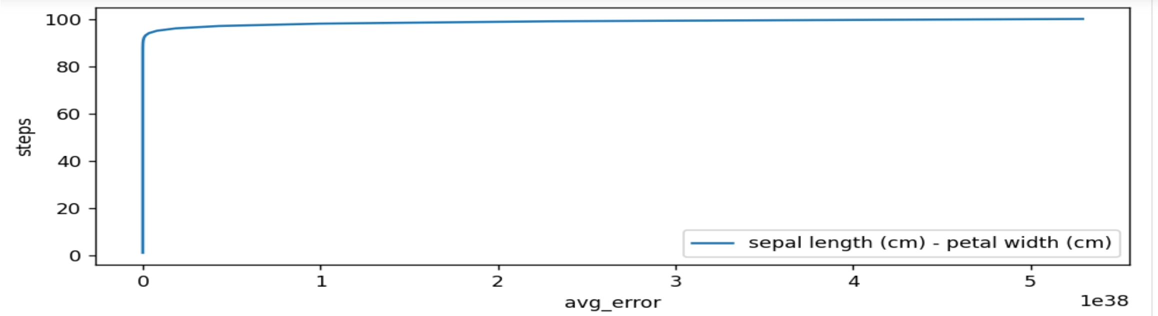
```
  precision    recall  f1-score   support

           0       1.00      1.00      1.00         6
           1       1.00      1.00      1.00         5
           2       1.00      1.00      1.00         4

    accuracy                           1.00        15
   macro avg       1.00      1.00      1.00        15
weighted avg       1.00      1.00      1.00        15

Confusion matrix for LDA
 [[6 0 0]
 [0 5 0]
 [0 0 4]]
accuracy_LDA: 1.000
precision_LDA: 1.000
recall_LDA : 1.000
f1-score_LDA : 1.000


Mean Absolute Error: 0.0
Mean Squared Error: 0.0
Mean Root Squared Error: 0.0
```

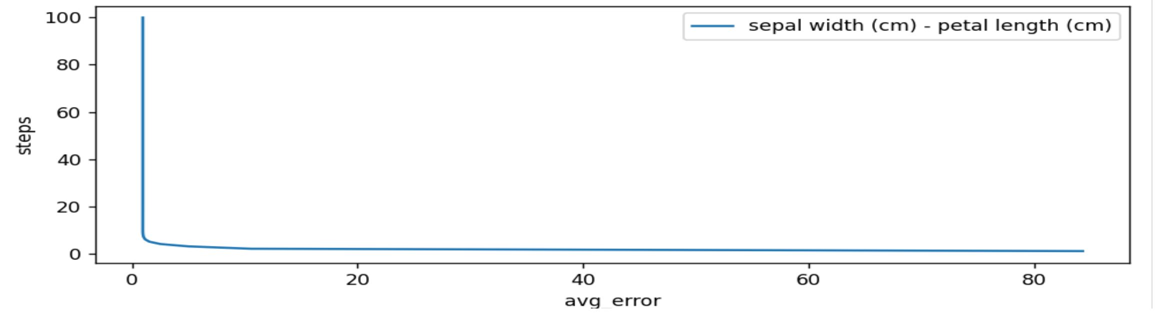
LDA Training_Plot:

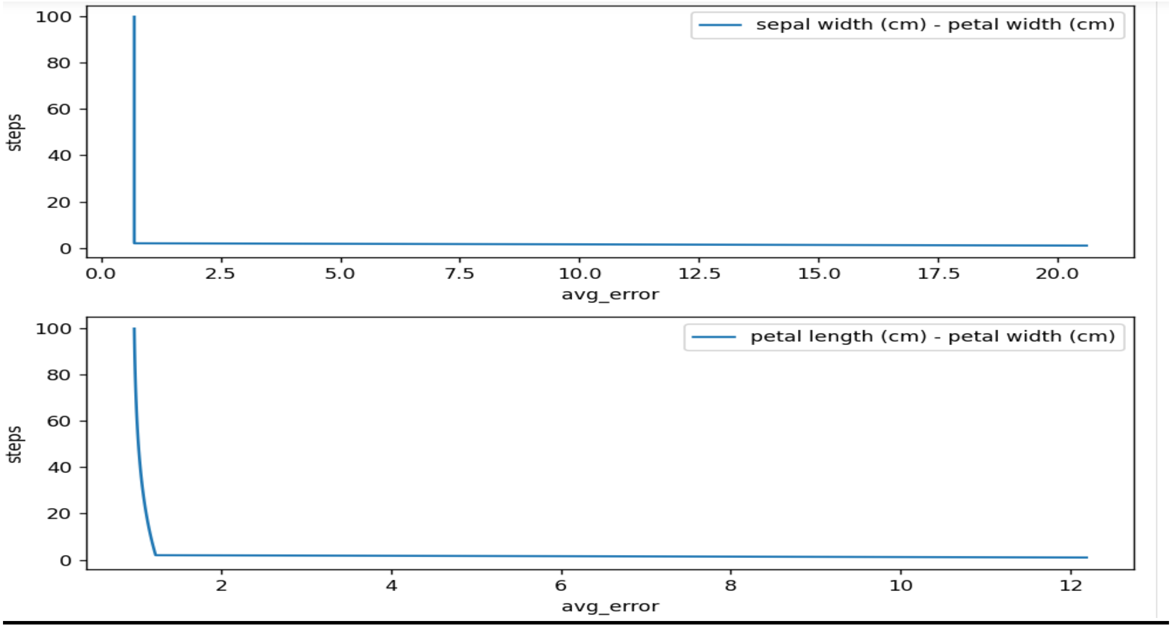
The training data set is split into batches of 32 each. Six different linear models are trained through these batches. Below are the plots for all six models (errors Vs count).


steps
100
80
60
40
20
0
sepal length (cm) - sepal width (cm)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
avg_error
1e121


steps
100
80
60
40
20
0
sepal length (cm) - petal length (cm)
0.0
0.5
1.0
1.5
2.0
2.5
avg_error
1e188


steps
100
80
60
40
20
0
sepal length (cm) - petal width (cm)
0
1
2
3
4
5
avg_error
1e38


steps
100
80
60
40
20
0
sepal width (cm) - petal length (cm)
0
20
40
60
80
avg_error

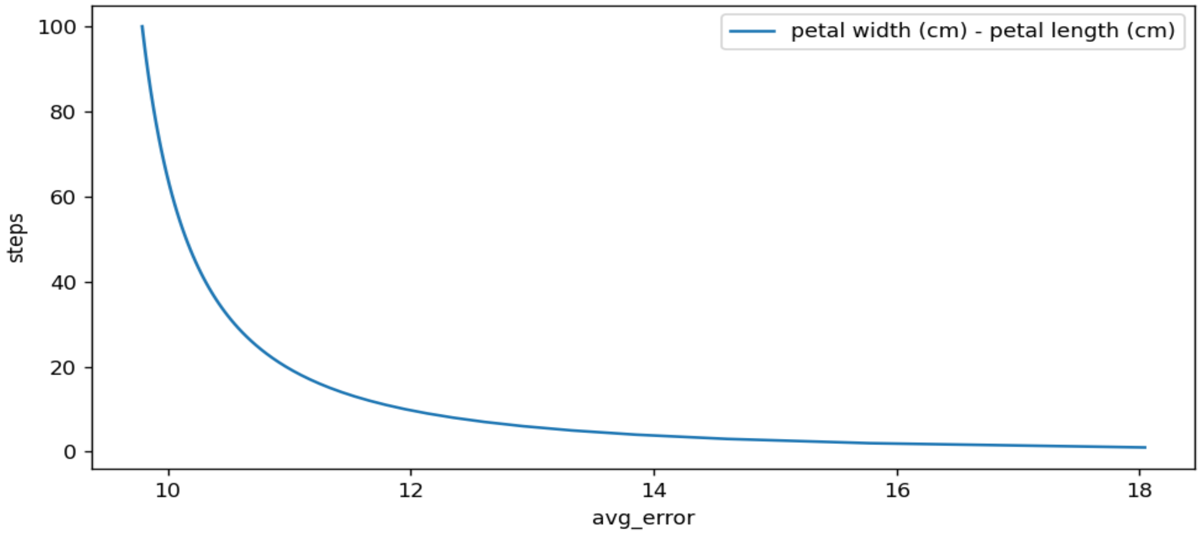


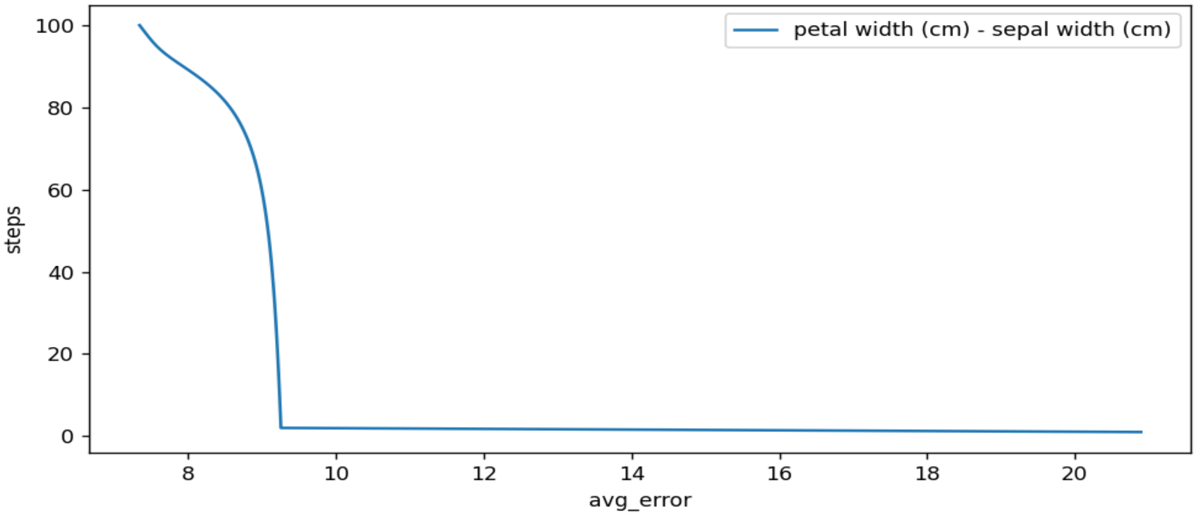
Testing output for LDA:

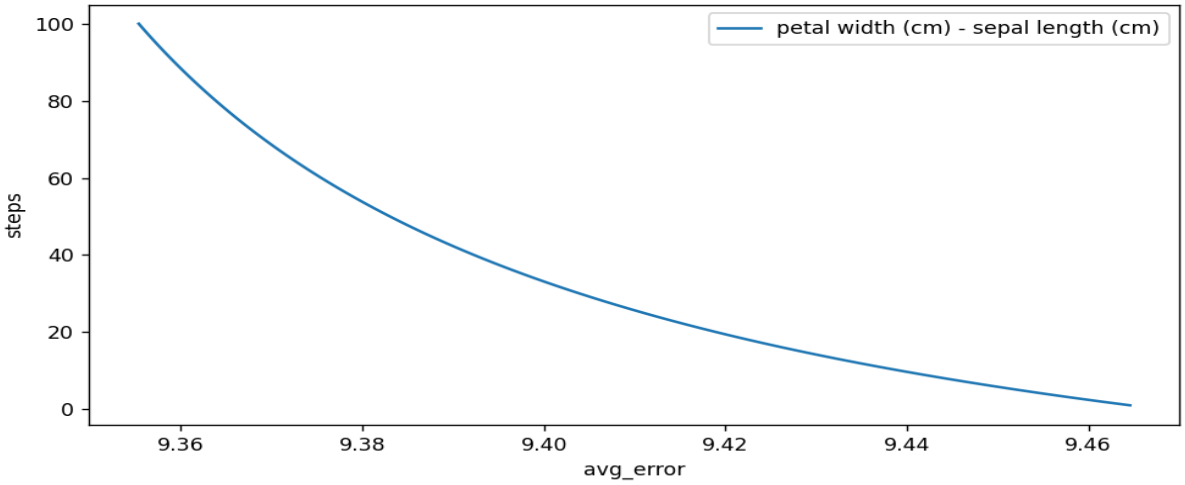
Model	Mean Square Error
sepal length (cm) - sepal width (cm)	1.903377955164063e+121
sepal length (cm) - petal length (cm)	2.1189393264193424e+188
sepal length (cm) - petal width (cm)	3.112995613395727e+38
sepal width (cm) - petal length (cm)	0.5331846195224962
sepal width (cm) - petal width (cm)	0.6513834730456635
petal length (cm) - petal width (cm)	0.7388076337868977

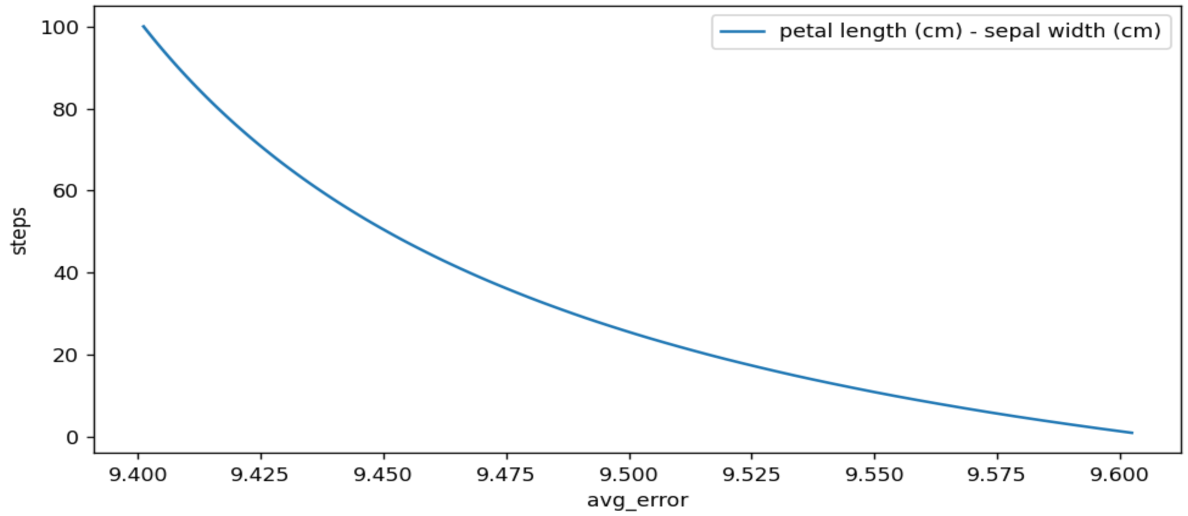
2) Logistic Regression Training Plot:

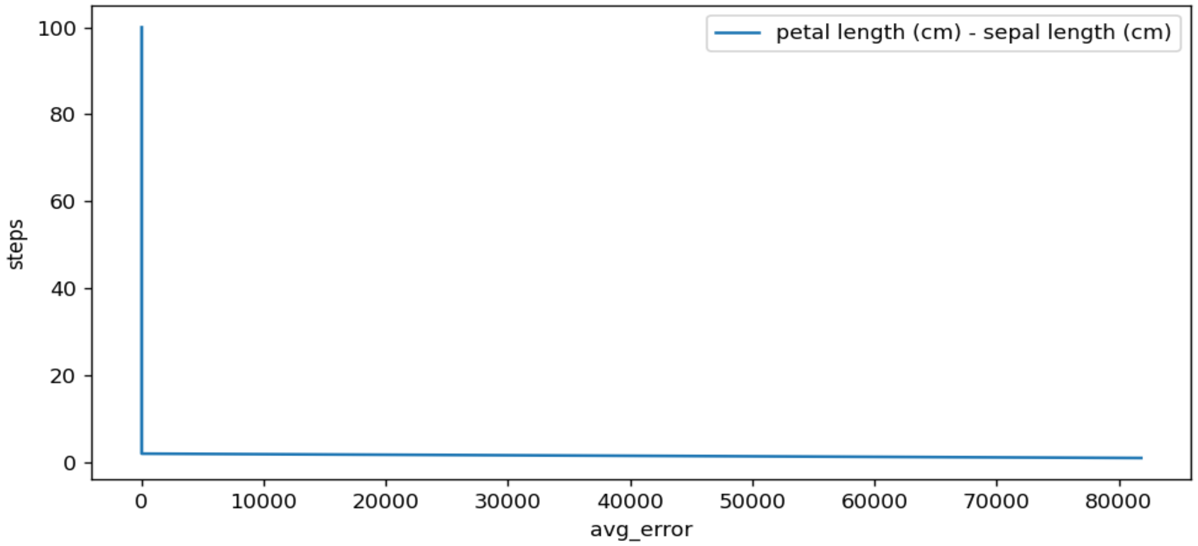
The training data set is split into batches of 32 each. Six different linear models are trained through these batches. Below are the plots for all six models (errors Vs count).

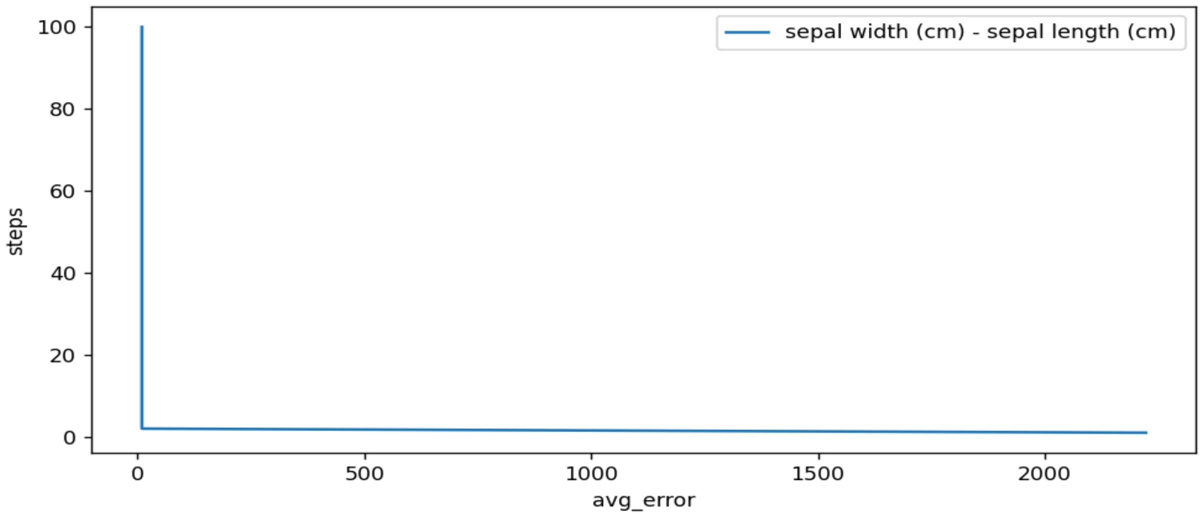

petal width (cm) - petal length (cm)
steps
100
80
60
40
20
0
10
12
14
16
18
avg_error


petal width (cm) - sepal width (cm)
steps
100
80
60
40
20
0
8
10
12
14
16
18
20
avg_error


petal width (cm) - sepal length (cm)
steps
100
80
60
40
20
0
9.36
9.38
9.40
9.42
9.44
9.46
avg_error


petal length (cm) - sepal width (cm)
steps
0
20
40
60
80
100
9.400
9.425
9.450
9.475
9.500
9.525
9.550
9.575
9.600
avg_error


petal length (cm) - sepal length (cm)
steps
0
20
40
60
80
100
0
10000
20000
30000
40000
50000
60000
70000
80000
avg_error


sepal width (cm) - sepal length (cm)
steps
0
20
40
60
80
100
0
500
1000
1500
2000
avg_error

Accuracy:

```
precision    recall  f1-score   support

           0       1.00      1.00      1.00         4
           1       1.00      1.00      1.00         5
           2       1.00      1.00      1.00         6

    accuracy                           1.00        15
   macro avg       1.00      1.00      1.00        15
weighted avg       1.00      1.00      1.00        15

Confusion matrix for Logistic Regression
 [[4 0 0]
 [0 5 0]
 [0 0 6]]
accuracy_Logistic Regression : 1.000
precision_Logistic Regression : 1.000
recall_Logistic Regression: 1.000
f1-score_Logistic Regression : 1.000


Mean Absolute Error: 0.0
Mean Squared Error: 0.0
Mean Root Squared Error: 0.0
```

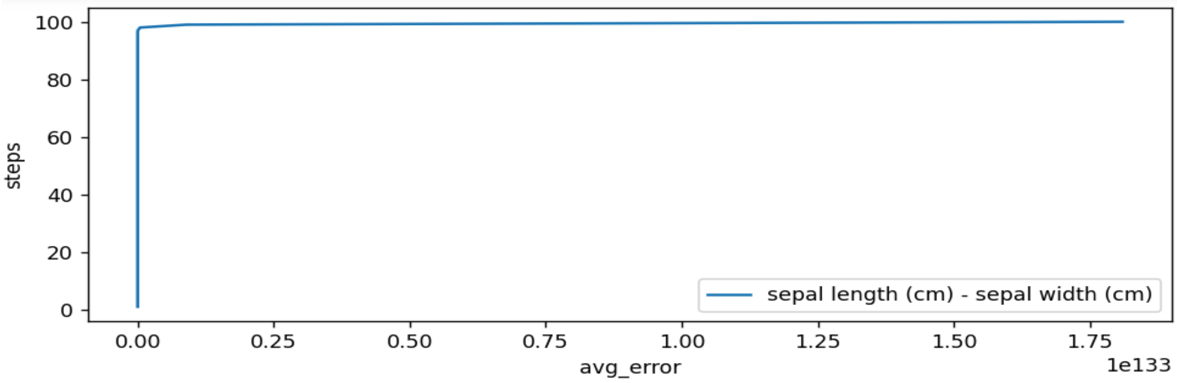
Testing output for Logistic Regression:

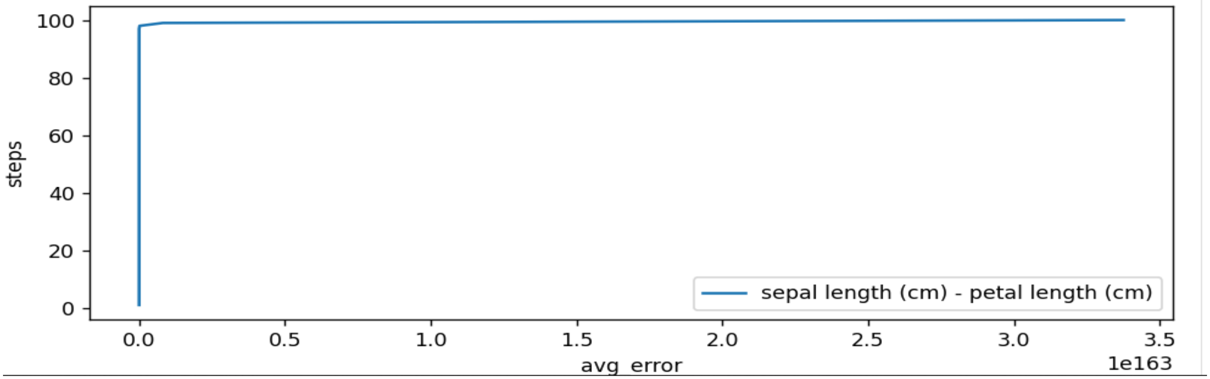
Model	Mean Square Error
petal width (cm) - petal length (cm)	3.6674620490145204
petal width (cm) - sepal width (cm)	2.7653218571175806
petal width (cm) - sepal length (cm)	3.512663580375092
petal length (cm) - sepal width (cm)	3.518767591473467
petal length (cm) - sepal length (cm)	3.5
sepal width (cm) - sepal length (cm)	3.5

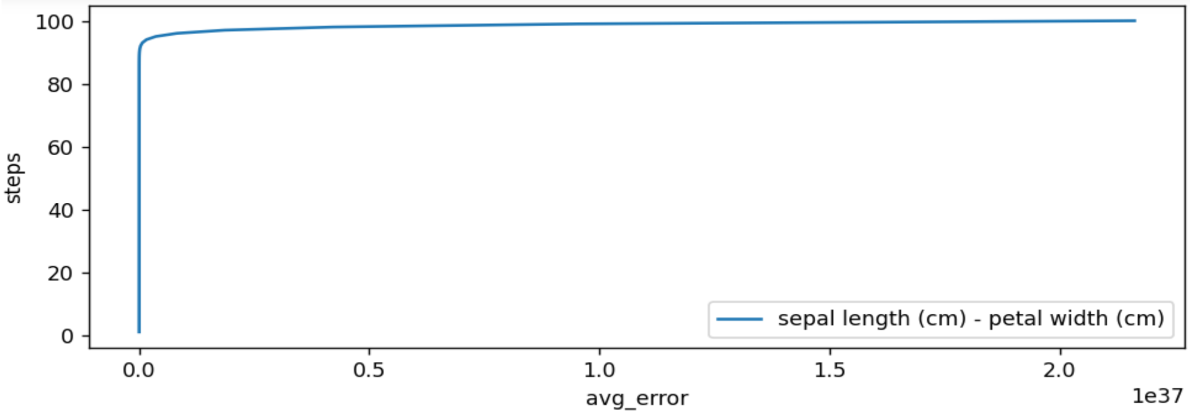
3) Naive Bayes:

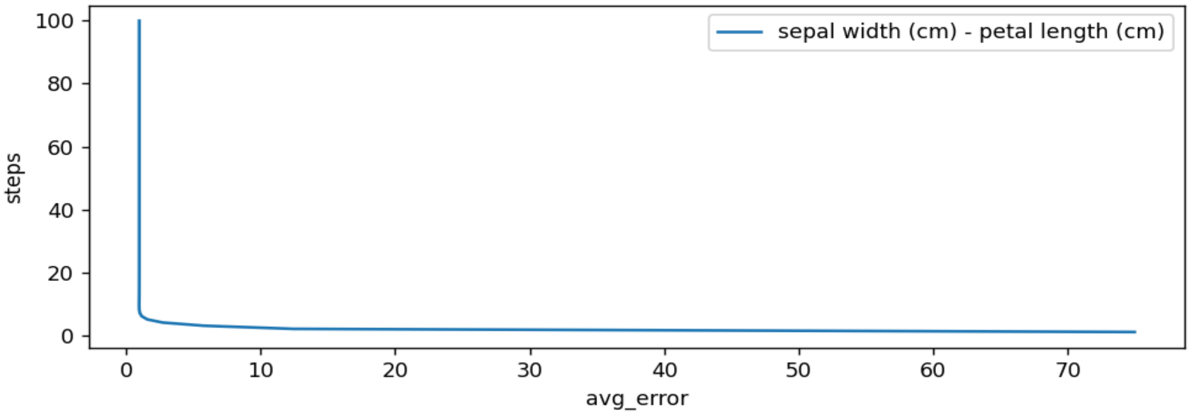
Naive Bayes Training Plot:

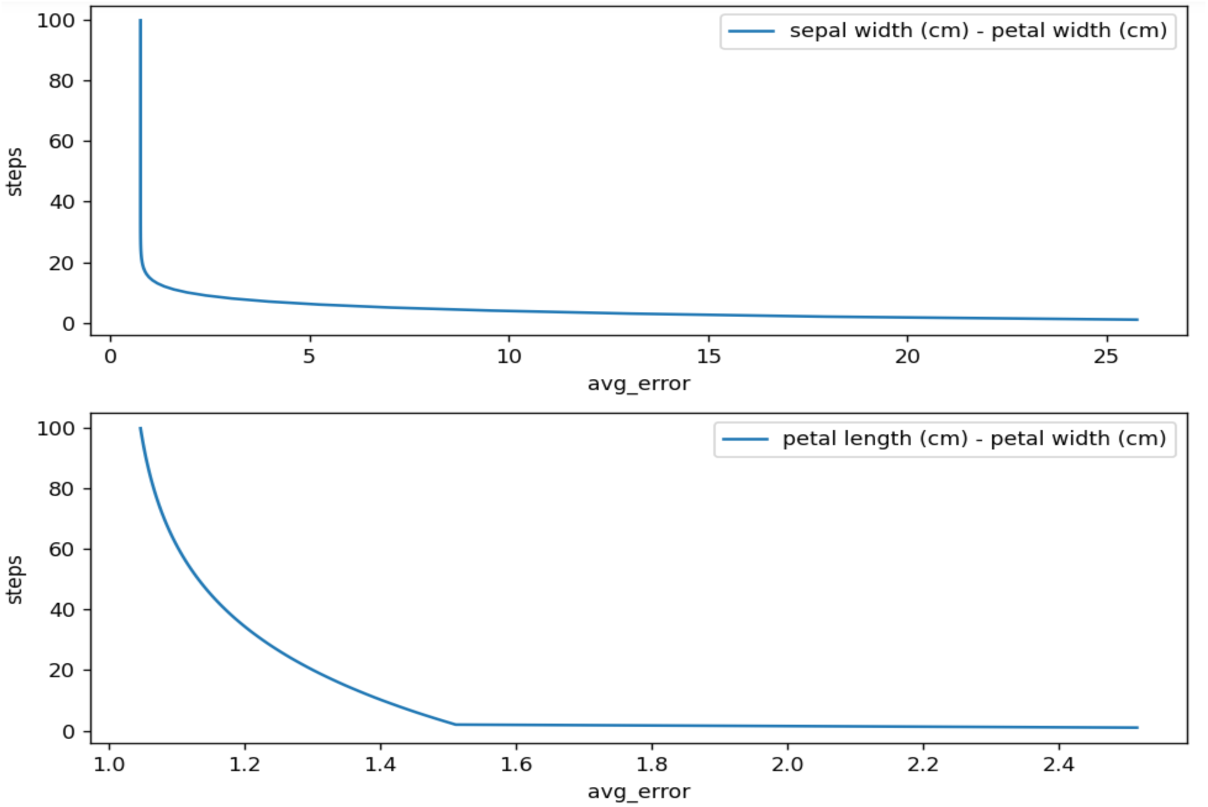
The training data set is split into batches of 32 each. Six different linear models are trained through these batches. Below are the plots for all six models (errors Vs count).


100
80
60
40
20
0
steps
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
avg_error
1e133
sepal length (cm) - sepal width (cm)


100
80
60
40
20
0
steps
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
avg_error
1e163
sepal length (cm) - petal length (cm)


100
80
60
40
20
0
steps
0.0
0.5
1.0
1.5
2.0
avg_error
1e37
sepal length (cm) - petal width (cm)


100
80
60
40
20
0
steps
0
10
20
30
40
50
60
70
avg_error
sepal width (cm) - petal length (cm)



Accuracy:

```
           precision    recall  f1-score   support

           0       1.00      1.00      1.00         3
           1       0.89      1.00      0.94         8
           2       1.00      0.75      0.86         4

    accuracy                           0.93        15
   macro avg       0.96      0.92      0.93        15
weighted avg       0.94      0.93      0.93        15

Confusion matrix for Naive Bayes
 [[3 0 0]
 [0 8 0]
 [0 1 3]]
accuracy_Naive Bayes: 0.933
precision_Naive Bayes: 0.933
recall_Naive Bayes: 0.933
f1-score_Naive Bayes : 0.933



Mean Absolute Error: 0.06666666666666667
Mean Squared Error: 0.06666666666666667
Mean Root Squared Error: 0.2581988897471611
```

Testing output for Naive Bayes:

Model	Mean Square Error
sepal length (cm) - sepal width (cm)	1.1662961729052648e+133
sepal length (cm) - petal length (cm)	1.340318424194045e+163
sepal length (cm) - petal width (cm)	9.940774082215085e+36
sepal width (cm) - petal length (cm)	0.6810111671779693
sepal width (cm) - petal width (cm)	0.30087984678313295
petal length (cm) - petal width (cm)	0.3335154084751074

Classification Accuracy:

Classification	Accuracy
LDA	1.00
Logistic Regression	1.00
Navie Bayes	0.933

Based on Classifier Accuracy Linear Discriminate Analysis and Logistic Regression has 100% accuracy.

References:

https://stackoverflow.com/questions/43159754/datasets-load-iris-in-python

https://stackoverflow.com/questions/464864/how-to-get-all-possible-combinations-of-a-list-s-elements

https://www.codegrepper.com/code-examples/python/X_train%2C+X_test%2C+Y_train%2C+Y_test+%3D+train_test_split%28X%2CY%2Ctest_size%3D0.25%2C+random_state%3D123%29

https://stackoverflow.com/questions/3584805/in-matplotlib-what-does-the-argument-mean-in-fig-add-subplot111

https://stackoverflow.com/questions/66902973/using-matplotlib-to-draw-i-use-ax-set-xlabelx-and-ax-set-ylabely-to-set

https://stackoverflow.com/questions/9603230/how-to-use-matplotlib-tight-layout-with-figure

https://gist.github.com/nuzrub/f1527654572b3e2da5125d0581e7bdad https://www.codegrepper.com/code-examples/python/how+to+calculate+mean+squared+error+in+python

https://realpython.com/gradient-descent-algorithm-python

https://numpy.org/doc/stable/reference/random/generated/numpy.random.uniform.html

https://data-flair.training/blogs/train-test-set-in-python-ml/

https://scikit-learn.org/stable/auto_examples/datasets/plot_iris_dataset.html

https://towardsdatascience.com/classification-basics-walk-through-with-the-iris-data-set-d46b0331bf82

https://hands-on.cloud/implementation-of-linear-discriminant-analysis-lda-using-python/

https://www.statology.org/linear-discriminant-analysis-in-python/

https://medium.com/@kgpvijaybg/logistic-regression-on-iris-dataset-48b2ecdfb6d3

https://medium.com/analytics-vidhya/linear-regression-using-iris-dataset-hello-world-of-machine-learning-b0feecac9cc1

https://sa2253.medium.com/classification-of-iris-dataset-b4310ddf0482

https://hands-on.cloud/implementation-of-linear-discriminant-analysis-lda-using-python/

https://towardsdatascience.com/classification-basics-walk-through-with-the-iris-data-set-d46b0331bf82

https://www.kaggle.com/code/vinayshaw/iris-species-100-accuracy-using-naive-bayes/notebook