

**CSCE4604: Practical deep machine Learning**

**Assignment 1- Report**

|  |  |
| --- | --- |
| **Submitted by:** | **Mohammed Abuelwafa** |
| **Student ID:** | **900172603** |

|  |  |
| --- | --- |
| **Date:** | 3rd October 2020 |

**Fall 2020**

# Part 1:

Chart, line chart

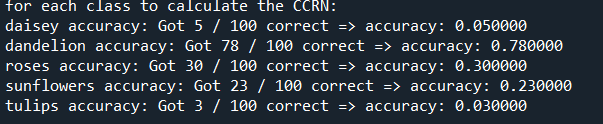
Description automatically generated2) Plot of 5-fold cross validation accuracy vs number of neighbors ‘k’:

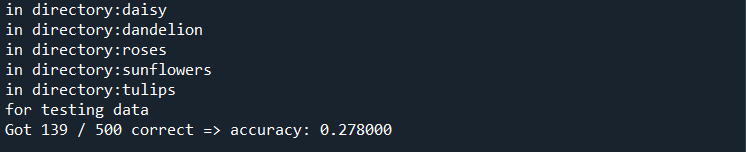
Chart, line chart

Description automatically generated  
Plot of 5-fold cross validation standard deviation vs number of neighbors ‘k’:

From the plots, we may conclude that the best k for this data set is k = 60 since it has the highest average accuracy and the lowest standard deviation. K = 70 is a good choice as well. however, the confidence of k = 60 is higher.

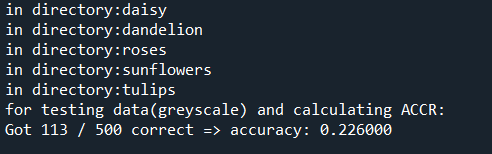
3. calculating CCRn for the classes:



4. Calculating ACCR for the testing dataset:

Got an accuracy of 27.8% for the testing data set.

5.calculating ACCR for the greyscale images used for training and testing:



We got an accuracy of 22.6% for the greyscale images. Hence, training and testing using RGB produces better results.

# Part 2:

b) plotting the root mean square error (Erms) vs the polynomial order:

Chart

Description automatically generated

I experimented with polynomial orders up to 20 (only first 15 are shown in this graph). The following observations are noticed:  
  
a) there is an outlier point at order 9 for the 5th fold that caused the average to increase. This is a result of the high values of w generated by the developed polynomial\_fit function that diverges after an order of 7 resulting in higher values for the model. Hence, higher errors.

b) The trend is generally conserved that the error is getting higher for the validation data and getting lower for the training data.  
  
due to the following observation, the best order according to this graph is order 8 due to having the lowest Erms error without overfitting.

3)

Chart, line chart

Description automatically generated

After trying different combinations, it appears that the 15th combination (1111) is the best combination since it has the lowest error. That combination is corresponding to columns (0,1,2,3) that corresponds to (AT-V-AP-RH) combination is the best.

Chart, line chart

Description automatically generated4)

From the plot, one can notice that order 6 produces the minimum error for the 4 input features plugged in the model. Moreover, at order 9 we notice the sign of overfitting.