Neeraj Vashistha - 190573735

School of Electronic Engineering and Computer Science

ECS797

Machine Learning for Visual Data Analysis

Lab 3: Age Estimation by Regression

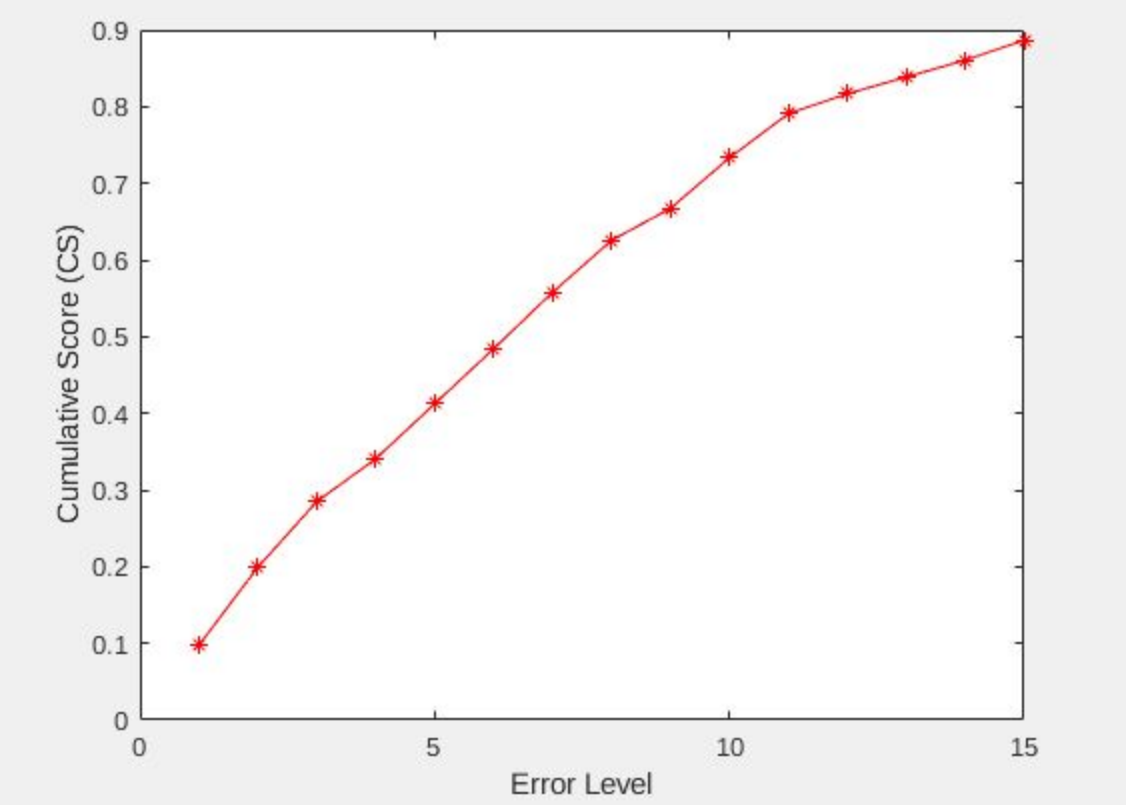
1. Read in the training and test data by loading the .mat file.
2. Call the Matlab built-in function regress(), which takes the training features and labels as input and learns a linear regression model. Read the Matlab document to understand which linear regression model is implemented.
3. Read in the test data, and apply the learned linear regression model to estimate the age for each test data point.
4. Compute the MAE and CS value (with a cumulative error level of 5) by comparing the estimated ages with the ground truth ages. You need to write your own code here.

```
%% Compute the MAE and CS value (with cumulative error level of 5) for linear regression
abs_error = abs(yhat_test-ytest); % calculate absolute error
mae = sum(abs_error)/size(ytest, 1); % use absolute error to calculate MAE
i = 5;
cum_sum = sum(abs_error < i > 0)/size(ytest, 1); % use absolute error to calculate cumulative error

fprintf('MAE Linear regression = %f\n', mae);
fprintf('Cumulative error of level 5 = %f\n', cum_sum);
```

```
MAE Linear regression = 7.704359
Cumulative error of level 5 = 0.412351
fx >>
```

5. Vary the cumulative error level from 1 to 15 and generate a plot of the CS value against the cumulative error level. You need to write your own code here. See the lecture slides for week 6 for an example of the plot.

```
%% generate a cumulative score (CS) vs. error level plot by varying the error level from 1 to 15.
for i = 1:15
    cum_sum(i) = sum(abs_error < i > 0)/size(ytest, 1);
end

plot (cum_sum, 'r-*')
xlabel('Error Level')
ylabel('Cumulative Score (CS)')
```



6. Compute the MAE and CS values (with cumulative error level of 5) for both the partial least square regression model and the regression tree model by using the Matlab built-in functions. You need to write your own code here.

```
% partial least square regression
[XL,yl,XS,YS,beta,PCTVAR,MSE,stats] = plsregress(xtrain, ytrain, 10); % training

yhat_test_plsr = [ones(size(xtest,1),1) xtest]*beta; % testing

abs_error_plsr = abs(yhat_test_plsr-ytest); % calculate absolute error
mae_plsr = sum(abs_error_plsr)/size(ytest, 1); % use absolute error to calculate MAE
i=5;
cum_sum_plsr = sum(abs_error_plsr < i > 0)/size(ytest, 1); % use absolute error to calculate cumulative

fprintf('MAE Partial Least Square Regression = %f\n', mae_plsr);
fprintf('Cumulative error of level 5 = %f\n', cum_sum_plsr);
```

```
% regression tree model
tree = fitrtree(xtrain, ytrain); % training
yhat_test_rtree = predict(tree, xtest); % testing

abs_error_rtree = abs(yhat_test_rtree-ytest); % calculate absolute error
mae_rtree = sum(abs_error_rtree)/size(ytest, 1); % use absolute error to calculate MAE
i=5;
cum_sum_rtree = sum(abs_error_rtree < i > 0)/size(ytest, 1); % use absolute error to calculate cumulative

fprintf('MAE Regression Tree = %f\n', mae_rtree);
fprintf('Cumulative error of level 5 = %f\n', cum_sum_rtree);
```

```
MAE Partial Least Square Regression = 6.070294
Cumulative error of level 5 = 0.525896
MAE Regression Tree = 8.235005
Cumulative error of level 5 = 0.496016
fx >>
```

7. Compute the MAE and CS values (with cumulative error level of 5) for Support Vector Regression by using the LIBSVM toolbox (http://www.csie.ntu.edu.tw/~cjlin/libsvm/). This step is worth 15% of the total mark for this lab.

```
%% Compute the MAE and CS value (with cumulative error level of 5) for Support Vector Regression by us
addpath(genpath('./softwares'));

st = svmtrain(ytrain, xtrain, '-s 3 -t 0'); % training
yhat_test_svm = svmpredict(ytest, xtest, st); % testing

abs_error_svr = abs(yhat_test_svr-ytest); % calculate absolute error
mae_svr = sum(abs_error_svr)/size(ytest, 1); % use absolute error to calculate MAE
i=5;
cum_sum_svr = sum(abs_error_svr < i > 0)/ size(ytest, 1); % use absolute error to calculate cumulative

fprintf('MAE SVR = %f\n', mae_svr);
fprintf('Cumulative error of level 5 = %f\n', cum_sum_svr);
```

```
MAE SVR = 5.731360
Cumulative error of level 5 = 0.537849
fx >>
```

In particular, you should include the CS plot for linear regression, and a table to compare the four different regression models using both MAE and CS **with an error level of 5.**

Model	**Mean Absolute Error**	**Cumulative Score**
Linear Regression	7.704359	0.412351
Partial Least Square Regression	6.070294	0.525896
Regression Tree	8.235005	0.496016
Support Vector Regression	5.731360	0.537849