Text Mining Final-Project by Thomas Wagner, Alexander Allen, MingYi Wang

Table of Contents

May 14 2015

Input Data

```
clear all
close all

% load 10% sample data and the test data
load DDISample.mat

% convert to normal format instead of sparse
Classp_train=full(Classp_train);
Classm_train=full(Classm_train);
Classp_test=full(Classp_test);
Classm_test=full(Classm_test);
```

PCA on data

```
Train = [Classp_train;Classm_train];
Test = [Classp_test;Classm_test];

[ptrain_m,ptrain_n]=size(Classp_train);
[mtrain_m,mtrain_n]=size(Classm_train);
[ptest_m,ptest_n]=size(Classp_test);
[mtest_m,mtest_n]=size(Classm_test);

train_mean = (1/ptrain_m+mtrain_m)*(ones(1,ptrain_m+mtrain_m)*Train);

Train = Train - ones(ptrain_m+mtrain_m,1)*train_mean;
Test = Test - ones(ptest_m+mtest_m,1)*train_mean;


YTrain = [ones(ptrain_m,1);-ones(mtrain_m,1)];
YTest = [ones(ptest_m,1);-ones(mtest_m,1)];

[eigenvectors, scores, eigenvalues] = pca(Train);
```

Get training scores and testing scores

```
Train_scores = Train*eigenvectors;
Test_scores = Test*eigenvectors;
```

KNN classifier

```
classifier=knnsearch(Train_scores,Test_scores);
total_error=0;
[s,z]=size(Test)
```

```
s =

        6238


z =

        6254
```

Calculating testing error

```
perror=0;
for i=1:ptest_m,
    if(YTest(i)~=YTrain(classifier(i)))
        perror=perror+1;
    end
end

merror=0;
for i=ptest_m+1:s,
    if(YTest(i)~=YTrain(classifier(i)))
        merror=merror+1;
    end
end
merror
perror
total_error = merror+perror
error_percent = total_error/s


%Result 13.58% testing error
```

```
merror =

   295


perror =
```

```
   552

total_error =

   847

error_percent =

    0.1358
```

Published with MATLAB® R2013a