



Classification (KNN) using Clusters of PCA matrix

```
In [249]: #Creating training and testing data
from sklearn import neighbors
from sklearn.model_selection import train_test_split

train, test, target_train, target_test = train_test_split(DTtrans, labels, test_size=
print(test.shape)

(134, 10)

In [250]: n_neighbors = 5

knnclf = neighbors.KNeighborsClassifier(n_neighbors, weights = 'distance')
knnclf.fit(train, target_train)

<ipython-input-250-2413ef3590fb>:14: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, 1), for example using ravel().
knnclf.fit(train, target_train)

Out[250]: KNeighborsClassifier(weights='distance')

In [251]: knnpreds_test = knnclf.predict(test)

In [252]: print(classification_report(target_test, knnpreds_test))

              precision    recall  f1-score   support

0               1.00        1.00        1.00         3
1               1.00        1.00        1.00         4
2               1.00        1.00        1.00         3
3               1.00        1.00        1.00         1
4               1.00        1.00        1.00        123

 accuracy          1.00          1.00          1.00        134
 macro avg          1.00          1.00          1.00        134
 weighted avg          1.00          1.00          1.00        134

In [253]: print(knnclf.score(test, target_test))

1.0

In [254]: print(knnclf.score(train, target_train))

1.0
```

Again, we are obtaining *perfect* accuracy on our classifier which is not what we want to see. Our model is extremely overfitting to the data and we need to further reduce our feature dimensions to fight the overfitting we continue to see.