Olivetti dataset

Labelled Faces in the Wild dataset

The Olivetti dataset will have better accuracy, as there is less variance between samples in a class, and the Olivetti samples are well-lit and nicely cropped.

Olivetti

Left column: model from first project

```
Predicting people's names on the test set
done in 0.008s
              precision    recall  f1-score   support

           0       1.00      1.00      1.00         3
           1       1.00      1.00      1.00         3
           2       0.75      1.00      0.86         3
           3       1.00      1.00      1.00         3
           4       1.00      1.00      1.00         3
           5       1.00      1.00      1.00         3
           6       1.00      1.00      1.00         3
           7       1.00      0.67      0.80         3
           8       1.00      1.00      1.00         3
           9       1.00      0.67      0.80         3
          10       1.00      1.00      1.00         3
          11       1.00      1.00      1.00         3
          12       1.00      1.00      1.00         3
          13       1.00      1.00      1.00         3
          14       1.00      1.00      1.00         3
          15       1.00      0.33      0.50         3
          16       1.00      1.00      1.00         3
          17       1.00      1.00      1.00         3
          18       1.00      1.00      1.00         3
          19       1.00      1.00      1.00         3
          20       0.60      1.00      0.75         3
          21       1.00      1.00      1.00         3
          22       1.00      1.00      1.00         3
          23       1.00      1.00      1.00         3
          24       1.00      1.00      1.00         3
          25       1.00      1.00      1.00         3
          26       1.00      1.00      1.00         3
          27       1.00      1.00      1.00         3
          28       1.00      1.00      1.00         3
          29       1.00      1.00      1.00         3
          30       1.00      1.00      1.00         3
          31       1.00      1.00      1.00         3
          32       1.00      1.00      1.00         3
          33       1.00      1.00      1.00         3
          34       1.00      1.00      1.00         3
          35       1.00      1.00      1.00         3
          36       1.00      1.00      1.00         3
          37       1.00      1.00      1.00         3
          38       1.00      1.00      1.00         3
          39       0.75      1.00      0.86         3

    accuracy                           0.97       120
   macro avg       0.98      0.97      0.96       120
weighted avg       0.98      0.97      0.96       120

[[3 0 0 ... 0 0 0]
 [0 3 0 ... 0 0 0]
 [0 0 3 ... 0 0 0]
 ...
 [0 0 0 ... 3 0 0]
 [0 0 0 ... 0 3 0]
 [0 0 0 ... 0 0 3]]
```

The Olivetti dataset has great accuracy across both models.

Right column: model from second project

```
              precision    recall  f1-score   support

           0       1.00      0.33      0.50         3
           1       1.00      1.00      1.00         3
           2       0.67      0.67      0.67         3
           3       1.00      1.00      1.00         3
           4       1.00      1.00      1.00         3
           5       1.00      1.00      1.00         3
           6       1.00      1.00      1.00         3
           7       0.67      0.67      0.67         3
           8       1.00      1.00      1.00         3
           9       1.00      0.67      0.80         3
          10       1.00      1.00      1.00         3
          11       1.00      1.00      1.00         3
          12       1.00      0.33      0.50         3
          13       1.00      1.00      1.00         3
          14       0.75      1.00      0.86         3
          15       0.60      1.00      0.75         3
          16       1.00      1.00      1.00         3
          17       1.00      1.00      1.00         3
          18       1.00      1.00      1.00         3
          19       1.00      1.00      1.00         3
          20       0.75      1.00      0.86         3
          21       1.00      0.67      0.80         3
          22       1.00      1.00      1.00         3
          23       1.00      1.00      1.00         3
          24       0.75      1.00      0.86         3
          25       1.00      0.67      0.80         3
          26       1.00      1.00      1.00         3
          27       1.00      1.00      1.00         3
          28       1.00      1.00      1.00         3
          29       1.00      1.00      1.00         3
          30       1.00      1.00      1.00         3
          31       1.00      0.67      0.80         3
          32       1.00      1.00      1.00         3
          33       1.00      1.00      1.00         3
          34       1.00      1.00      1.00         3
          35       1.00      1.00      1.00         3
          36       1.00      1.00      1.00         3
          37       1.00      1.00      1.00         3
          38       0.75      1.00      0.86         3
          39       0.60      1.00      0.75         3

    accuracy                           0.92       120
   macro avg       0.94      0.92      0.91       120
weighted avg       0.94      0.92      0.91       120

========== LDA RESULT ==========
Accuracy score:0.94

========== LR RESULT ==========
Accuracy score:0.93

========== NB RESULT ==========
Accuracy score:0.85

========== KNN RESULT ==========
Accuracy score:0.79

========== DT RESULT ==========
Accuracy score:0.60

========== SVM RESULT ==========
Accuracy score:0.92
```

LFW, min samples/person = 100

```
Predicting people's names on the test set
done in 0.060s
              precision    recall  f1-score   support

           0       0.84      0.94      0.89        71
           1       0.96      0.75      0.84        36
           2       0.86      0.96      0.90       159
           3       0.96      0.70      0.81        33
           4       0.88      0.67      0.76        43

    accuracy                           0.87       342
   macro avg       0.90      0.80      0.84       342
weighted avg       0.88      0.87      0.87       342

[[ 67   0   3   0   1]
 [  2  27   6   0   1]
 [  7   0 152   0   0]
 [  1   1   6  23   2]
 [  3   0  10   1  29]]
```

With many samples per class, and few classes, both perform relatively well. First model performs better.

```
              precision    recall  f1-score   support

           0       0.90      0.75      0.82        71
           1       1.00      0.42      0.59        36
           2       0.68      0.98      0.80       159
           3       1.00      0.48      0.65        33
           4       0.95      0.49      0.65        43

    accuracy                           0.76       342
   macro avg       0.91      0.62      0.70       342
weighted avg       0.82      0.76      0.75       342
========== LDA RESULT ==========
Accuracy score:0.87

========== LR RESULT ==========
Accuracy score:0.83

========== NB RESULT ==========
Accuracy score:0.72

========== KNN RESULT ==========
Accuracy score:0.66

========== DT RESULT ==========
Accuracy score:0.40

========== SVM RESULT ==========
Accuracy score:0.76
```

LFW, min samples/person = 100, reduce classes to 4

```
Predicting people's names on the test set
done in 0.051s
              precision    recall  f1-score   support

           0       0.86      0.94      0.90        71
           1       0.87      0.72      0.79        36
           2       0.92      0.95      0.93       159
           3       1.00      0.82      0.90        33

    accuracy                           0.91       299
   macro avg       0.91      0.86      0.88       299
weighted avg       0.91      0.91      0.90       299

[[ 67   0   4   0]
 [  6  26   4   0]
 [  5   3 151   0]
 [  0   1   5  27]]
```

Reducing the number of classes increases the accuracy.

```
              precision    recall  f1-score   support

           0       0.93      0.79      0.85        71
           1       1.00      0.44      0.62        36
           2       0.75      0.98      0.85       159
           3       1.00      0.48      0.65        33

    accuracy                           0.82       299
   macro avg       0.92      0.67      0.74       299
weighted avg       0.85      0.82      0.80       299
========== LDA RESULT ==========
Accuracy score:0.88

========== LR RESULT ==========
Accuracy score:0.86

========== NB RESULT ==========
Accuracy score:0.78

========== KNN RESULT ==========
Accuracy score:0.73

========== DT RESULT ==========
Accuracy score:0.53

========== SVM RESULT ==========
Accuracy score:0.82
```

LFW, min samples/person = 100, reduce samples to 100

```
Predicting people's names on the test set
done in 0.008s
              precision    recall  f1-score   support

           0       0.73      0.73      0.73        30
           1       0.81      0.87      0.84        30
           2       0.79      0.63      0.70        30
           3       0.81      0.83      0.82        30
           4       0.76      0.83      0.79        30

    accuracy                           0.78       150
   macro avg       0.78      0.78      0.78       150
weighted avg       0.78      0.78      0.78       150

[[22  2  2  1  3]
 [ 0 26  1  2  1]
 [ 4  3 19  2  2]
 [ 1  1  1 25  2]
 [ 3  0  1  1 25]]
```

Reducing the number of samples decreases the accuracy (even though we are equalizing the number of samples for all classes). However, the models have around the same accuracy now (will come back to this potential error at the end).

```
              precision    recall  f1-score   support

           0       0.73      0.73      0.73        30
           1       0.81      0.83      0.82        30
           2       0.83      0.83      0.83        30
           3       0.71      0.67      0.69        30
           4       0.71      0.73      0.72        30

    accuracy                           0.76       150
   macro avg       0.76      0.76      0.76       150
weighted avg       0.76      0.76      0.76       150
========== LDA RESULT ==========
Accuracy score:0.76

========== LR RESULT ==========
Accuracy score:0.73

========== NB RESULT ==========
Accuracy score:0.71

========== KNN RESULT ==========
Accuracy score:0.53

========== DT RESULT ==========
Accuracy score:0.39

========== SVM RESULT ==========
Accuracy score:0.76
```

LFW, min samples/person = 20, reduce classes to 40, reduce samples to 20

```
Predicting people's names on the test set
done in 0.036s
```

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.50 | 0.50 | 0.50 | 6 |
| 1 | 0.50 | 0.17 | 0.25 | 6 |
| 2 | 0.50 | 0.17 | 0.25 | 6 |
| 3 | 0.33 | 0.33 | 0.33 | 6 |
| 4 | 0.27 | 0.50 | 0.35 | 6 |
| 5 | 1.00 | 0.83 | 0.91 | 6 |
| 6 | 0.00 | 0.00 | 0.00 | 6 |
| 7 | 0.57 | 0.67 | 0.62 | 6 |
| 8 | 0.43 | 0.50 | 0.46 | 6 |
| 9 | 0.75 | 0.50 | 0.60 | 6 |
| 10 | 0.33 | 0.17 | 0.22 | 6 |
| 11 | 0.40 | 0.33 | 0.36 | 6 |
| 12 | 0.12 | 0.17 | 0.14 | 6 |
| 13 | 0.57 | 0.67 | 0.62 | 6 |
| 14 | 0.57 | 0.67 | 0.62 | 6 |
| 15 | 0.18 | 0.33 | 0.24 | 6 |
| 16 | 1.00 | 0.67 | 0.80 | 6 |
| 17 | 0.33 | 0.50 | 0.40 | 6 |
| 18 | 0.20 | 0.33 | 0.25 | 6 |
| 19 | 0.60 | 0.50 | 0.55 | 6 |
| 20 | 0.00 | 0.00 | 0.00 | 6 |
| 21 | 0.20 | 0.17 | 0.18 | 6 |
| 22 | 0.11 | 0.17 | 0.13 | 6 |
| 23 | 1.00 | 0.33 | 0.50 | 6 |
| 24 | 0.29 | 0.33 | 0.31 | 6 |
| 25 | 0.57 | 0.67 | 0.62 | 6 |
| 26 | 0.67 | 0.33 | 0.44 | 6 |
| 27 | 0.11 | 0.17 | 0.13 | 6 |
| 28 | 0.60 | 0.50 | 0.55 | 6 |
| 29 | 0.50 | 0.67 | 0.57 | 6 |
| 30 | 1.00 | 0.67 | 0.80 | 6 |
| 31 | 0.40 | 0.67 | 0.50 | 6 |
| 32 | 0.50 | 0.33 | 0.40 | 6 |
| 33 | 0.60 | 0.50 | 0.55 | 6 |
| 34 | 1.00 | 0.50 | 0.67 | 6 |
| 35 | 0.80 | 0.67 | 0.73 | 6 |
| 36 | 1.00 | 0.67 | 0.80 | 6 |
| 37 | 0.83 | 0.83 | 0.83 | 6 |
| 38 | 0.50 | 0.17 | 0.25 | 6 |
| 39 | 0.36 | 0.67 | 0.47 | 6 |
| accuracy | | | 0.44 | 240 |
| macro avg | 0.51 | 0.44 | 0.45 | 240 |
| weighted avg | 0.51 | 0.44 | 0.45 | 240 |

```
[[3 0 0 ... 0 0 1]
 [0 1 0 ... 0 0 1]
 [0 0 1 ... 0 0 0]
 ...
 [0 0 0 ... 5 0 0]
 [0 0 0 ... 0 1 0]
 [0 0 0 ... 0 0 4]]
```

With many classes and few samples, both models perform poorly.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.22 | 0.33 | 0.27 | 6 |
| 1 | 0.38 | 0.50 | 0.43 | 6 |
| 2 | 1.00 | 0.67 | 0.80 | 6 |
| 3 | 0.33 | 0.17 | 0.22 | 6 |
| 4 | 0.29 | 0.33 | 0.31 | 6 |
| 5 | 1.00 | 0.50 | 0.67 | 6 |
| 6 | 0.18 | 0.33 | 0.24 | 6 |
| 7 | 1.00 | 0.67 | 0.80 | 6 |
| 8 | 0.29 | 0.33 | 0.31 | 6 |
| 9 | 0.67 | 0.67 | 0.67 | 6 |
| 10 | 0.50 | 0.50 | 0.50 | 6 |
| 11 | 0.29 | 0.33 | 0.31 | 6 |
| 12 | 0.50 | 0.33 | 0.40 | 6 |
| 13 | 1.00 | 0.33 | 0.50 | 6 |
| 14 | 0.43 | 0.50 | 0.46 | 6 |
| 15 | 0.16 | 0.50 | 0.24 | 6 |
| 16 | 0.67 | 0.33 | 0.44 | 6 |
| 17 | 0.00 | 0.00 | 0.00 | 6 |
| 18 | 0.24 | 0.67 | 0.35 | 6 |
| 19 | 0.25 | 0.17 | 0.20 | 6 |
| 20 | 0.60 | 0.50 | 0.55 | 6 |
| 21 | 0.50 | 0.33 | 0.40 | 6 |
| 22 | 0.00 | 0.00 | 0.00 | 6 |
| 23 | 0.50 | 0.50 | 0.50 | 6 |
| 24 | 0.29 | 0.33 | 0.31 | 6 |
| 25 | 0.75 | 0.50 | 0.60 | 6 |
| 26 | 0.60 | 0.50 | 0.55 | 6 |
| 27 | 0.33 | 0.33 | 0.33 | 6 |
| 28 | 1.00 | 0.17 | 0.29 | 6 |
| 29 | 0.80 | 0.67 | 0.73 | 6 |
| 30 | 1.00 | 0.50 | 0.67 | 6 |
| 31 | 0.20 | 0.17 | 0.18 | 6 |
| 32 | 0.33 | 0.50 | 0.40 | 6 |
| 33 | 1.00 | 0.33 | 0.50 | 6 |
| 34 | 0.62 | 0.83 | 0.71 | 6 |
| 35 | 1.00 | 0.33 | 0.50 | 6 |
| 36 | 0.50 | 0.83 | 0.62 | 6 |
| 37 | 1.00 | 0.50 | 0.67 | 6 |
| 38 | 0.20 | 0.17 | 0.18 | 6 |
| 39 | 0.40 | 0.33 | 0.36 | 6 |
| accuracy | | | 0.41 | 240 |
| macro avg | 0.53 | 0.41 | 0.43 | 240 |
| weighted avg | 0.53 | 0.41 | 0.43 | 240 |

```
========== LDA RESULT ==========
Accuracy score:0.47

========== LR RESULT ==========
Accuracy score:0.43

========== NB RESULT ==========
Accuracy score:0.39

========== KNN RESULT ==========
Accuracy score:0.20

========== DT RESULT ==========
Accuracy score:0.11

========== SVM RESULT ==========
Accuracy score:0.41
```

LFW, min samples/person = 20, reduce classes to 10, reduce samples to 20

```
Predicting people's names on the test set
done in 0.002s
```

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.33 | 0.33 | 0.33 | 6 |
| 1 | 0.80 | 0.67 | 0.73 | 6 |
| 2 | 0.60 | 0.50 | 0.55 | 6 |
| 3 | 0.00 | 0.00 | 0.00 | 6 |
| 4 | 0.50 | 0.50 | 0.50 | 6 |
| 5 | 0.56 | 0.83 | 0.67 | 6 |
| 6 | 0.20 | 0.33 | 0.25 | 6 |
| 7 | 1.00 | 0.67 | 0.80 | 6 |
| 8 | 0.57 | 0.67 | 0.62 | 6 |
| 9 | 0.60 | 0.50 | 0.55 | 6 |
| accuracy | | | 0.50 | 60 |
| macro avg | 0.52 | 0.50 | 0.50 | 60 |
| weighted avg | 0.52 | 0.50 | 0.50 | 60 |

```
[[2 0 1 0 0 1 2 0 0 0]
 [0 4 0 1 0 0 1 0 0 0]
 [2 0 3 0 1 0 0 0 0 0]
 [0 1 0 0 0 0 3 0 0 2]
 [0 0 1 1 3 0 1 0 0 0]
 [1 0 0 0 0 5 0 0 0 0]
 [0 0 0 1 2 0 2 0 1 0]
 [0 0 0 0 0 1 0 4 1 0]
 [0 0 0 0 0 2 0 0 4 0]
 [1 0 0 0 0 0 1 0 1 3]]
```

Reducing the number of classes increases the accuracy (in model 2, as expected).

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.75 | 0.50 | 0.60 | 6 |
| 1 | 0.75 | 1.00 | 0.86 | 6 |
| 2 | 0.71 | 0.83 | 0.77 | 6 |
| 3 | 0.43 | 0.50 | 0.46 | 6 |
| 4 | 0.60 | 0.50 | 0.55 | 6 |
| 5 | 0.62 | 0.83 | 0.71 | 6 |
| 6 | 0.50 | 0.33 | 0.40 | 6 |
| 7 | 0.67 | 0.67 | 0.67 | 6 |
| 8 | 0.83 | 0.83 | 0.83 | 6 |
| 9 | 0.80 | 0.67 | 0.73 | 6 |
| accuracy | | | 0.67 | 60 |
| macro avg | 0.67 | 0.67 | 0.66 | 60 |
| weighted avg | 0.67 | 0.67 | 0.66 | 60 |

```
========== LDA RESULT ==========
Accuracy score:0.65

========== LR RESULT ==========
Accuracy score:0.50

========== NB RESULT ==========
Accuracy score:0.48

========== KNN RESULT ==========
Accuracy score:0.38

========== DT RESULT ==========
Accuracy score:0.33

========== SVM RESULT ==========
Accuracy score:0.67
```

LFW, min samples/person = 20, reduce classes to 5, reduce samples to 20

```
Predicting people's names on the test set
done in 0.001s
              precision    recall  f1-score   support

           0       0.38      0.50      0.43         6
           1       0.40      0.33      0.36         6
           2       0.60      0.50      0.55         6
           3       0.33      0.33      0.33         6
           4       0.50      0.50      0.50         6

    accuracy                           0.43        30
   macro avg       0.44      0.43      0.43        30
weighted avg       0.44      0.43      0.43        30

[[3 0 0 3 0]
 [1 2 1 1 1]
 [1 0 3 0 2]
 [2 2 0 2 0]
 [1 1 1 0 3]]
```


Further reducing the number of classes does not increase the accuracy. Model 1 performs worse, will come back to this potential error at the end.

```
              precision    recall  f1-score   support

           0       0.50      0.67      0.57         6
           1       0.50      0.67      0.57         6
           2       1.00      1.00      1.00         6
           3       0.67      0.33      0.44         6
           4       0.80      0.67      0.73         6

    accuracy                           0.67        30
   macro avg       0.69      0.67      0.66        30
weighted avg       0.69      0.67      0.66        30
========== LDA RESULT ==========
Accuracy score:0.70

========== LR RESULT ==========
Accuracy score:0.70

========== NB RESULT ==========
Accuracy score:0.57

========== KNN RESULT ==========
Accuracy score:0.53

========== DT RESULT ==========
Accuracy score:0.33

========== SVM RESULT ==========
Accuracy score:0.67
```

LFW, min samples/person = 100, reduce samples to 20

```
Predicting people's names on the test set
done in 0.001s
              precision    recall  f1-score   support

           0       0.40      0.33      0.36         6
           1       0.25      0.33      0.29         6
           2       0.75      1.00      0.86         6
           3       0.60      0.50      0.55         6
           4       0.25      0.17      0.20         6

    accuracy                           0.47        30
   macro avg       0.45      0.47      0.45        30
weighted avg       0.45      0.47      0.45        30

[[2 2 1 0 1]
 [3 2 0 0 1]
 [0 0 6 0 0]
 [0 1 1 3 1]
 [0 3 0 2 1]]
```


With classes that look more similar to each other, the accuracy decreases (in model 2, as expected).

```
              precision    recall  f1-score   support

           0       0.42      0.83      0.56         6
           1       0.60      0.50      0.55         6
           2       0.67      0.67      0.67         6
           3       0.50      0.33      0.40         6
           4       0.00      0.00      0.00         6

    accuracy                           0.47        30
   macro avg       0.44      0.47      0.43        30
weighted avg       0.44      0.47      0.43        30
========== LDA RESULT ==========
Accuracy score:0.43

========== LR RESULT ==========
Accuracy score:0.40

========== NB RESULT ==========
Accuracy score:0.50

========== KNN RESULT ==========
Accuracy score:0.33

========== DT RESULT ==========
Accuracy score:0.40

========== SVM RESULT ==========
Accuracy score:0.47
```

**Potential error:** for all the tests where the number of samples were reduced, model 1 (the left column) behaved much worse relative the model 2. When the samples weren't reduced, model 1 performed better. When the samples were reduced, model 2 performed the same or worse while displaying irregular behavior, while model 2 exhibited the expected behavior. Will test this potential error by formulating curated dataset, so that the reduceClassesAndSamples() function will not have to be used.

Now using the make_dataset() function (no need to use reduceClassesAndSamples() anymore

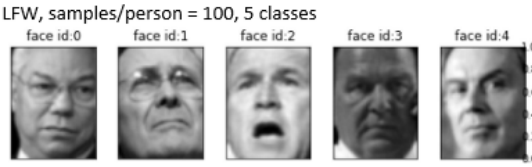
LFW, min samples/person = 100, 5 classes


```
Predicting people's names on the test set
done in 0.063s
              precision    recall  f1-score   support

           0       0.82      0.93      0.87        71
           1       0.96      0.72      0.83        36
           2       0.87      0.94      0.90       159
           3       0.96      0.70      0.81        33
           4       0.84      0.74      0.79        43

    accuracy                           0.87       342
   macro avg       0.89      0.81      0.84       342
weighted avg       0.87      0.87      0.87       342

[[ 66   0   4   0   1]
 [  2  26   5   1   2]
 [  8   0 150   0   1]
 [  1   1   6  23   2]
 [  3   0   8   0  32]]
```

```
              precision    recall  f1-score   support

           0       0.89      0.79      0.84        71
           1       1.00      0.42      0.59        36
           2       0.68      0.98      0.80       159
           3       1.00      0.39      0.57        33
           4       0.95      0.49      0.65        43

    accuracy                           0.76       342
   macro avg       0.90      0.61      0.69       342
weighted avg       0.82      0.76      0.75       342
========== LDA RESULT ==========
Accuracy score:0.86

========== LR RESULT ==========
Accuracy score:0.82

========== NB RESULT ==========
Accuracy score:0.72

========== KNN RESULT ==========
Accuracy score:0.69

========== DT RESULT ==========
Accuracy score:0.41

========== SVM RESULT ==========
Accuracy score:0.76
```

Accuracy of both models are the same as when fetch_lfw_people() was used to create the dataset (serves as the control).

LFW, samples/person = 100, 5 classes

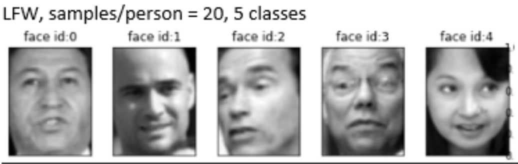



```
Predicting people's names on the test set
done in 0.008s
              precision    recall  f1-score   support

           0       0.93      0.93      0.93        30
           1       0.83      0.83      0.83        30
           2       0.84      0.87      0.85        30
           3       0.85      0.73      0.79        30
           4       0.85      0.93      0.89        30

    accuracy                           0.86       150
   macro avg       0.86      0.86      0.86       150
weighted avg       0.86      0.86      0.86       150

[[28  0  0  1  1]
 [ 2 25  2  0  1]
 [ 0  3 26  1  0]
 [ 0  2  3 22  3]
 [ 0  0  0  2 28]]
```

```
              precision    recall  f1-score   support

           0       0.84      0.90      0.87        30
           1       0.80      0.67      0.73        30
           2       0.79      0.87      0.83        30
           3       0.77      0.80      0.79        30
           4       0.72      0.70      0.71        30

    accuracy                           0.79       150
   macro avg       0.79      0.79      0.78       150
weighted avg       0.79      0.79      0.78       150
========== LDA RESULT ==========
Accuracy score:0.79

========== LR RESULT ==========
Accuracy score:0.71

========== NB RESULT ==========
Accuracy score:0.69

========== KNN RESULT ==========
Accuracy score:0.58

========== DT RESULT ==========
Accuracy score:0.43

========== SVM RESULT ==========
Accuracy score:0.79
```

Model 2 performed around the same as before, and model 1 performed much better, confirming our theory that reduceClassesAndSamples was breaking its functionality.

LFW, samples/person = 20, 5 classes

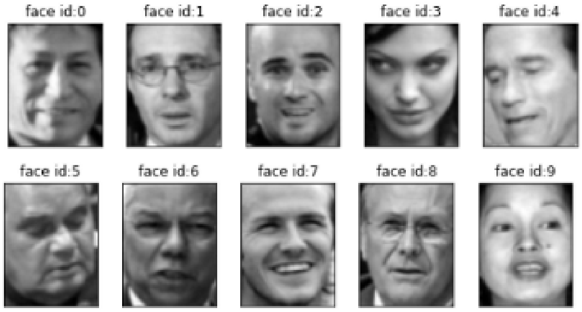



```
Predicting people's names on the test set
done in 0.001s
              precision    recall  f1-score   support

           0       0.30      0.50      0.37         6
           1       0.71      0.83      0.77         6
           2       0.67      0.33      0.44         6
           3       0.67      0.67      0.67         6
           4       1.00      0.67      0.80         6

    accuracy                           0.60        30
   macro avg       0.67      0.60      0.61        30
weighted avg       0.67      0.60      0.61        30

[[3 2 1 0 0]
 [0 5 0 1 0]
 [3 0 2 1 0]
 [2 0 0 4 0]
 [2 0 0 0 4]]
```

```
              precision    recall  f1-score   support

           0       1.00      0.67      0.80         6
           1       0.60      0.50      0.55         6
           2       0.45      0.83      0.59         6
           3       0.80      0.67      0.73         6
           4       1.00      0.83      0.91         6

    accuracy                           0.70        30
   macro avg       0.77      0.70      0.71        30
weighted avg       0.77      0.70      0.71        30
========== LDA RESULT ==========
Accuracy score:0.73

========== LR RESULT ==========
Accuracy score:0.63

========== NB RESULT ==========
Accuracy score:0.53

========== KNN RESULT ==========
Accuracy score:0.60

========== DT RESULT ==========
Accuracy score:0.53

========== SVM RESULT ==========
Accuracy score:0.70
```

Model 2 performed around the same as before, and model 1 better but still worse than model 2. Hypothesis: model 2 is better when there are fewer samples.

LFW, samples/person = 20, 10 classes

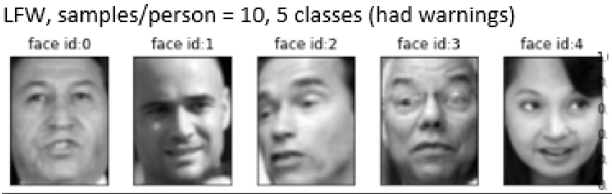



```
Predicting people's names on the test set
done in 0.002s
              precision    recall  f1-score   support

           0       0.50      0.67      0.57         6
           1       0.44      0.67      0.53         6
           2       0.40      0.33      0.36         6
           3       0.40      0.67      0.50         6
           4       0.71      0.83      0.77         6
           5       0.50      0.67      0.57         6
           6       1.00      0.50      0.67         6
           7       0.50      0.33      0.40         6
           8       1.00      0.50      0.67         6
           9       1.00      0.50      0.67         6

    accuracy                           0.57        60
   macro avg       0.65      0.57      0.57        60
weighted avg       0.65      0.57      0.57        60

[[4 0 0 1 1 0 0 0 0 0]
 [0 4 1 0 1 0 0 0 0 0]
 [0 2 2 0 0 2 0 0 0 0]
 [0 1 0 4 0 0 0 1 0 0]
 [0 1 0 0 5 0 0 0 0 0]
 [1 0 1 0 0 4 0 0 0 0]
 [2 0 0 1 0 0 3 0 0 0]
 [1 1 1 0 0 1 0 2 0 0]
 [0 0 0 1 0 1 0 1 3 0]
 [0 0 0 3 0 0 0 0 0 3]]
```

```
              precision    recall  f1-score   support

           0       0.67      0.33      0.44         6
           1       0.80      0.67      0.73         6
           2       0.33      0.33      0.33         6
           3       0.67      0.67      0.67         6
           4       0.33      0.50      0.40         6
           5       0.50      0.33      0.40         6
           6       0.83      0.83      0.83         6
           7       0.14      0.17      0.15         6
           8       0.57      0.67      0.62         6
           9       0.71      0.83      0.77         6

    accuracy                           0.53        60
   macro avg       0.56      0.53      0.53        60
weighted avg       0.56      0.53      0.53        60
========== LDA RESULT ==========
Accuracy score:0.60

========== LR RESULT ==========
Accuracy score:0.52

========== NB RESULT ==========
Accuracy score:0.48

========== KNN RESULT ==========
Accuracy score:0.37

========== DT RESULT ==========
Accuracy score:0.15

========== SVM RESULT ==========
Accuracy score:0.53
```

Model 2 performed much worse, and model 1 around the same. Assuming that the performance of model 2 can serve as the control, we hypothesize that the quality of samples can have a huge impact on model performance.

LFW, samples/person = 10, 5 classes (had warnings)




```
Predicting people's names on the test set
done in 0.002s
              precision    recall  f1-score   support

           0       0.50      0.67      0.57         3
           1       0.50      0.67      0.57         3
           2       0.00      0.00      0.00         3
           3       0.60      1.00      0.75         3
           4       1.00      0.67      0.80         3

    accuracy                           0.60        15
   macro avg       0.52      0.60      0.54        15
weighted avg       0.52      0.60      0.54        15

[[2 1 0 0 0]
 [0 2 0 1 0]
 [1 1 0 1 0]
 [0 0 0 3 0]
 [1 0 0 0 2]]
```

```
              precision    recall  f1-score   support

           0       0.75      1.00      0.86         3
           1       1.00      0.67      0.80         3
           2       0.50      0.33      0.40         3
           3       0.60      1.00      0.75         3
           4       1.00      0.67      0.80         3

    accuracy                           0.73        15
   macro avg       0.77      0.73      0.72        15
weighted avg       0.77      0.73      0.72        15
========== LDA RESULT ==========
Accuracy score:0.87

========== LR RESULT ==========
Accuracy score:0.73

========== NB RESULT ==========
Accuracy score:0.73

========== KNN RESULT ==========
Accuracy score:0.73

========== DT RESULT ==========
Accuracy score:0.60

========== SVM RESULT ==========
Accuracy score:0.73
```

Model 2 performed unexpectedly very well. Perhaps these samples were of high quality. However, model 1 had warning when plotting the eigenfaces, so there might still be some broken functionality in some cases.

LFW, samples/person = 10, 10 classes


```
Predicting people's names on the test set
done in 0.001s
              precision    recall  f1-score   support

           0       0.50      0.67      0.57         3
           1       0.50      0.33      0.40         3
           2       1.00      0.33      0.50         3
           3       0.50      0.33      0.40         3
           4       0.50      0.33      0.40         3
           5       0.14      0.33      0.20         3
           6       0.60      1.00      0.75         3
           7       0.00      0.00      0.00         3
           8       0.33      0.33      0.33         3
           9       1.00      0.67      0.80         3

    accuracy                           0.43        30
   macro avg       0.51      0.43      0.44        30
weighted avg       0.51      0.43      0.44        30

[[2 0 0 1 0 0 0 0 0 0]
 [0 1 0 0 0 1 1 0 0 0]
 [0 1 1 0 0 1 0 0 0 0]
 [0 0 0 1 0 2 0 0 0 0]
 [0 0 0 0 1 0 1 0 1 0]
 [1 0 0 0 1 1 0 0 0 0]
 [0 0 0 0 0 0 3 0 0 0]
 [0 0 0 0 0 2 0 0 1 0]
 [0 0 0 0 0 0 0 2 1 0]
 [1 0 0 0 0 0 0 0 0 2]]
```

```
              precision    recall  f1-score   support

           0       0.25      0.33      0.29         3
           1       0.50      0.67      0.57         3
           2       0.29      0.67      0.40         3
           3       0.00      0.00      0.00         3
           4       1.00      0.33      0.50         3
           5       0.20      0.33      0.25         3
           6       0.75      1.00      0.86         3
           7       0.00      0.00      0.00         3
           8       0.50      0.33      0.40         3
           9       0.00      0.00      0.00         3

    accuracy                           0.37        30
   macro avg       0.35      0.37      0.33        30
weighted avg       0.35      0.37      0.33        30
========== LDA RESULT ==========
Accuracy score:0.60

========== LR RESULT ==========
Accuracy score:0.47

========== NB RESULT ==========
Accuracy score:0.50

========== KNN RESULT ==========
Accuracy score:0.43

========== DT RESULT ==========
Accuracy score:0.37

========== SVM RESULT ==========
Accuracy score:0.37
```

We have reached to limitations of our model, there are too many classes and too few samples to predict samples to an acceptable level. One outlier is the LDA results in model 2 is still pretty good.

**Key takeaways:**

- reduceSamplesAndClasses() was negatively affecting the performance of model 1, so we will not use that code moving forwards
- Model 2 seems to perform better than model 1 when there are fewer samples, and more consistently overall. Furthermore, LDA in model 2 seemed to perform the most consistently, and achieved great results with only a few samples. Will use LDA moving forwards.
- The quality of the samples significantly affects our either model's performance
- We need at least 20 samples per person (14 labelled images) for 10 classes, or 10 samples per person (7 labelled images) for 5 classes to achieve decent performance

After recording these results, I briefly experimented using different parameters in the models, and trying LDA for model 1. The notebook on GitHub is up to date with these efforts. LDA with model 1 did seem to perform better, but LDA with model 2 still performed the best. Therefore, we will continue to use model 2 LDA for the application, but have other models ready as back up.