Yale Eigenfaces

AMATH563: Homework 3

Dan Gorringe
dngrrng@uw.edu

May 28, 2020

Sec. I. Introduction and Overview

Supervised, and unsupervised clustering algorithms are used on the extended yale faces dataset. Individual identification is explored, as are gender groupings to an extent. The homework finds that generalisable models can be created using this dataset, and a discussion on facial dimensions is had.

The findings come from a Python Notebook that can be found on my Github[1].

Sec. II. Theoretical Background

Singular Value Decomposition

Lots of datasets used contain relatively low entropy - all the possible images of humanoid faces is proportionally zero when compared to all the possible pixel and weighting combinations that make the full image space. Many problems are established on lower dimensions than their accompanying data.

Singular Value Decomposition, SVD, decomposes complex matrices uniquely. A diagonal matrix, $\underline{\underline{\Sigma}}$, is pre and post multiplied by unitary matrices.

$$\underline{\underline{X}} = \underline{\underline{U}}\underline{\underline{\Sigma}}\underline{\underline{V}}^*$$

The modes of the dataset, $\underline{\underline{U}}$, are weighted by their sigma value, and the contributing factors in $\underline{\underline{V}}^*$. The beautiful significance of SVD is that using the modes accounting for a large proportion of energy in $\underline{\underline{\Sigma}}$ is able to represent high dimensional data accurately with many less dimensions. Datasets can be dimensionally reduced using truncated SVD, with little loss.

Clustering

The task of finding labels is a significant task in the domain of machine learning. In supervised learning from historical data future classifications are desired, whereas in unsupervised learning unlabelled data is given in hopes to find suitable categories and patterns. Clustering is the process of finding similar samples in data - either to predict new entries of find underlying relationships.

k-means

k-means is a simplistic unsupervised clustering algorithm. On guessing a number of clusters, k clusters are random placed. Iteratively the centroids move to the weighted centre of all the observations that are nearest.

k-Nearest Neighbours

As a supervised algorithm, k-Nearest Neighbours uses local real observations to classify new observations. A predicted label is derived from taking the k nearest labels. This can produce non-linear decision boundaries without the traditionally associated computational complexity, however without considering the system holistically outliers in training data are able to skew classification.

Naïve Bayes

Conditional probabilities are approximated, by assuming Independence, and using Bayes rule. Supervised data is used to construct a prior probability distribution to label new data. Real data is infrequently in information orthogonal dimensions, the independence assumption is dubious, as such naïve Bayes can perform poorly.

Support Vector Machines

Support vector machines, SVMs, find hyperplane decision boundaries that maximise the distance between labelled clusters. The more valuable nonlinear SVMs initially map data to a higher dimensional space before finding a classification boundary. Traditionally nonlinear SVMs are computationally intensive, the kernel trick reduces the cost. Different kernels help find different types of clustering, to name a few: Radial basis function, Polynomial and hyperbolic tangent.

Sec III. Algorithm Implementation and Development

Data Wrangling

The Extended Yale faces data given in cropped and uncropped formats came in differing directory layouts, so different functions were devised to load the series of PGM files.

The dataset contains multiple pictures for several people. For generability creating training and testing sets that included subsets of both people and pictures of these people was key. People were left out of the training set, so that the individual was new when apraising the model with the test set. This is beneficial in reviewing the utility when applied to faces that are outside of Yale faces, and that our models are not simply recognising the particular individuals.

[1] https://github.com/DanGorringe/AMATH563_inferringComplexSystems

Classifier Implementation

With the data wrangled into the appropriate format, classifiers were trivially imported and trained using `sklearn`.

Classifier Evaluation

In the following binary classifiers it is important to track not only the accuracy of the model, but the true negative and false negative rates also. To visualise confusion matrices were used. For models trained, and evaluated with unequally sized population sets it was paramount in order to see the utility in diagnosing the minority group.

Cross Validation

Of paramount importance in sharing machine learning models is cross validation. Each model discussed is run multiple times, and their averaged confusion matrices is shown.

Sec IV. Computational Results

Question 1: Do an SVD analysis of the images

A sample of the uncropped images can be seen in Figure 1. On SVD analysis it is found that approximately 25 modes are capable of representing the set, see Figure 2.

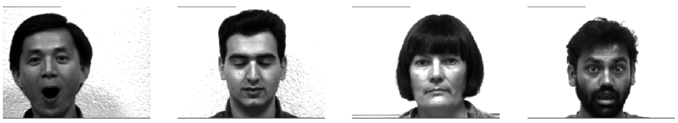

Figure 1: Random selection of images from the original Yale faces dataset given.

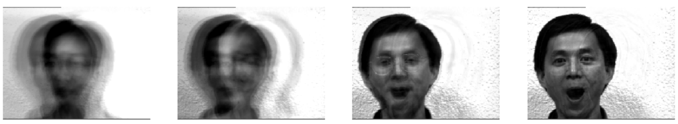

Figure 2: Using 1, 5, 25 and 125 first modes of the SVD to compose the first image from Figure 1.

SVD returns a set of modes, singular values denoting the importance of the modes, and how much each mode contributes to each element in the dataset. The amount each mode contributes to each image is in $\underline{\underline{V}}$. The modes in $\underline{\underline{U}}$, are the eigenfaces. SVD does not produce orthogonal modes, but is in an interesting pseudo-philosophical or aesthetical question as to how many truly orthogonal faces there are. Further the small-worlds extension to the question as to how many orthogonal faces one expects to see in their lifetime.

Looking at the plot of the singular values, Figure 3, it seems like the first 100 modes will recognisably remodel the images.

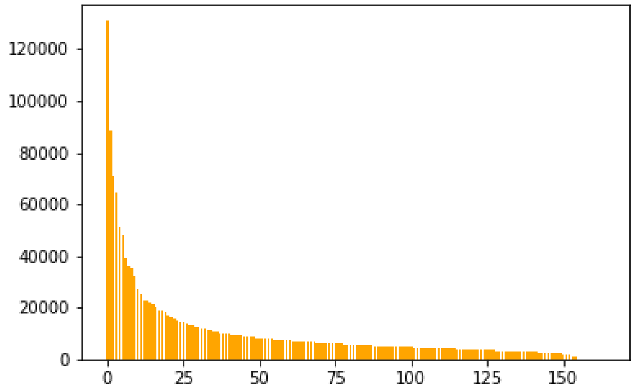

Figure 3: The singular value weightings for uncropped images.

In Figure 5 a sample of the cropped images are shown. There appears to be a series of very shadowy images, although being uninteresting to the human eye it is assumed they help learn from sparse data.

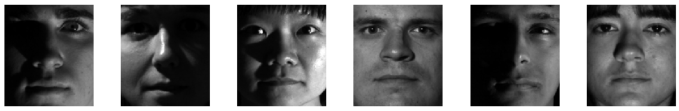

Figure 4: Random selection of images from the cropped Yale faces dataset given.

Figure 5: Using 1, 5, 25, 50, 200 and 1000 first modes of the SVD to compose the first image from Figure 5.

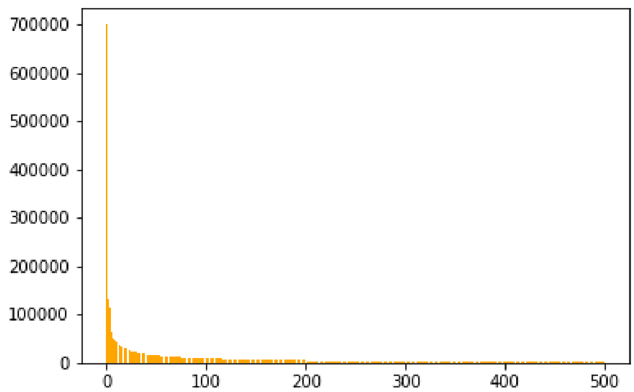

Figure 6: First 500 singular value weightings for uncropped images.

It is initially surprising that the singular value plots for both cropped and uncropped look to tail away after the same number of modes. Remembering that the cropped images set is more than fourteen times larger than uncropped it is impressive that both can be represented in almost equal. Using less modes to encompass more images would indicate both that human faces truly have low entropy, and that the dataset exhibits low dimensionality.

Figure 7: First 100 singular value weightings for uncropped images.

Find the first 6 modes of the uncropped data in Figure 7.

Question 2: Face Classification

Feature selection is used to produce a low dimension space for clustering. SVD provides useful mode features. In Figure 8 the 2D combinations of the first 8 modes are plotted - with desire to find a printable space that works nicely for visualisation.

Concurring with Figure 6, the complete lack of distinct clusterings in Figure 8 shows that multiple modes will be needed to find distinct clusters. Following 100 features are used, $\underline{\underline{U}}$ is truncated to 100 modes.

KNN

Initially, in order to distinguish a single face, group2 was composed of a single person and the remaining 38 were in group1. The confusion matrix, Figure 9, shows the average accuracy in labelling the individual being very low. It is suspected that the total imbalance in amount of training data is responsible.

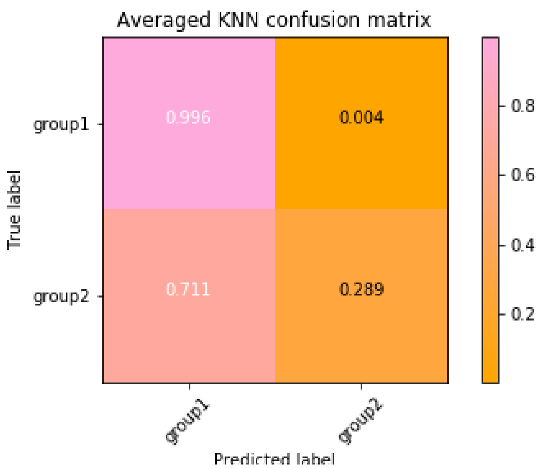

Figure 9: Confusion matrix averaged over 256 folds, of a KNN model predicting an individual. The training set contains all individuals.

Using less 'other' persons in group one training data makes the model predict the individual easier, see the updated confusion matrix in Figure 10. With increased group2 classification accuracy, comes inaccuracy in classifying group1. With KNN it turns out to be difficult to create a generalisable model to find an individual.

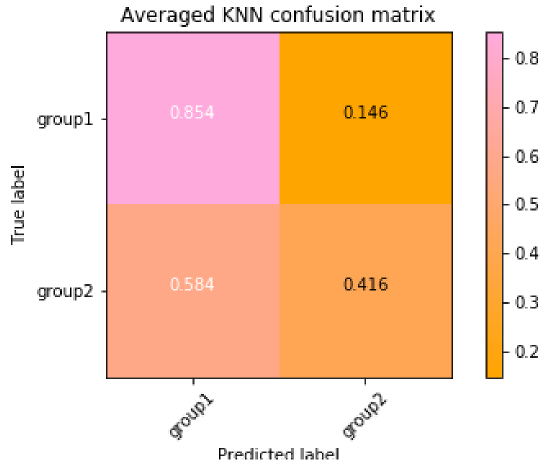

Figure 10: Confusion matrix averaged over 256 folds, of a KNN model predicting an individual. The training set contains only contains 8 individuals of 39.

Naïve Bayes

Continuing with 8 individuals, the same experiment was conducted, using `sklearn`'s naïve Bayes classifier.

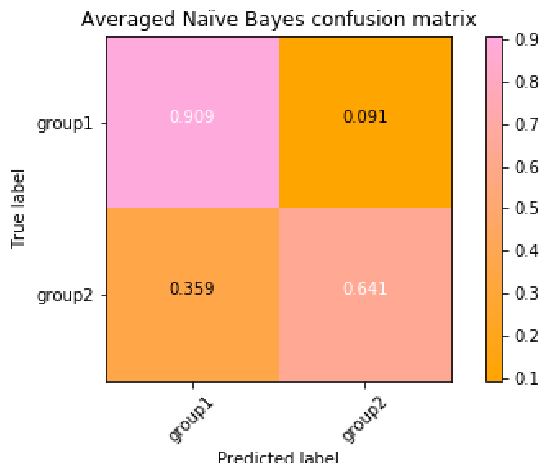

Figure 11: Confusion matrix averaged over 256 folds, of a naïve Bayes model predicting an individual. The training set contains 8 individuals.

Contrary to KNN, naïve Bayes is able to classify the group1 individual with higher accuracy than a regular coin toss. The ability for naïve Bayes to account for outliers pays dividends in this use case.

SVM

A SVM model is also trialled, confusion matrix in Figure 12. Like the naïve Bayes model, it outperforms KNN - there is minimal difference in their confusion matrices so is difficult to rank between.

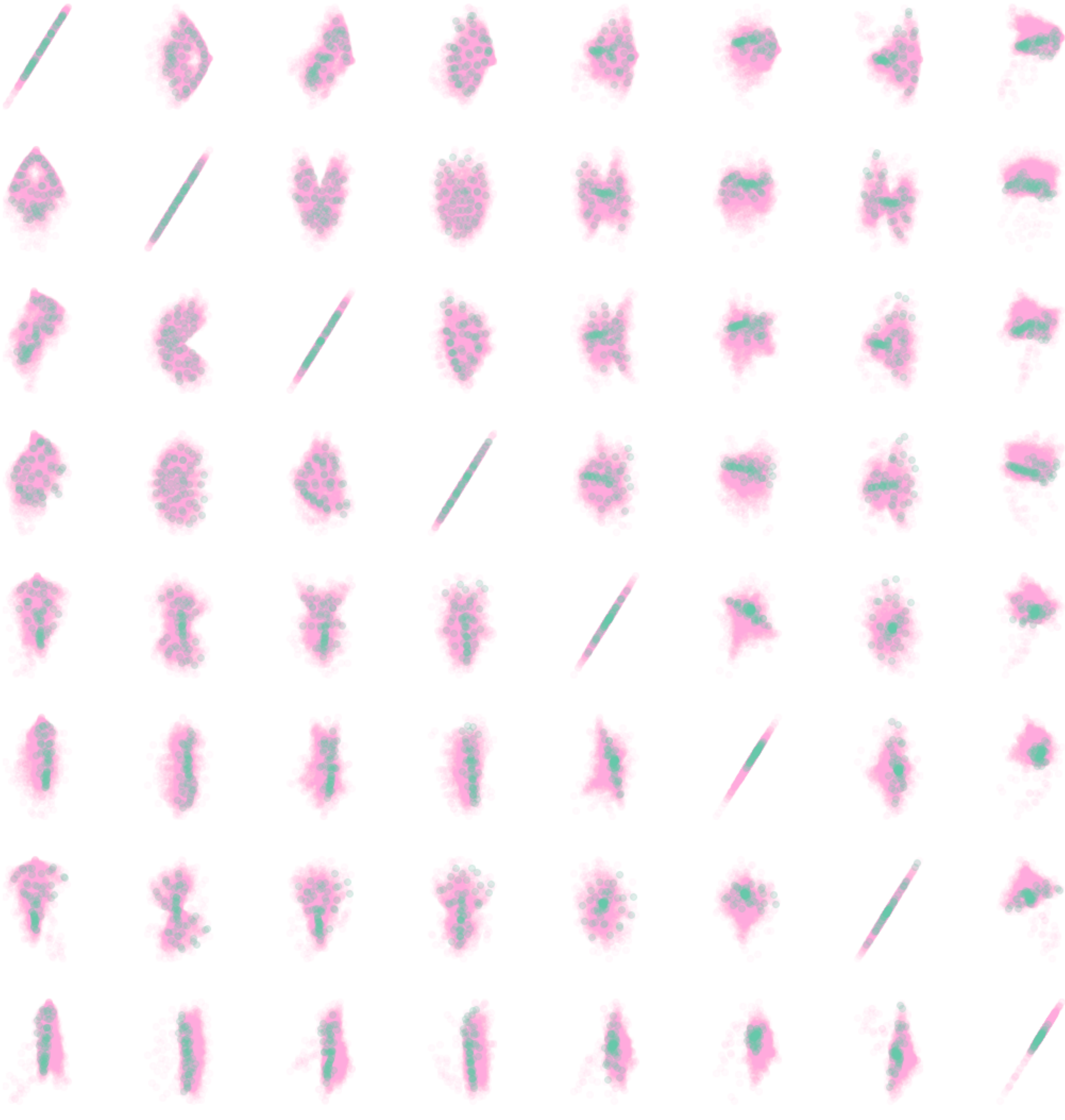

Figure 8: The first 8 SVD modes are plotted against each other for yaleB01, to investigate their utility at clustering in 2 dimensions. Whole population is plotted in millenial pink, and the individual in medium aquamarine

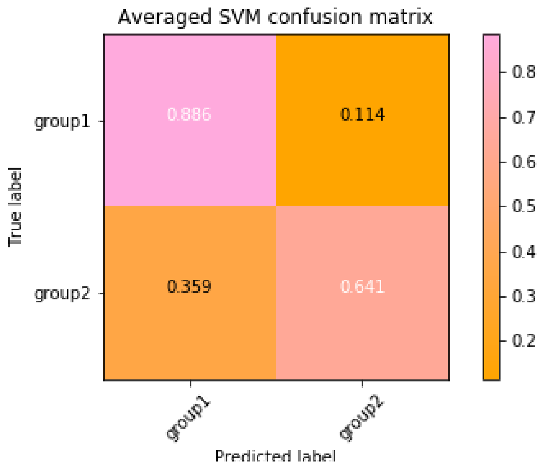

Figure 12: Confusion matrix averaged over 256 folds, of a SVM model predicting an individual. The training set contains 8 individuals.

Question 3: Gender Classification

Test 2 initially seems morally questionable, against American[2], British[3] and International[4] professional bodies' codes of ethics. To be clear, in the following experiments two training groups are chosen based on my personal naïve interpretations of gender expression, however without any self-declarations or verification there is no evidence to suggest validity even if subscribing to a binary interpretation of gender. Without permission, accuracy or willingness to discritise other's genders normative labels are emitted - group1 and group2 contain individuals whom are perceived to be of the same gender expression, but this decision is not shared.

Feature selection is trialled again, to see whether there is a 2-dimensional representation that provides a space to cluster well, see Figre 13. Unsurprisingly, the phenomena of gender predictions transcends two modes of SVD on a limited dataset. It is interesting to consider whether normative gender expression stereotypes are expected to be in a higher or lower space than identifying an individual. It would come naturally that strongly-held, yet wholly inadequate and troublesome, stereotypes of jewellery and hair might limit dimensions required - further study of the dimensionality required to distinguish folk gender determination across time might be sociological and anthropologically insightful.

100 modes were chosen to cluster in, all three models used in test 1 where carried out for group1 and group2 classification.

[2]ACM Code of Ethics and Professional Conduct 1.1, 1.2, 1.4 and 1.6
[3]Royal Academy of Engineering Statement of Ethical Principles 2
[4]IEEE Code of Ethics 7.8.3

KNN

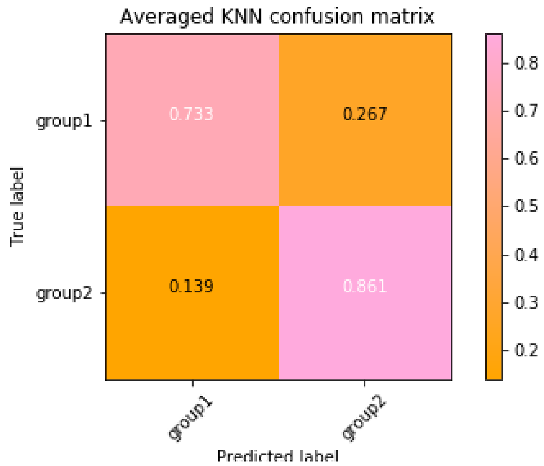

Figure 14: Confusion matrix averaged over 256 folds, of a KNN model predicting group membership. The training set contains all individuals.

It is interesting that despite, groups with skewed membership unlike for a simple individual identification KNN was able to accurately predict both groupings, Figure 14. The implication here might be that contrary to an individual, these groupings share modes that are not perturbed by outlying instances. Normative stereotypes of gender seem to manifest themselves in important modes of the SVD such that in this small dataset there is no difficult in overlapping regions that tend themselves to mis-labelling.

Naïve Bayes

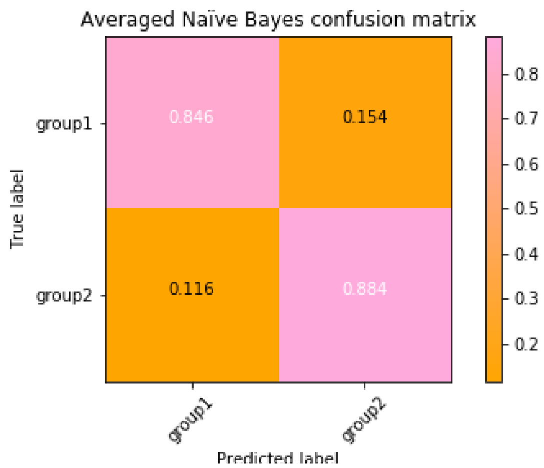

Figure 15: Confusion matrix averaged over 256 folds, of a Naïve Bayes model predicting group membership. The training set contains all individuals.

Figure 15, shows that Naïve Bayes it able to improve on the group1 accuracy in comparison to KNN. This would indicate that rather than no overlapping regions, that they are reduced in respect to identifying an individual. Improvements in both misclassifcations are seen.

Figure 13: The first 8 SVD modes are plotted against each other for yaleB01, to investigate their utility at clustering in 2 dimensions. Group1 is plotted in orange, and Group2 in medium aquamarine

SVM

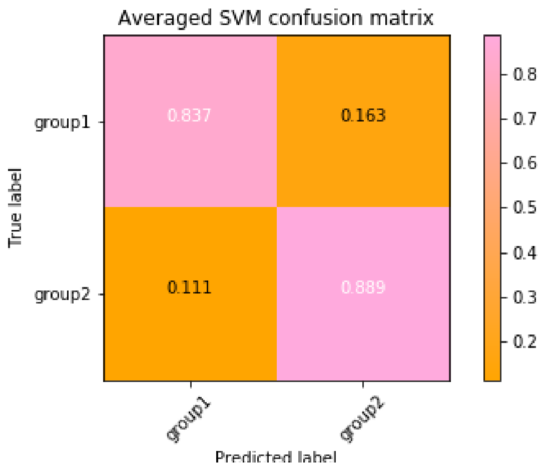

Figure 16: Confusion matrix averaged over 256 folds, of a SVM model predicting group membership. The training set contains all individuals.

In Figure 16 we see, again, that SVM produces similar results.

Question 3: Unsupervised algorithms

For the unsupervised approach, there are no assumptions initially made on the patterns intended to be found.

Figure 17: Sample of 18 images that are clustered into the same group using k-means with 2 clusters on the full dataset truncated to 100 SVD modes. Group0

Figure 18: Sample of 18 images that are clustered into the same group using k-means with 2 clusters on the full dataset truncated to 100 SVD modes. Group1

k-means was used with 2 clusters on the full dataset truncated to 100 modes. Figure 17 & 18 show samples from the clusterings found. It would seem that the images in Group1 tend to be more in shadow than those in Group0 - k-means with 2 clusters roughly finds those images fully engulfed in shadow in one group.

The same experiment is run, but with 3 clusters, Figure 19, 20, & 21. Interestingly we see Group1 clearly show a clustering of well lit images, however Group0 and Group2 seemingly without pattern.

Figure 19: Sample of 18 images that are clustered into the same group using k-means with 2 clusters on the full dataset truncated to 100 SVD modes. Group0

Figure 20: Sample of 18 images that are clustered into the same group using k-means with 2 clusters on the full dataset truncated to 100 SVD modes. Group1

Figure 21: Sample of 18 images that are clustered into the same group using k-means with 2 clusters on the full dataset truncated to 100 SVD modes. Group2

The unsupervised k-means is able to independently discover patterns in lighting.

Sec V. Summary and Conclusions

Despite a dataset limited to only twice the number of total Village Persons (19) the clustering techniques exhibited were able to find patters, either with stereo-typically labelled data or independently. The data wrangling functions left out individuals, and used them in testing to verify that the models found had a degree of generability. The dataset and algorithms used are strong enough to create meaningful low-dimension pattern classifiers on human faces.

An extended discussion on the philosophical and sociological dimensionality of both empirically an image collection representative of people, and the depth required in people's perception would be incredibly interesting and fun to read.