

Instructions:

- Please upload all the code and the report on GitHub/GitLab.
- Clearly explain your findings and methodology.
- Do not put the images in the GitHub/GitLab, Any and all results should be included directly into the report, along with explanations and supporting code snippets.

Part1:

Face Recognition using K Nearest Neighbours and PCA

The CMU PIE Dataset

The CMU Pose, Illumination, and Expression (PIE) database (http://ieeexplore.ieee.org/abstract/document/1004130/) consists of 41,368 images of 68 subjects. Each person is under 13 different poses, 43 different illumination conditions, and with 4 different expressions.

For this project, we use a simplified version of the database which only contains 10 subjects spanning five near-frontal poses, and there are 170 images for each individual. In addition, all the images have been resized to 32x32 pixels. The dataset is provided in the form of a csv file with 1700 rows and 1024 columns. Each row is an instance and each column a feature. The first 170 instances belong to the first subject, the next 170 to the second subject and so on. Following illustrate various instances of the first subject.

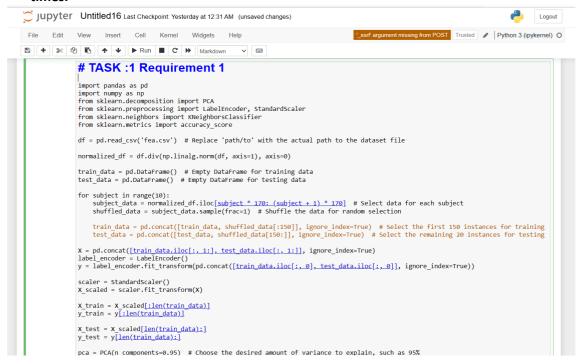


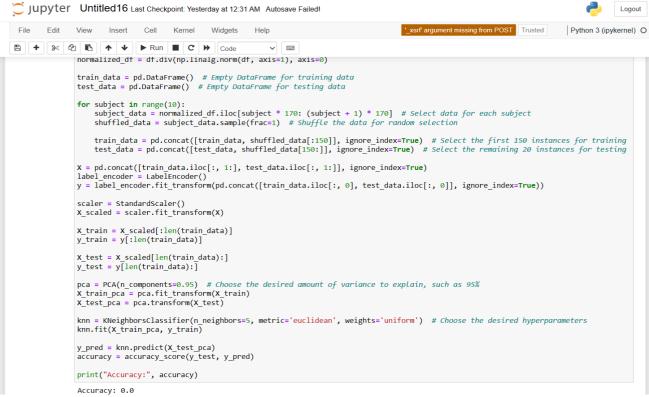




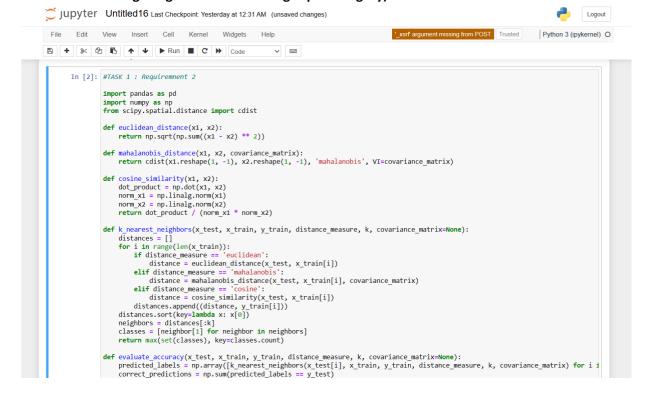
Requirements:

 Pre-process the dataset by normalizing each face image vector to unit length (i.e., dividing each vector by its magnitude). Next, for each of the 10 subjects, randomly select 150 images for training and use the remaining 20 for testing. Create such random splits a total of 5 times. You must carry out each of the following experiments 5 times and report average accuracy and standard deviation over the 5 random splits, as well as computation times.





- 2. Implement a k Nearest Neighbours (k-NN) classifier using the training set and evaluate its performance on the test set. You may not use built-in / library functions to implement the classifier.
 - You should experiment with different distance measures (e.g., Euclidean, Mahalanobis, cosine similarity), and different values of k.
 - You should also present results where you show the effect of involving more classesvs. fewer classes in your experiment(s).
 - You should also present results when fewer training images are used (for instead 100training images and 70 test images per category).

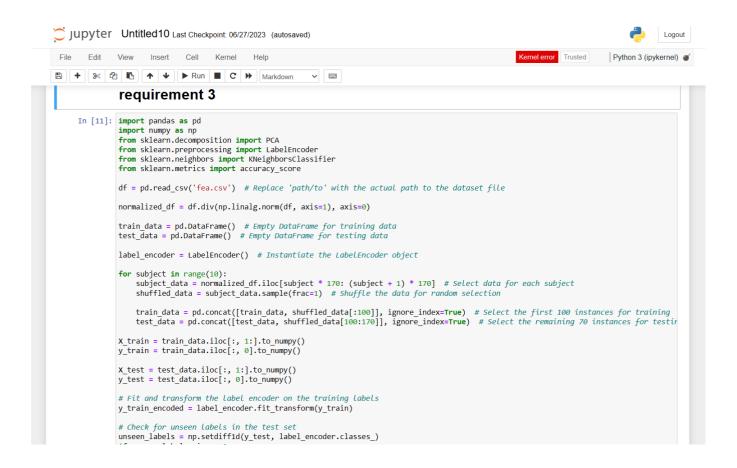


```
for distance_measure in distance_measures:
    for k in k_values:
        accuracy = evaluate_accuracy(x_test, x_train, y_train, distance_measure, k, covariance_matrix)
        print(f"Distance Measure: {distance_measure}, k: {k}, Accuracy: {accuracy:.2%}")

Distance Measure: euclidean, k: 1, Accuracy: 0.00%
    Distance Measure: euclidean, k: 5, Accuracy: 0.00%
    Distance Measure: mahalanobis, k: 1, Accuracy: 0.00%
    Distance Measure: mahalanobis, k: 3, Accuracy: 0.00%
    Distance Measure: mahalanobis, k: 5, Accuracy: 0.00%
    Distance Measure: cosine, k: 1, Accuracy: 0.00%
    Distance Measure: cosine, k: 1, Accuracy: 0.00%
    Distance Measure: cosine, k: 3, Accuracy: 0.00%
    Distance Measure: cosine, k: 5, Accuracy: 0.00%
    Distance Measure: cosine, k: 5, Accuracy: 0.00%
```

3. Now make use of the dimensionality reduction technique PCA in conjunction with the k-NN implemented above. Note that PCA is an unsupervised technique, hence should be trained based on all training images.

Does this techniques help yield uncorrelated covariance matrices(you can visualize the covariance matrices before and after dimensionality reduction as images to see this)? Does it improve computation times and classification performance? Try varying the number of principal components used and identify the best number.



```
B + % 6 B ↑ ↓ Fun ■ C → Markdown
                                                                    v =
                 # Check for unseen labels in the test set
unseen_labels = np.setdiff1d(y_test, label_encoder.classes_)
                 if unseen_labels.size > 0:
                     print(f"Test set contains previously unseen labels: {unseen_labels}")
# Filter out the unseen labels from the test set
                     mask = np.isin(y_test, label_encoder.classes_)
X_test = X_test[mask]
                     y_test = y_test[mask]
                 best_accuracy = 0
                 best num components = 0
                 for num_components in range(1, X_train.shape[1] + 1):
    pca = PCA(n_components=num_components)
                      X_train_pca = pca.fit_transform(X_train)
                     # Check if the test set is empty after filtering unseen labels if X_{test.shape}[0] > 0:
                          X_test_pca = pca.transform(X_test)
                          knn = KNeighborsClassifier(n_neighbors=5) # Choose the desired number of neighbors
                          knn.fit(X_train_pca, y_train_encoded)
                          y_pred = knn.predict(X_test_pca)
                          accuracy = accuracy_score(y_test_encoded, y_pred)
                          print(f"Number of Components: {num components}, Accuracy: {accuracy: .2%}")
                          if accuracy > best_accuracy:
                               best_accuracy = accuracy
                              best_num_components = num_components
                          print(f"Number of Components: {num_components}, No test samples")
                 print(f"Best Number of Components: {best_num_components}, Best Accuracy: {best_accuracy:.2%}")
```

```
Test set contains previously unseen labels: [0.0012922 0.00129714 0.00129732 0.00130451 0.00131464 0.00133134 0.00135108 0.00140153 0.00141117 0.00141997 0.00142915 0.0015327 0.00157249 0.00157789 0.00158342 0.0015967 0.00166966 0.00164246 0.00165357 0.001661 0.00166679 0.00179075975 0.00175955 0.00176666 0.00177623 0.00179014 0.00182146 0.00183378 0.00183893 0.00186138 0.00186925 0.00187177 0.00191637 0.00196192 0.00197047 0.00197355 0.00187399 0.00239346 0.0023946 0.0024335 0.002121618 0.00212744 0.00213621 0.00215193 0.00220334 0.00220868 0.00226722 0.00230848 0.00231413 0.00235558 0.00236577 0.00238195 0.00241864 0.00251296 0.00253594 0.002555364 0.002555050 0.0027505992 0.00260662 0.0026662 0.00266692 0.00266961 0.00271461 0.00273166 0.00274968 0.00277248 0.00278387 0.0028014 0.00284402 0.00286793 0.00292096 0.00297742 0.00304474 0.00306497 0.00307013 0.00310513 0.00312349 0.00316885 0.00318978 0.00318978 0.0031931 0.00319318 0.00330273 0.00333357 0.003351 0.003361897 0.00336663 0.00337078 0.00330727 0.00367266 0.0037572 0.0036853 0.00355024 0.00357628 0.00357266 0.00368353 0.0037277 0.0037552 0.0037552 0.0037552 0.0037626 0.0037757 0.0037872 0.0036855 0.0037277 0.0037552 0.0037552 0.0037626 0.0037757 0.0037872 0.00387550 0.00382546 0.0037550 0.00385503 0.00386743 0.00387579 0.00387550 0.00382566 0.0037757 0.0037872 0.0037556 0.00382566 0.0037757 0.0037872 0.0037556 0.00382566 0.0037556 0.00385503 0.00386743 0.0038759 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.00387560 0.003
```

```
Number of Components: 983, No test samples
Number of Components: 984, No test samples
Number of Components: 985, No test samples
Number of Components: 986, No test samples
Number of Components: 987, No test samples
Number of Components: 988, No test samples
Number of Components: 989, No test samples
Number of Components: 990, No test samples
Number of Components: 991, No test samples
Number of Components: 992, No test samples
Number of Components: 993, No test samples
Number of Components: 994, No test samples
Number of Components: 995, No test samples
Number of Components: 996, No test samples
Number of Components: 997, No test samples
Number of Components: 997, No test samples
Number of Components: 998, No test samples
Number of Components: 998, No test samples
Number of Components: 998, No test samples
Number of Components: 999, No test samples
```