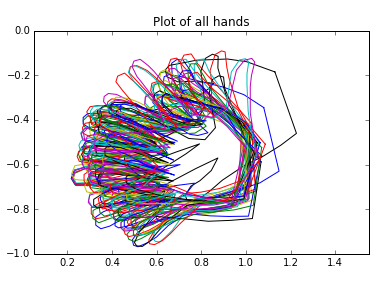
Assignment 2

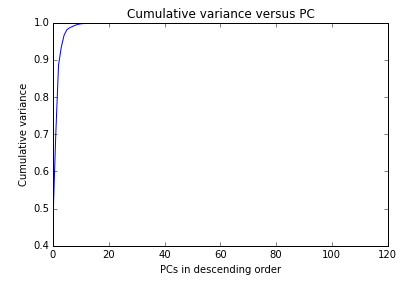
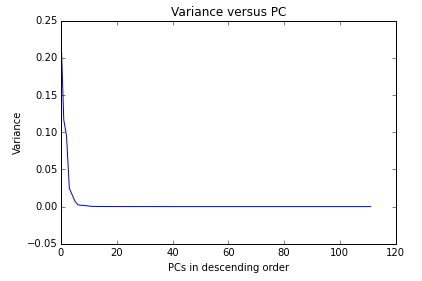
# Exercise 1

## ../../../Desktop/Screen%20Shot%202016-03-07%20at%2020.29.01.



The plot of all hands show that most of the hands are similar in that the fingers are pointing to the left, although there are some outliners that are slightly rotated anticlockwise and positioned more to the right from the rest of the hands.

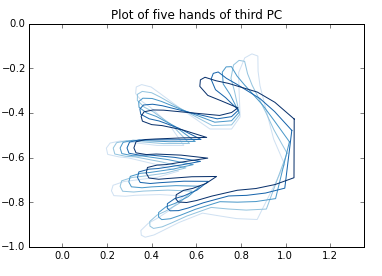
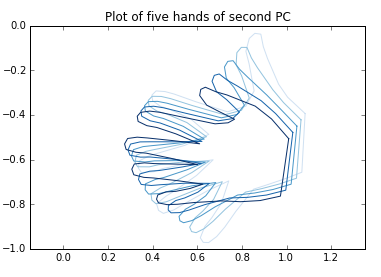
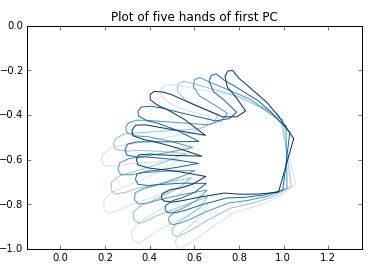
# Exercise 2



The number of dimensions (PCs) needed to capture 90% and 95% of the variance in the dataset is 4 and 5, respectively.

[justify answer]

# Exercise 3



The plot of five hands for the first PC are the most different from each other, suggesting that the first PC contains the largest variance compared to the second PC and the third PC. The plot of five hands for the third PC shows less variation than the plot of five hands for the second PC, which is expected because the variance is supposed to decrease from the first PC to the third PC. Also, the variation of the third PC tends to be along the length of the fingers compared with the variation of the second PC, which tends to be around the thumb and little finger.

# Exercise 4

b) The accuracy of the k-NN classifier on the training data for k = 1, 3, 5 were 100%, 88.8% and 82.7%, respectively. The accuracy of the k-NN classifier on the testing data for k = 1, 3, 5 were 90.0%, 85.6% and 82.5%, respectively.

For both data sets, the first nearest neighbour provided the most accurate prediction of the labels and increasing the number of neighbours introduced more errors, suggesting that a larger neighbourhood may have included points from other labels. The 100% accuracy on the training data for 1-NN classifier is expected because 1-NN classifier always has zero error on the training data.

# Exercise 5

b) The accuracy of the k-NN classifier on the centred and normalised training data for k = 1, 3, 5 were 100%, 95.9% and 90.7%, respectively. The accuracy of the k-NN classifier on the centred and normalised testing data for k = 1, 3, 5 were 87.8%, 93.9% and 90.0%, respectively.

[use a table]

# Exercise 6

b) The optimal k on centred and normalised training set was 3.

c) The error percentage of the optimal k on the centred and normalised test set was 6.1%.