



LOTI.05.046 PATTERN RECOGNITION

HOMEWORK 6 - PCA AND MINIMUM DISTANCE CLASSIFIER

April 22, 2019

Quazi Saimoon Islam
Msc. Robotics and Computer Engineering

1 INTRODUCTION

Principal Component Analysis (PCA) is arguably one of the most widely used statistical methods. It has applications in nearly all areas of statistics and machine learning including clustering, dimensionality reduction, face recognition, signal processing, image compression, visualisation and prediction. Without complicating things too much, PCA is able to recognise underlying patterns in high dimensional data and the outputs of PCA can be used to highlight both the similarities and differences within a dataset. Where PCA really shines, however, is in making sense of high-dimensional data. It can be hard to describe patterns in data with high dimensions, the results of PCA quite often lead to simple interpretations.

Minimum distance classifier however is used to classify unknown data to classes which minimize the distance between the data and the class in multi-feature space. The distance is defined as an index of similarity so that the minimum distance is identical to the maximum similarity.

2 METHODOLOGY

The given methodology was followed and are as follows:

1. Split the data in 80-20% for training and test dataset respectively.
2. Perform minimum distance classification by taking the mean of the training data and computing the distances of the test data (Both Manhattan and Euclidean).
3. Perform PCA and repeat the minimum distance classification this time using the transformed data from the PCA to find the distances.

3 RESULTS

3.1 Did PCA help to improve the results?

PCA definitely improved the results as both Euclidean distance and Manhattan distance accuracies increased from 93% & 90% to 97% each. It was also noticable that after performing PCA, the accuracies did not vary from either method used (Euclidean and Manhattan).

3.2 Why the results improved (or did not improve)?

The main reason that the results seems to be improved following PCA is that the dimensionality reduction has helped shift the data points closer to the mean. Of course for small dataset like the Fisher iris data, this improvement is minimal but still apparent.