**PCA Dimension Reduction**

# **Background Information**

This script can be run with a simple “python PCA\_dim\_reduction.py” and to get more information about how it is running, you can set the variable “DO\_DEBUG” to True before executing the code. The Secom data set is a text file of 1567 data points of 590 features, containing noise and machine process sensor data. The data is pre-processed to be centered and the noise (NaN) data is replaced with the mean of its feature (column). The goal of this project is to reduce the 590 features into a smaller dimension of features without losing any of the data.

# **Explanation**

The “new” projected features are made of eigenvector combinations of the previous ones, so no data is lost upon projection and the process if reversable if desired. The eigenvalues associated with these matrix operations relate to the percentage of data variance that that principal component in question is responsible for. The closer to the beginning of the matrix the component is, the more it accounts for during the calculations. To project the original data into the new dimension, you do a matrix multiplication on the original data (centered) with the eigenvectors you want to use the cumulative variance of. In this case, the first 17 eigenvectors are used to account for 99% of the variance in the data. This method has great uses in storage compression and in reducing training time on large data sets. The results found are listed below.

# **Results**

|  |  |  |
| --- | --- | --- |
| **Principle Component Number** | **% Variance of Principal** | **% Cumulative Variance** |
| 1 | 59.25405798 | 59.25405798 |
| 2 | 24.12381887 | 83.37787685 |
| 3 | 9.150013589 | 92.52789044 |
| 4 | 2.300578517 | 94.82846896 |
| 5 | 1.459192346 | 96.2876613 |
| 6 | 0.518817533 | 96.80647884 |
| 7 | 0.322658095 | 97.12913693 |
| 8 | 0.314676654 | 97.44381359 |
| 9 | 0.263079533 | 97.70689312 |
| 10 | 0.231306658 | 97.93819978 |
| 11 | 0.217534578 | 98.15573436 |
| 12 | 0.207281995 | 98.36301635 |
| 13 | 0.169083769 | 98.53210012 |
| 14 | 0.125590662 | 98.65769078 |
| 15 | 0.120353402 | 98.77804418 |
| 16 | 0.114092102 | 98.89213629 |
| 17 | 0.111115412 | 99.0032517 |