Reykjavik University

Master Degree in Computer Science
Course of Data Mining and Machine Learning

PCA

Author

Davide Parente

Instructor

Eyjólfur Ingi Ásgeirsson
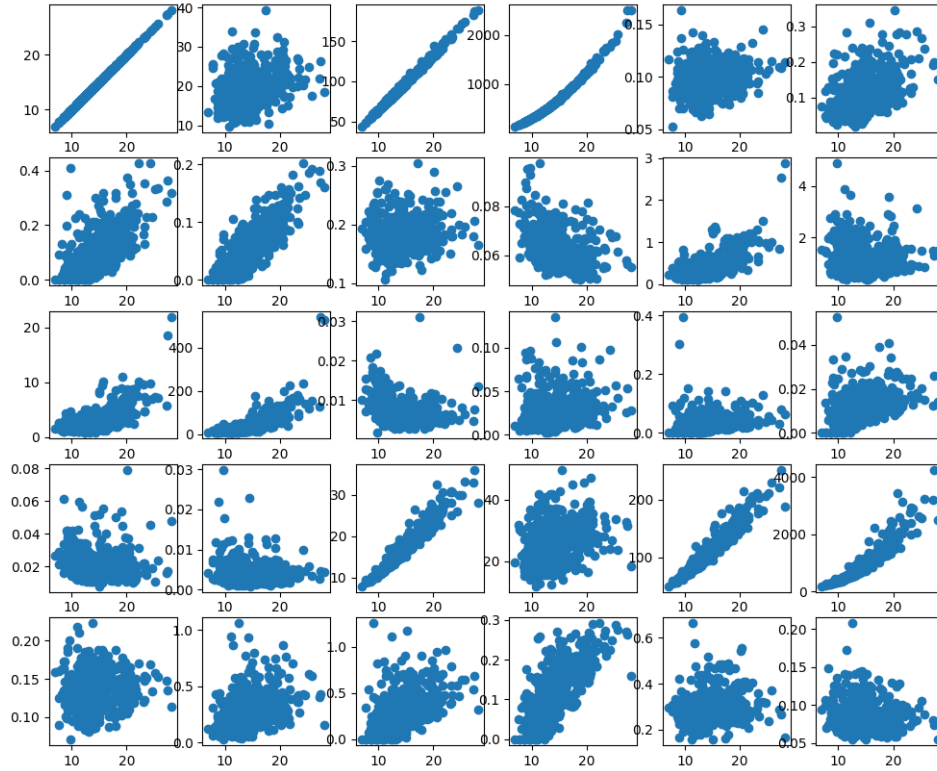
06. November 2022

1 Section 1.3

Figure 1: 1_3_1.png

2 Section 1.4

The measures that are more correlate with the first one are the number 3,4 respectively named : "mean perimeter", and "mean area". The correlation is showed because these measures are linear among them.

These measure are correlate because the first measure is the mean_radius, so these three measures are directly related because we are speaking about perimeter, area, and radius.
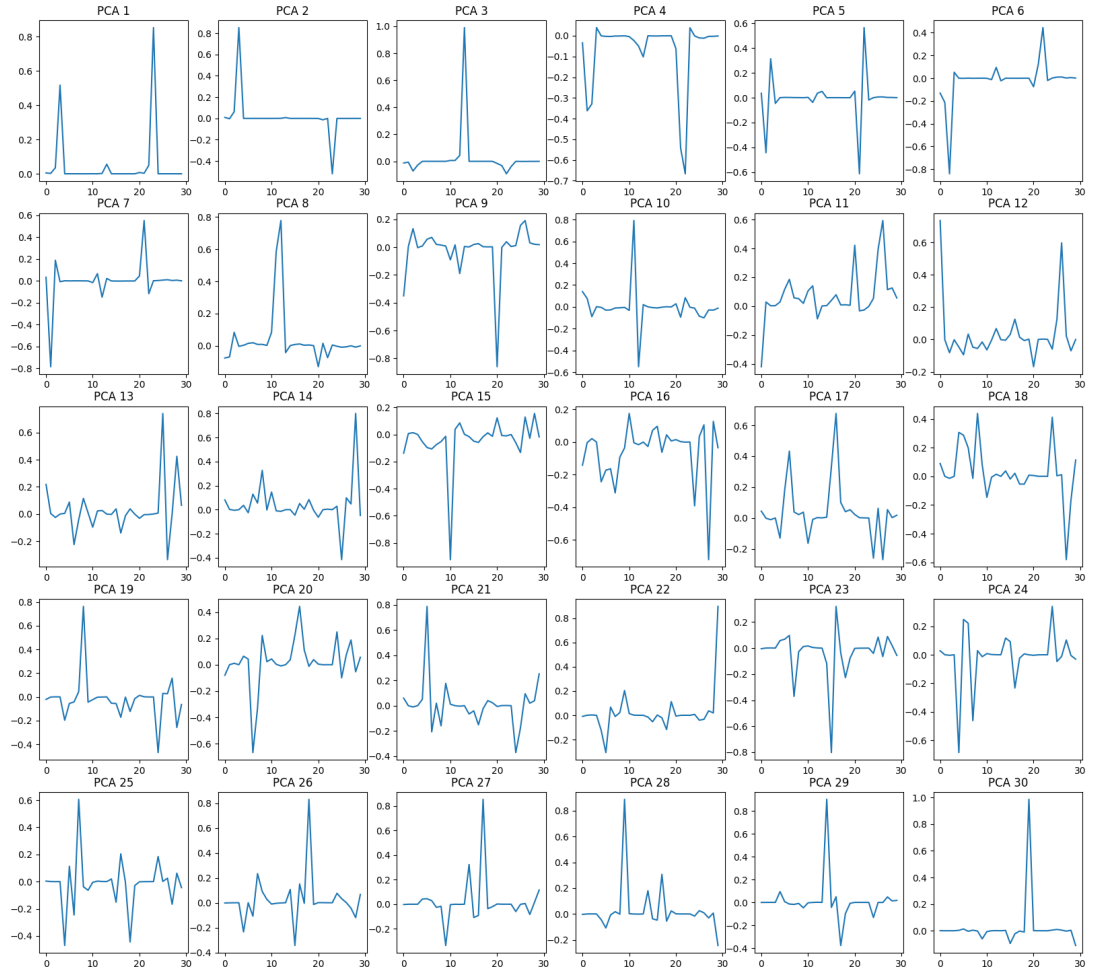
3   Section 2.1

Figure 2: 2_1_1.png

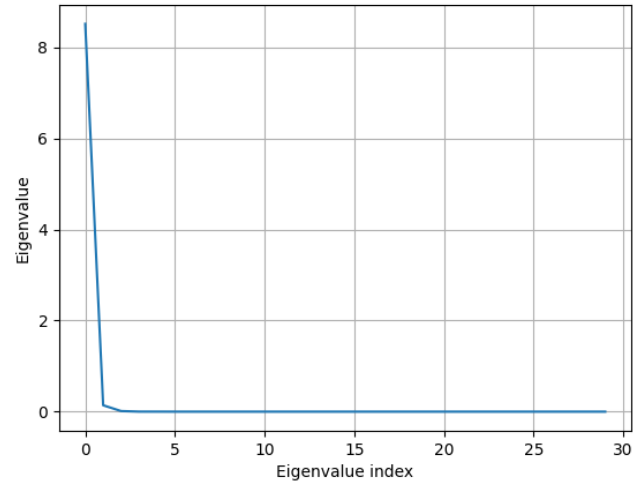# 4 Section 3.1.1

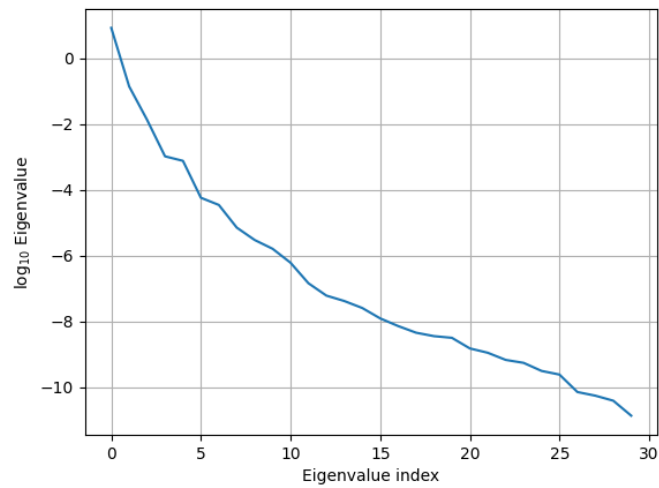Figure 3: 3_1_1.png

5 Section 3.1.2

Figure 4: 3_1_2.png
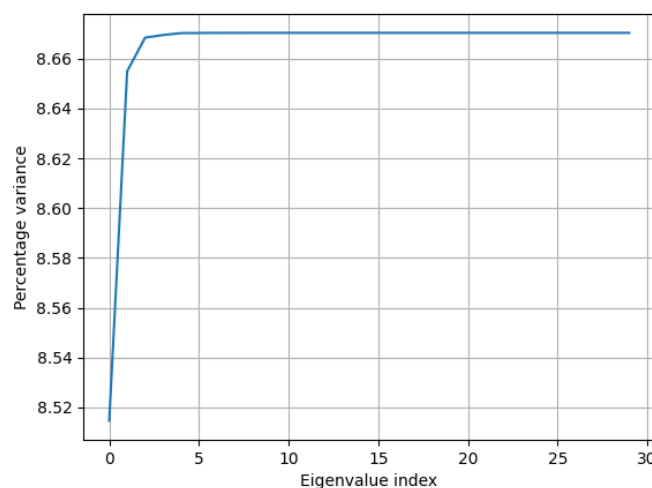
6   Section 3.1.3

Figure 5: 3_1_3.png

7   Section 3.1.4

Eigenvalues are coefficients applied to eigenvectors that give the vectors their length or magnitude.

The explained variance will tell what is the contribution to the percent variance of the entire model contributed by each components so in other words how much does the PC_i contribute to the overall variance of the data set.

The cumulative variance meaning that at each point of the graph it is going to show the sum of each variance with the last. So for point PC1 we have just the variance of PC1 while for PC2 it will be variance of PC1 and variance of PC2 and so on.

Taking in consideration *3_2_1* and *3_2_3* we will see that the cumulative is the sum of all the variance, and it is right.

In this case the results we got are right because the dataset has a sparse value.

8   Indep

In this independent part I decided to change the dataset so I used the Iris dataset. The plot I got follows:
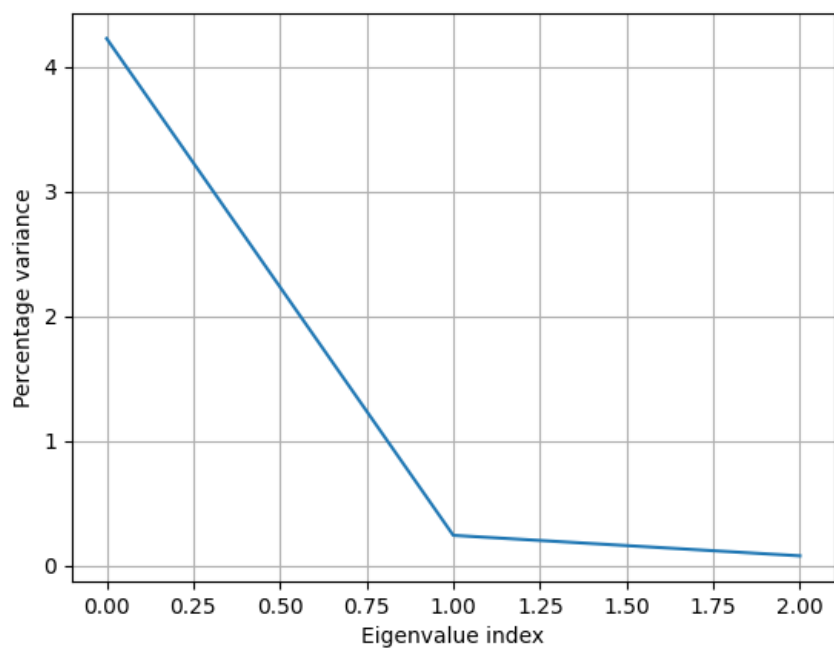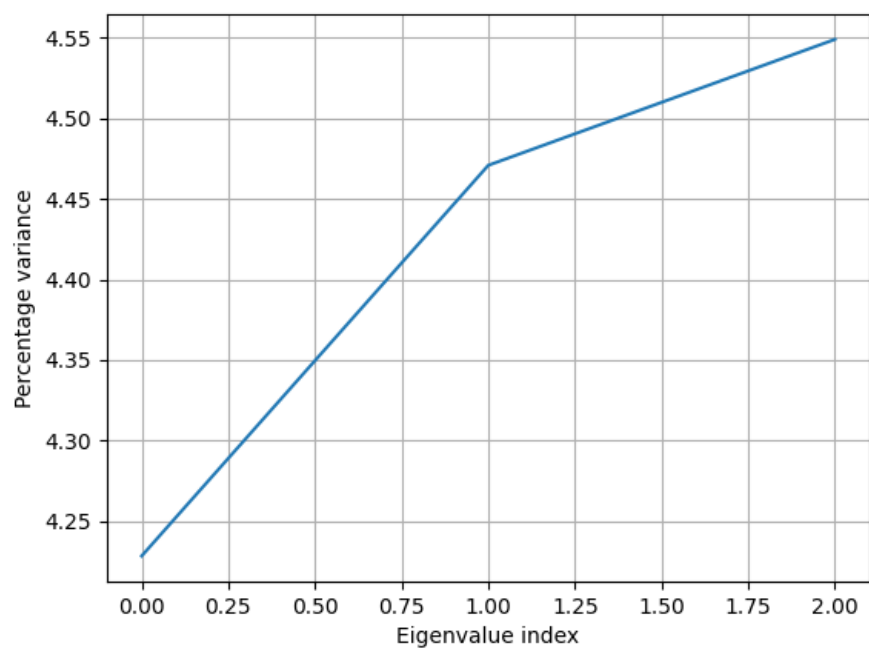
Figure 6: Variance for Iris dataset

Figure 7: Cumulative variance for Iris dataset