

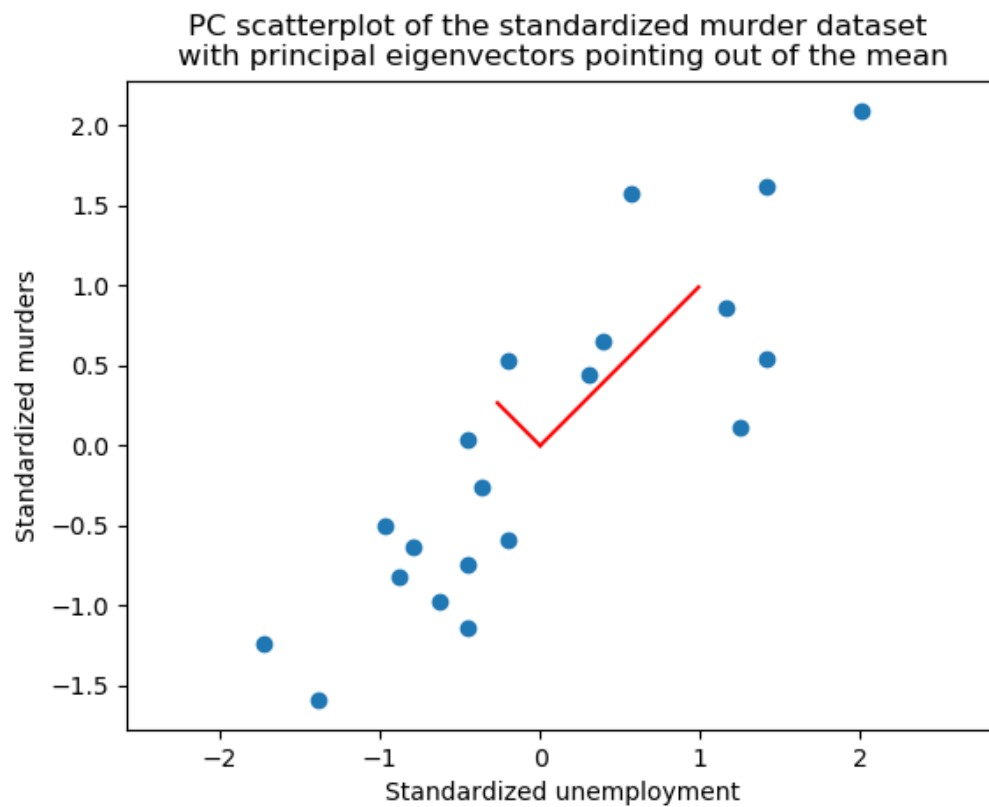
# Stefano Pellegrini - Assignment 3

March 4, 2020

## Exercise 1

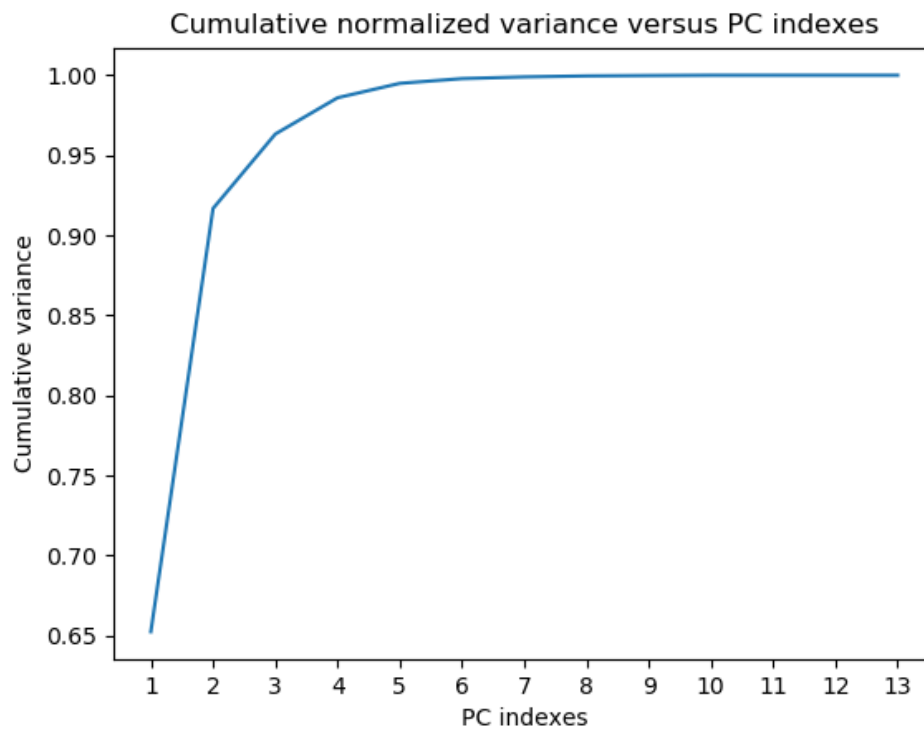
### 1.2. PC scatterplot of the standardized murders dataset

I normalized the murders dataset because its two features are in totally different units.

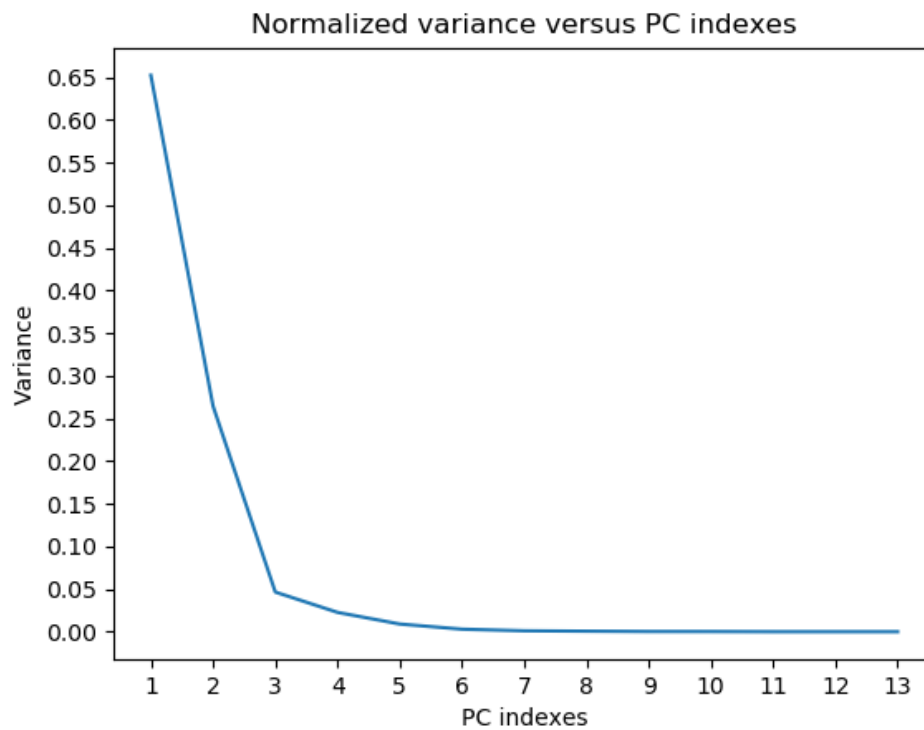


### 1.3. Variance and PCs (pesticides centered training dataset)

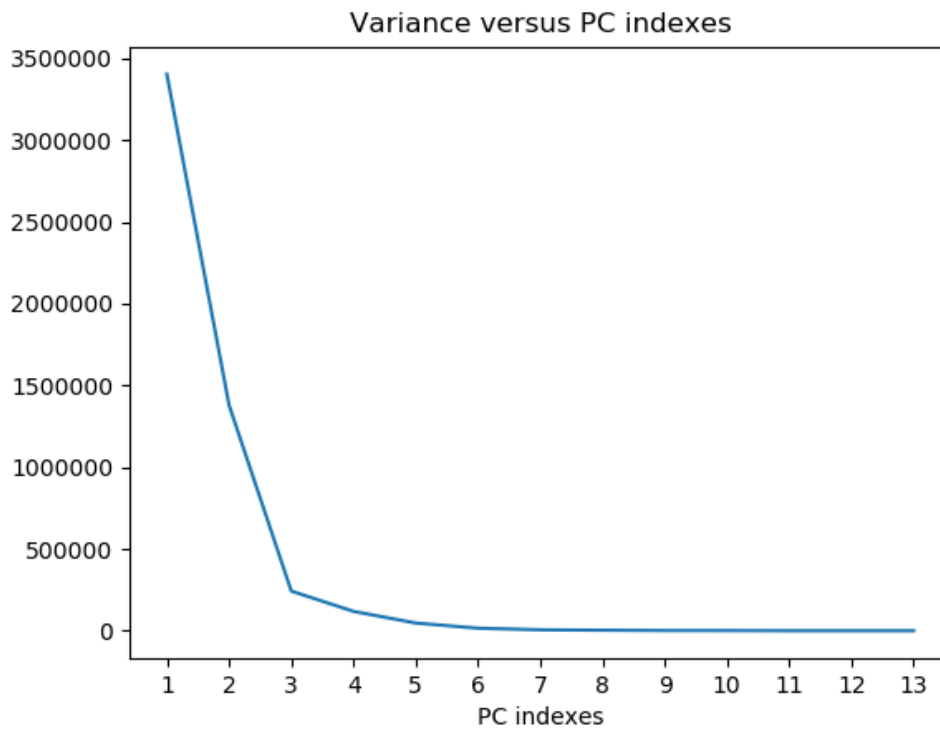
- Plot of cumulative normalized variance versus PC



- Plot of normalized variance versus PC



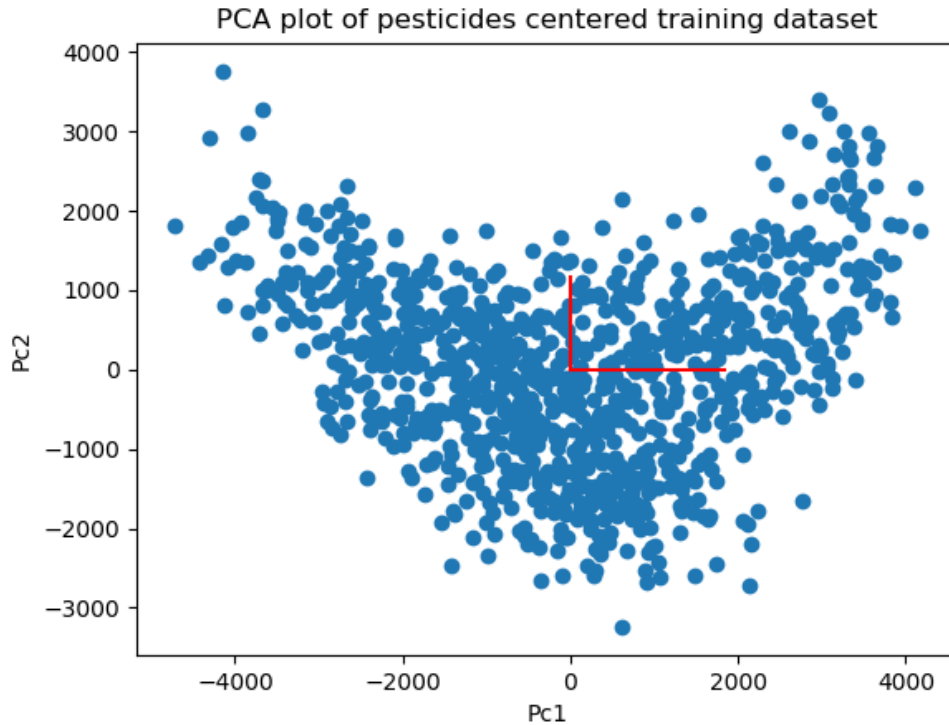
- Plot of variance versus PC



- The numbers of dimensions needed to capture 90% and 95% in pesticides training dataset
  - To capture 90% of variance we need 2 PCs
  - To capture 95% of variance we need 3 PCs

## Exercise 2

- PCA plot of pesticides centered training dataset  
I added the two principal eigenvectors in red, they define the space where the data is now projected.



### Exercise 3

- **Description of software used**

I used my own implementation of the unsupervised k-means clustering algorithm. As requested in the exercise, I initialized the algorithm placing the k centroids at the k first points of the dataset. To evaluate the quality of the clusters I used my implemented loss function that computes the sum of the squared errors among them. Based on the distances between the points and the centroids, I assign each point to the nearest cluster. Then I update the new centroids computing the mean of the new assigned points, and finally, I compute the loss of the new clusters. The algorithm continues this process until the centroids converge, which occurs when the loss converge.

- **Two cluster centers obtained from centered pesticides training dataset**

– The first centroid =

(1.86626496e+00, 1.87792821e+01, 3.29138120e+02, 1.30606740e+03, 1.06726675e+02, -6.72401222e+02, -3.77579821e+02, -2.66757684e+02, -2.26995889e+02, -1.36750966e+02, -6.75217692e+01, -1.33063846e+01, -1.26669231e+00)

– The second centroid =

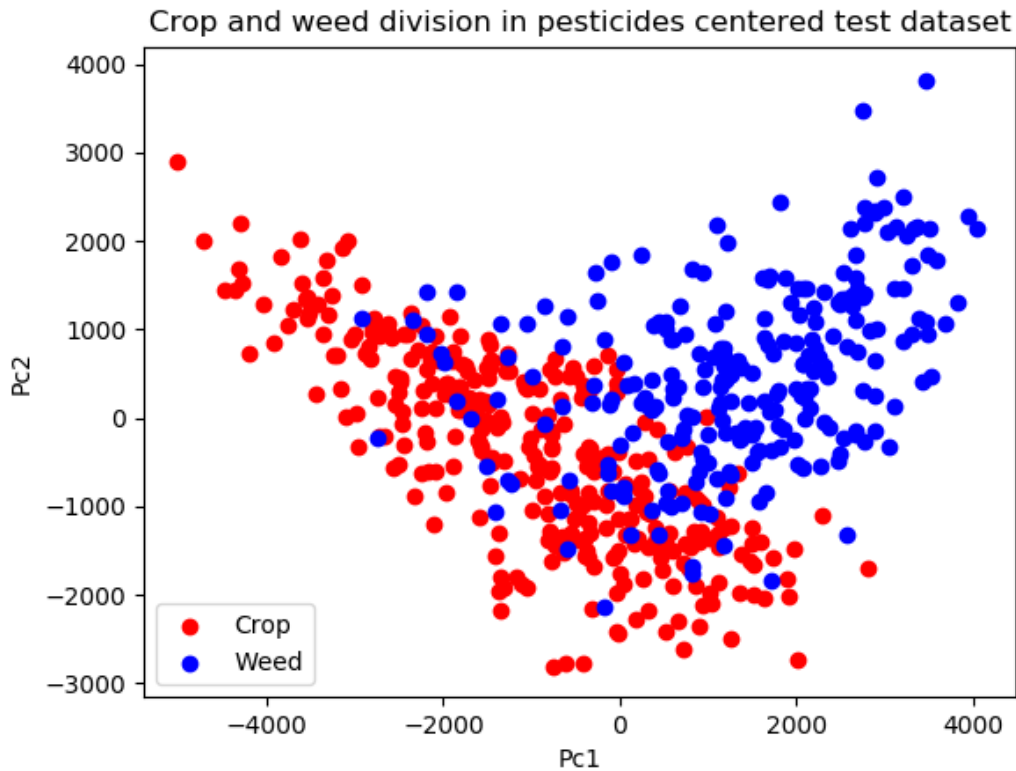
(-1.64175188e+00, -1.65201203e+01, -2.89542556e+02, -1.14894651e+03, -9.38873759e+01, 5.91510850e+02, 3.32156684e+02, 2.34666534e+02, 1.99688113e+02, 1.20299722e+02, 5.93988496e+01, 1.17056165e+01, 1.11430827e+00)

## Extra - Clusters prediction and plotting

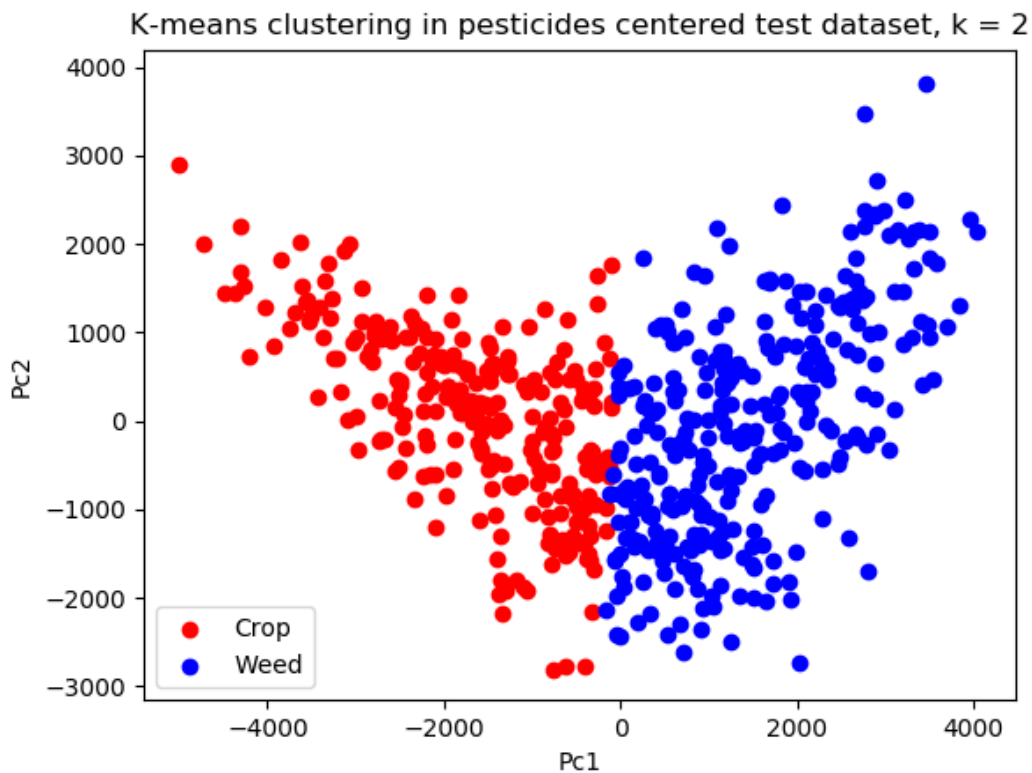
### 0.1 Plot the original dataset and the predicted clusters

I added this extra and not requested exercise to evaluate the performance of my k-means clustering algorithm. I used the centroids obtained from the pesticides training data to identify clusters in the pesticides test dataset. First I plot the test dataset with real classes division, then I plot the predicted clusters. Finally I compare the predicted clusters to the real classes of the data to obtain the accuracy of the prediction.

- Plot classes division of pesticides test dataset



- Plot test set clustering



- Compute accuracy of the prediction on pesticides test dataset

The accuracy of the model on the pesticides test dataset is 76.48%.

## Exercise 4

### 4.1. How is probability interpreted differently in the frequentist and Bayesian views?

The Bayesian view interprets probability as a degree of belief or a measure of certainty, the alternative frequentist view interprets probability as a frequency.

**4.2. Cheap, efficient computers played a major role in making Bayesian methods mainstream. Why?** The posterior probability is often unavailable in closed analytic form but it can be approximated using Monte Carlo methods based on sampling, or by variational Bayesian methods. These methods are computationally heavy and require fast computers to process data.

**4.3. What is the difference between a Bayesian credible interval and a frequentist confidence interval?** A frequentist confidence interval is not a probability, it is a statement, based on a large number of repeated experiments, about the frequency of the parameter being in the interval. In this interpretation the data is random and the parameter is fixed. In the Bayesian view, a certain confidence interval represents the probability, based on the posterior distribution, that the parameter value given the observed data is contained in the interval. In this alternative interpretation, the parameter is random and the data is fixed.

**4.4. How does a maximum likelihood estimate approximate full Bayesian inference?**

In Bayesian inference is possible to define a loss function that measures the cost of using a point estimate instead of the true parameter value. A maximum likelihood estimate can approximate the full Bayesian inference if we assume a zero-one loss function and uniform prior, so all values of the parameter should be equally likely. This is often a good approximation but in certain cases it can be badly wrong.

**4.5. When will point estimates be a good approximation of full Bayesian inference?**

A point estimate can be a good approximation if the distribution of the parameter is a unimodal distribution with one clear peak.