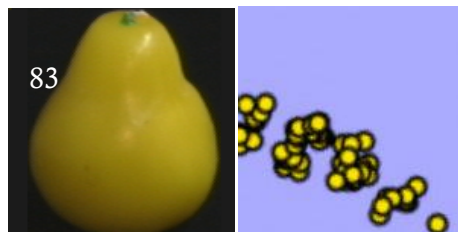
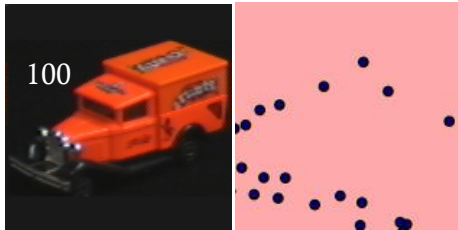


# Homework 1: PCA

## Categories

The chosen categories of COIL-100 are:

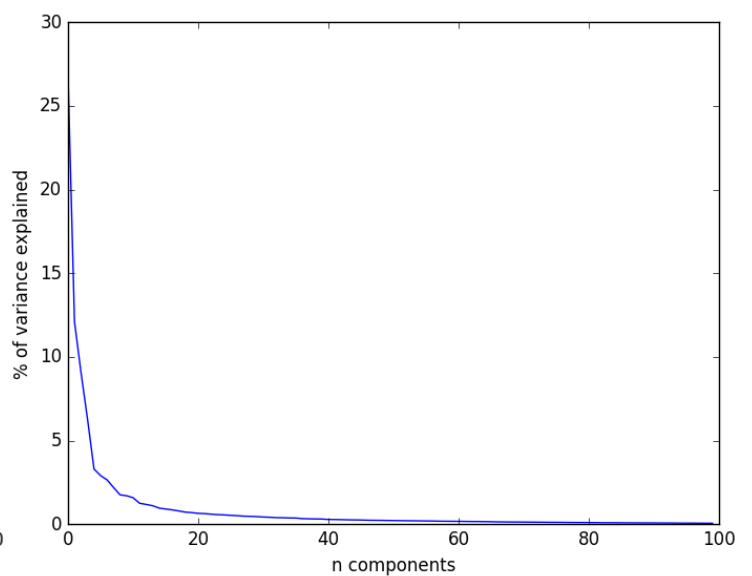
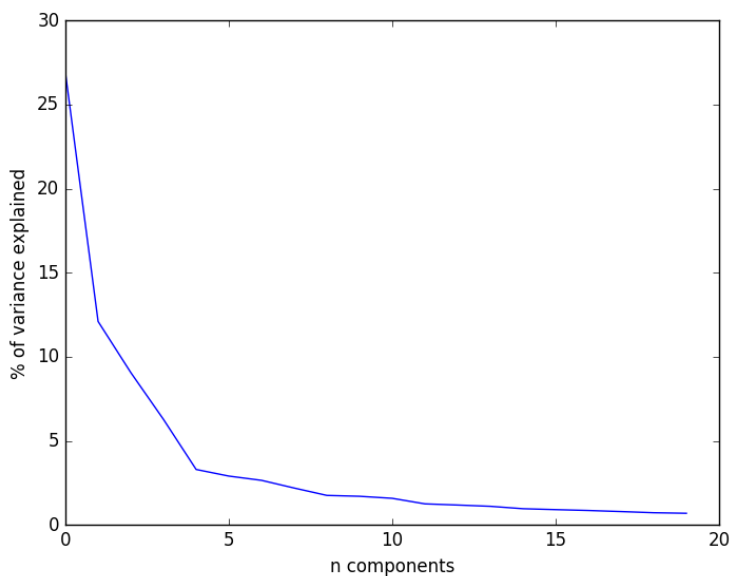


## Data preparation

Following the provided instructions all images (72 samples \* 4 classes) are converted into vectors, then the 288 vectors are organized in a matrix, each row corresponds to an image, each column to a feature, and finally all features are standardized.

## PCA

The chosen configurations for the analysis are four, everyone considers two principal components (1-2, 3-4, 19-20, 99-100). As expected by the theory the fraction of explained variance by each principal component decrease with the grade of the component, this trend is evident in the following plots:



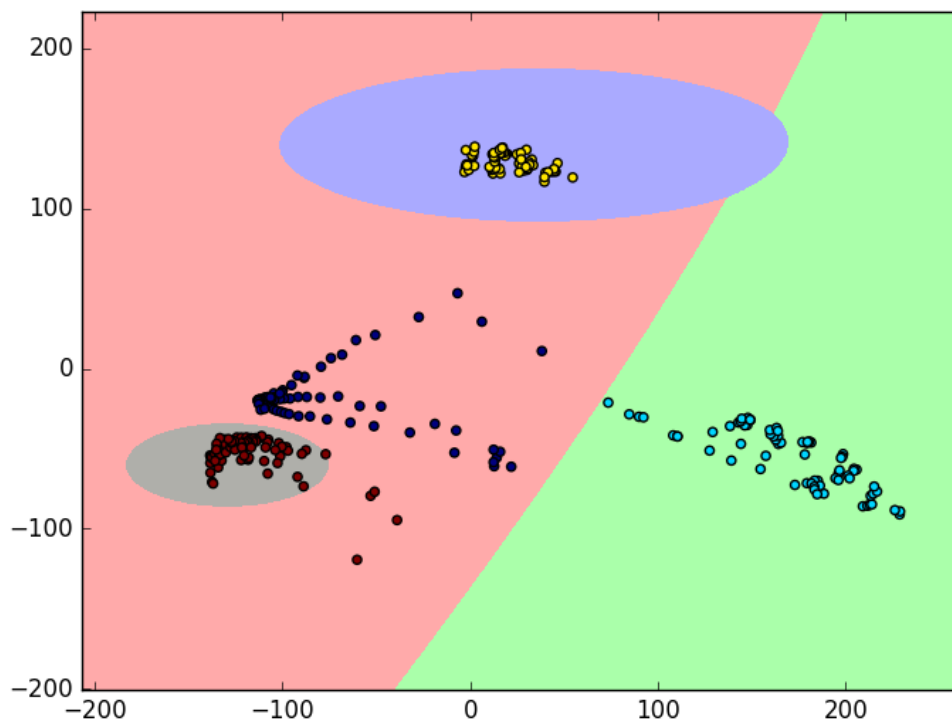
The total explained variance until the  $n$ th component is shown in the table below, this number is fundamental in the choice of the number of principal components to consider in a reduction of the dimensionality to preserve the required information. Instead if the aim is the visualization of the data, the number of principal components is fixed to two or three.

<i>Until:</i>	<i>2nd</i>	<i>4th</i>	<i>20th</i>	<i>100th</i>
<i>Total Explained variance</i>	38.99	54.33	78.92	96.80

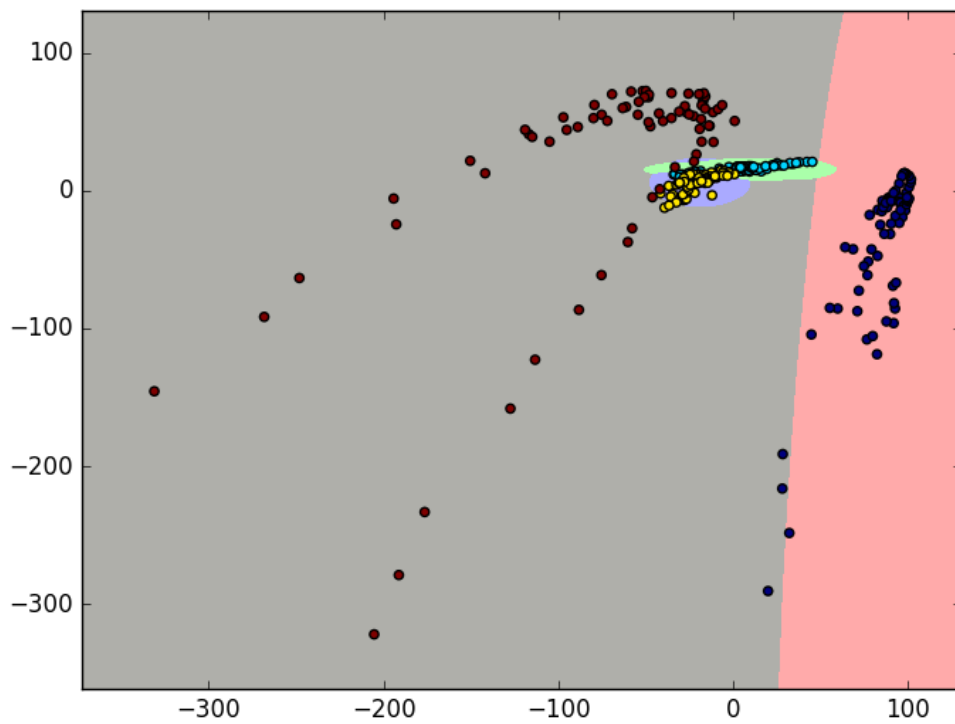
Before the scatter plots of the couples of principal components, in the table below are shown the performances of the Naïve Bayes classifier, as expected the lower the component the higher the performance:

<i>Principal comonents</i>	<i>1-2</i>	<i>3-4</i>	<i>19-20</i>	<i>99-100</i>
<i>Score of Naïve Bayes classifier</i>	0.97	0.87	0.56	0.41

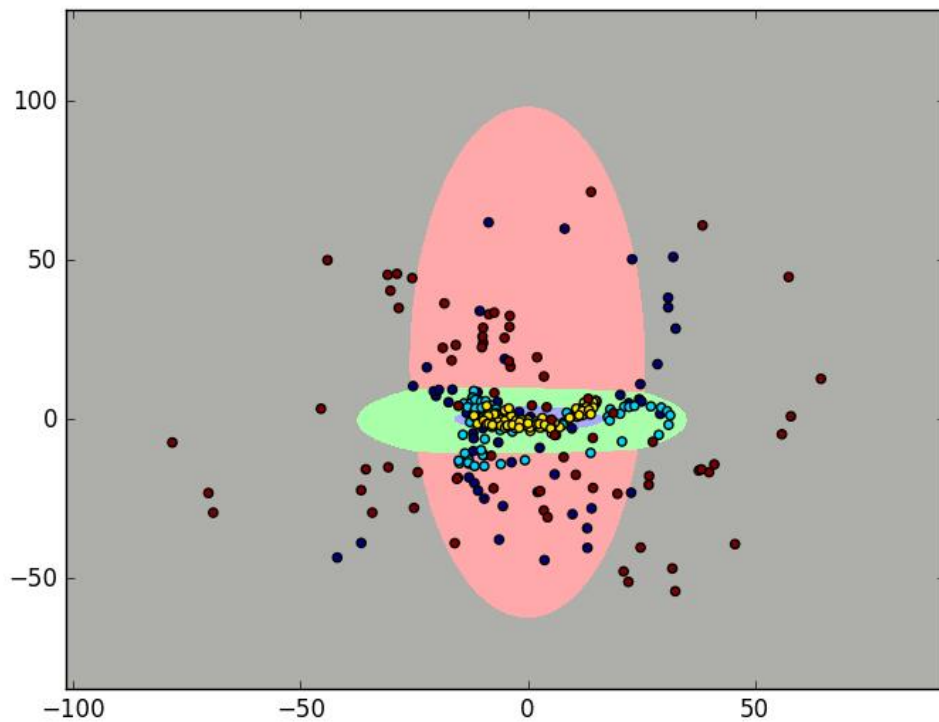
In the following plot are shown the first two principal components (1<sup>st</sup> on x, 2<sup>nd</sup> on y) and the decision boundaries of a Naïve Bayes classifier, the classification is more than acceptable:



In the following plot are shown the third and fourth principal components (3<sup>rd</sup> on x, 4<sup>th</sup> on y) and the decision boundaries of a Naïve Bayes classifier, the classification is decent:



In the following plot are shown the nineteenth and twentieth principal components (19<sup>th</sup> on x, 20<sup>th</sup> on y) and the decision boundaries of a Naïve Bayes classifier, the classification is really poor:



In the following plot are shown the ninety-ninth and hundredth principal components (99<sup>th</sup> on x, 100<sup>th</sup> on y) and the decision boundaries of a Naïve Bayes classifier, the classification is really poor again. As

expected the boundaries of the distributions become smaller with the growth of the grade of the principal component chosen:

