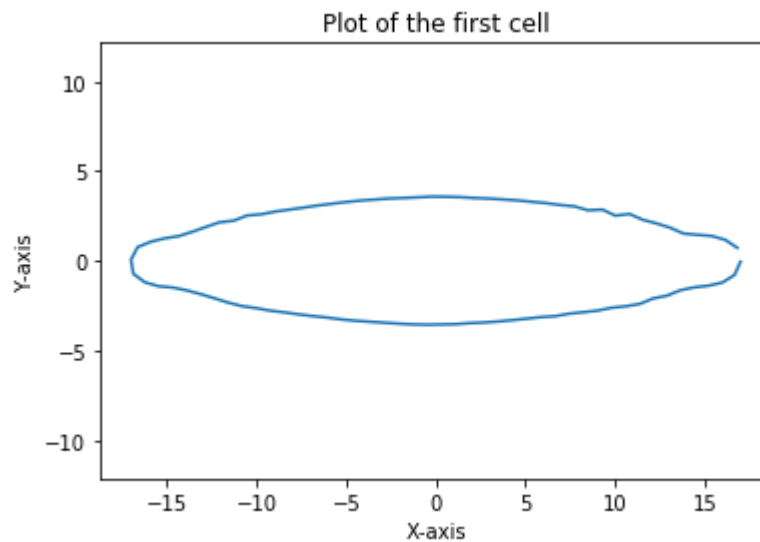# IDS - Assignment 4

Bjarke Kingo Iversen - SWD457

March 2018
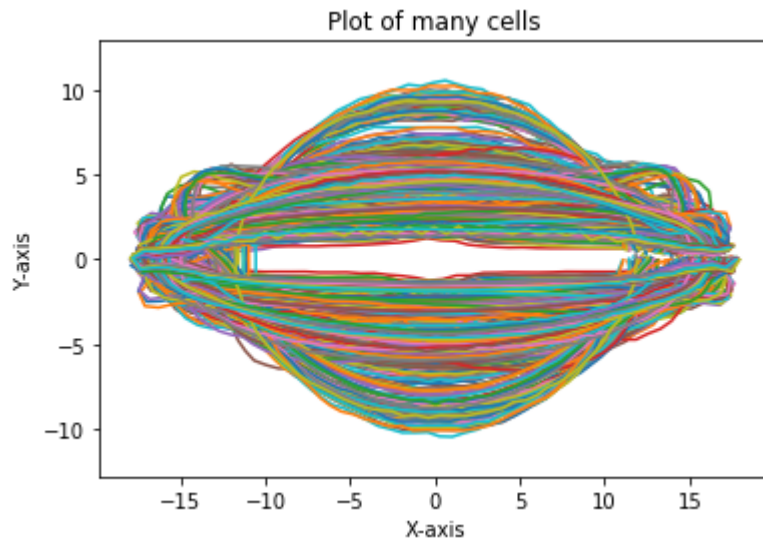
Introduction

For all my code, please see the Jupyter notebook file **Assignment 4.ipynb** in the **src.zip** folder.
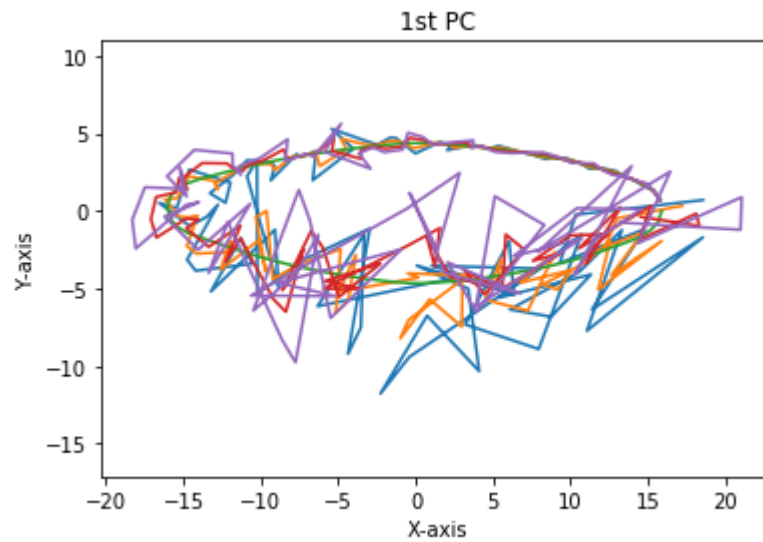
1 Plotting cell shapes

Here we see one cell. One thing we notice is that it is fairly longer on the x-axis.

Plot of many cells

We see that we actually have quite some variation, some of cells being long and thin, some of them having a larger body stretching along the y-axis.
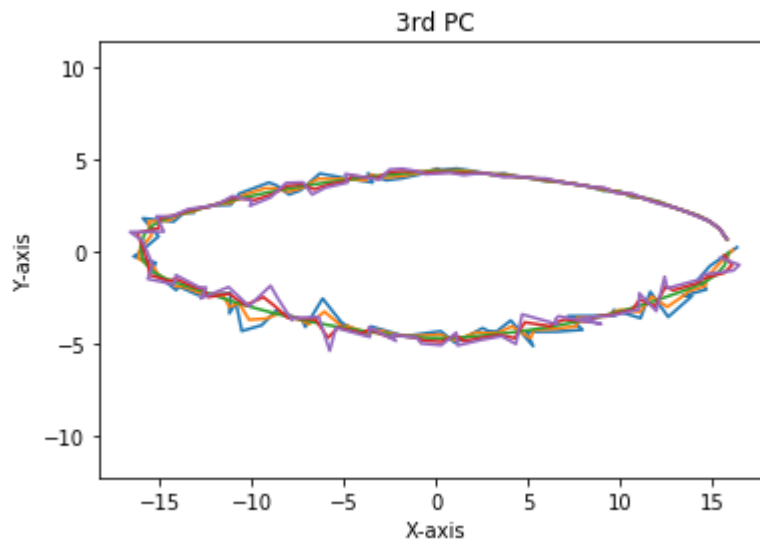
2 Visualizing variance in visual data

1st PC

The first component seem to capture alot of variance as one would expect. This variance is seen with the farily large irregular lines indicating the cell-wall.

2nd PC
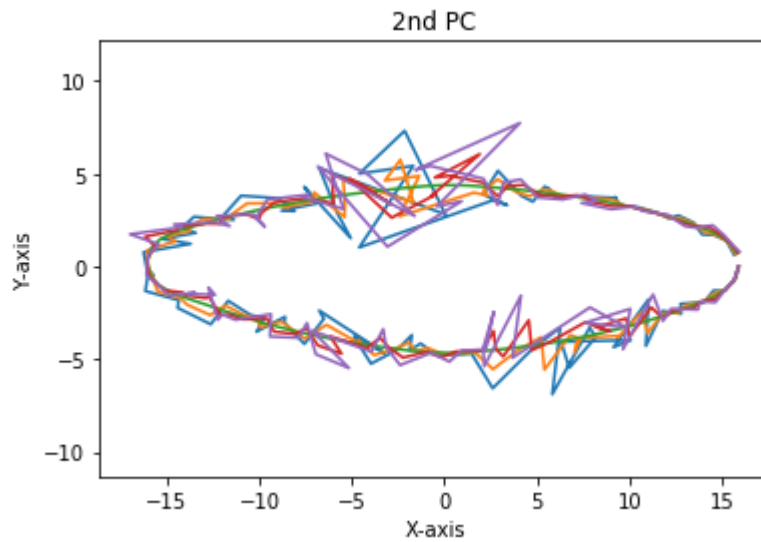
Less variance is captured by the second component.

3rd PC

At the third component we see that most of the variance is gone, and we get something that really looks alike a mean-cell.

3 Critical thinking

3.1 a

Assume that you perform each of the following preprocessing steps prior to performing PCA. What is the effect on the PCA result? Is it a good idea?

- **Centering**: Doesn't affect the PCA since we implicitly does the centering when computing the covariance-matrix.

- **Standardization** This could be useful when you want units to 'weight' the same.

- **Whitening** For some algorithms whitening may be necessary in order to make data-input less redundant

3.2 b

4 Clustering II
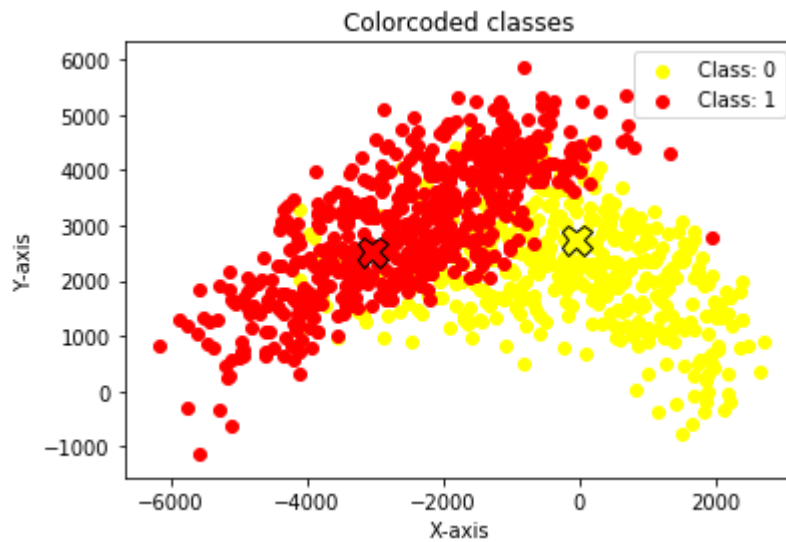
For calculating the clusteres i used KMeans algorithm from sklearn. We see 2 clusters and the color-coded classes. We see that there definitely is some sort of seperation between the classes looking at the clusters.