Assignment: Foundations of Machine Learning (AMMI)

African Institute for Mathematical Sciences

Quantum Leap Africa

Foundations of Machine Learning

Author: Lionel Ngoupeyou Tondji

Date: November 13, 2018

Density Estimation with Gaussian Mixture Models

Task 1

See the python file.

Task 2

You are given a two-dimensional dataset of about 30 geolocations (latitude/-longitude) of birthplaces. Model the data as well as possible. Describe your approach, justify your choices and interpret your results.

First of all we are trying to plot the two-dimensional dataset, to see how they look like.

The next task that we perform is to normalize our data set, So we will reduce by the mean and subtract by the standard deviation in order to have numerical stability. We obtain the following graph :

We create a function that help us visualize the locations and shapes of the Gaussian Mixture Model clusters by drawing ellipses based on the Gaussian Mixture Model output:

The first one we consider 2 clusters :

The second one we consider 3 clusters :

The third one we consider 4 clusters :

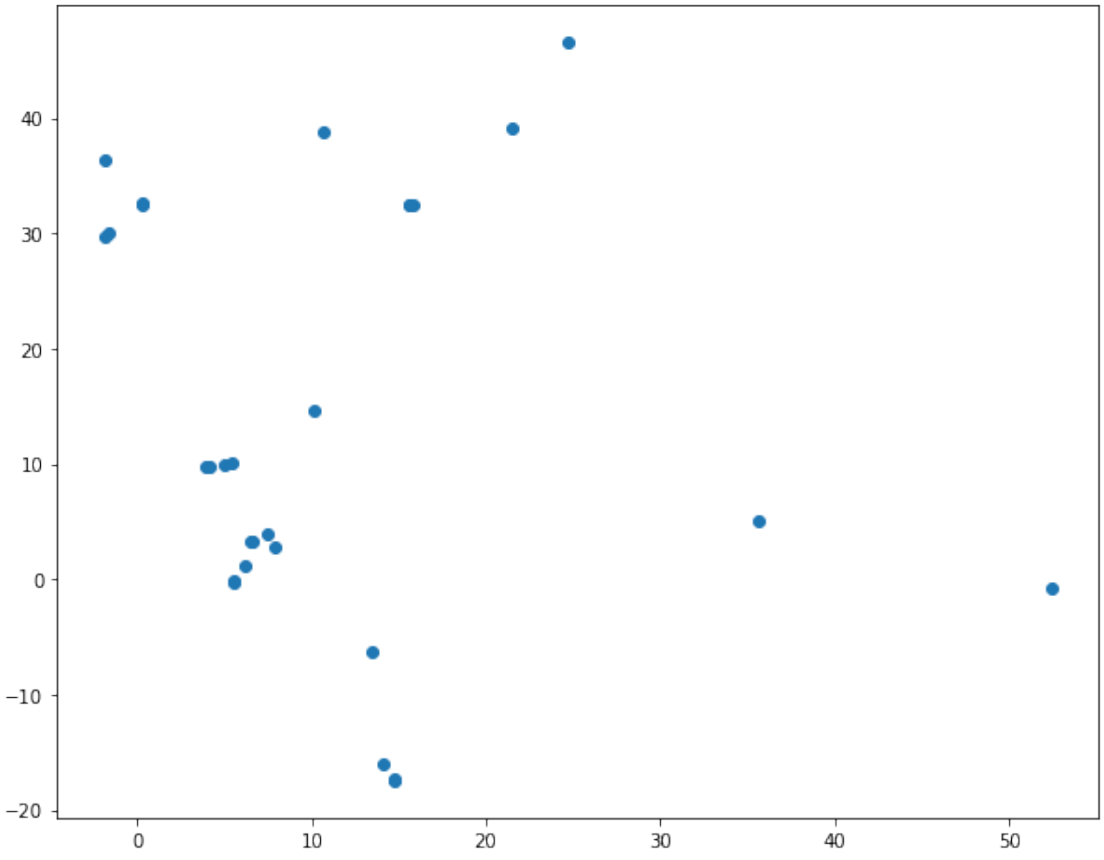

Figure 1: visualizing our data set

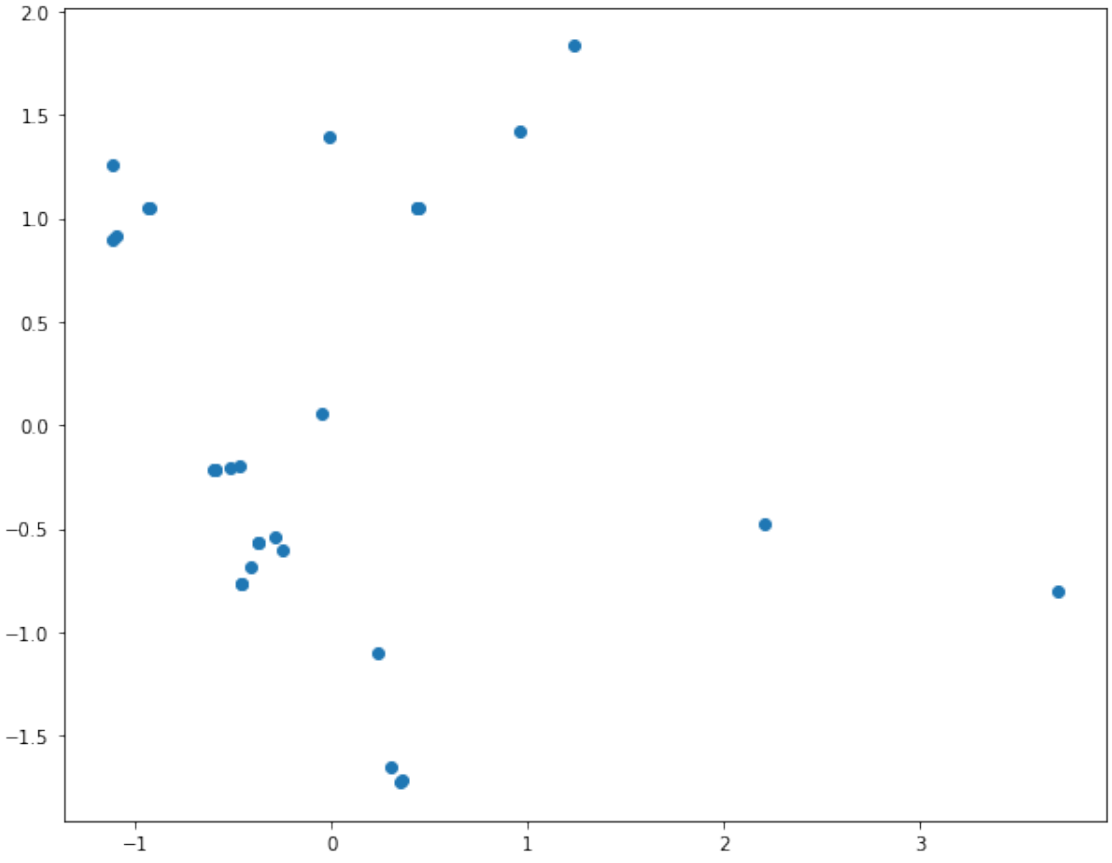

Figure 2: normalized Data set

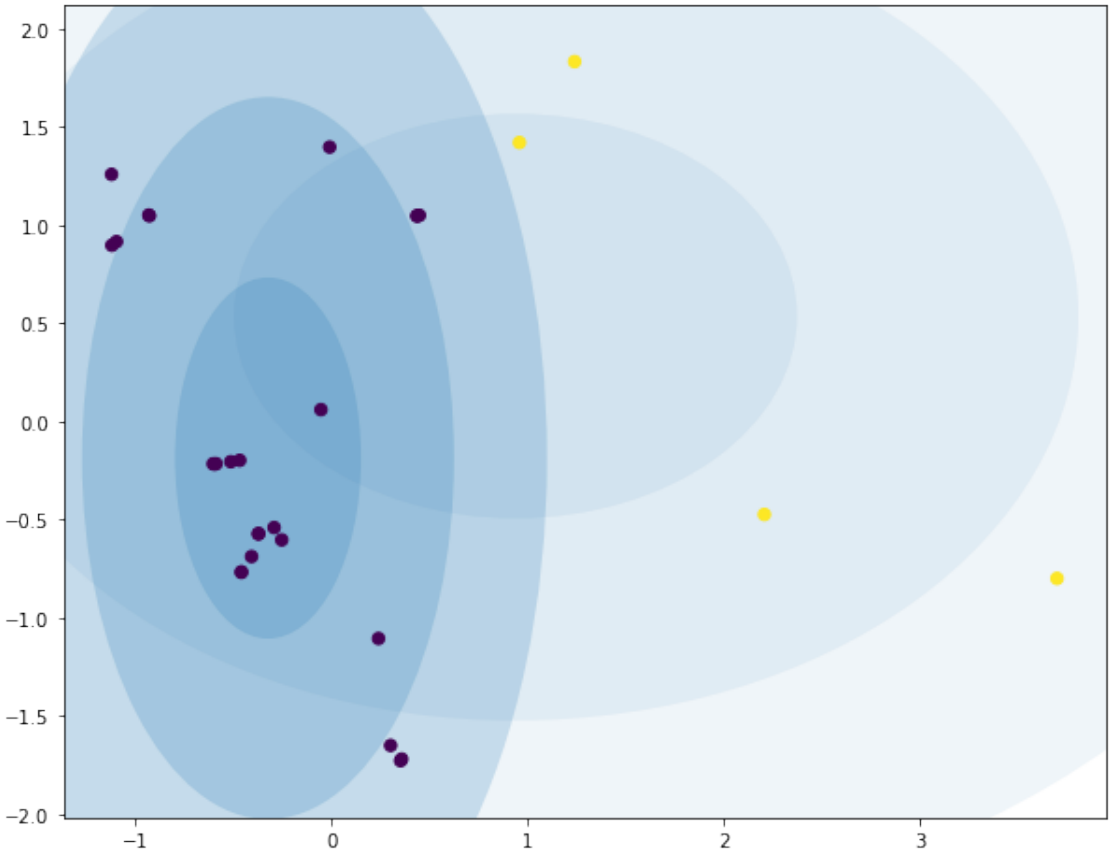

Figure 3: Ellipses based on the 2 Gaussian Mixture Model output

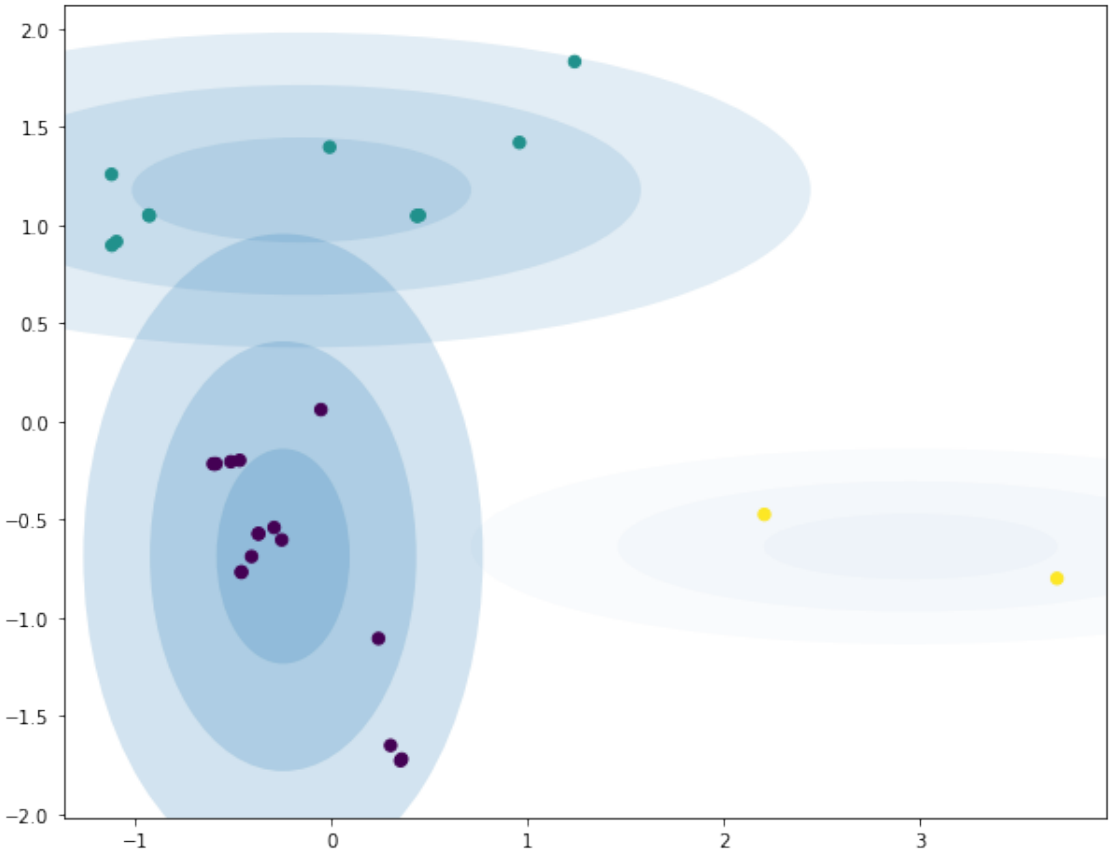

Figure 4: Ellipses based on the 3 Gaussian Mixture Model output

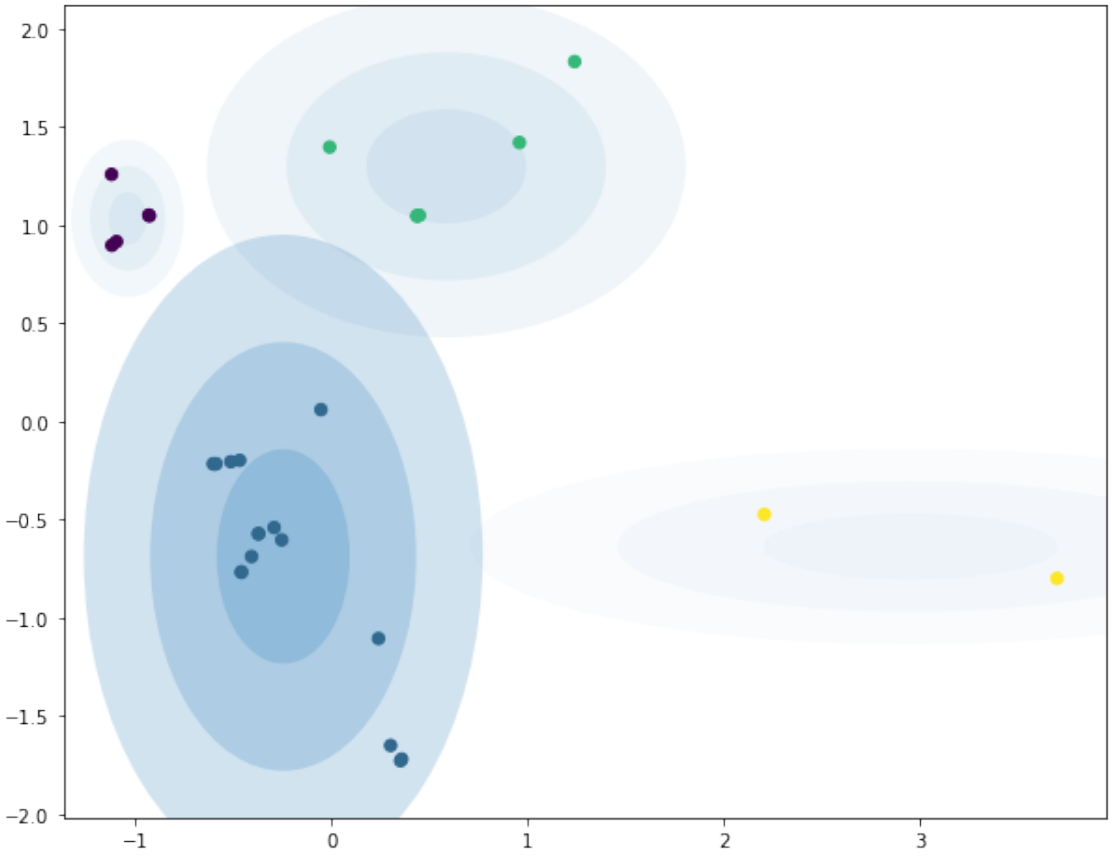

Figure 5: Ellipses based on the 4 Gaussian Mixture Model output

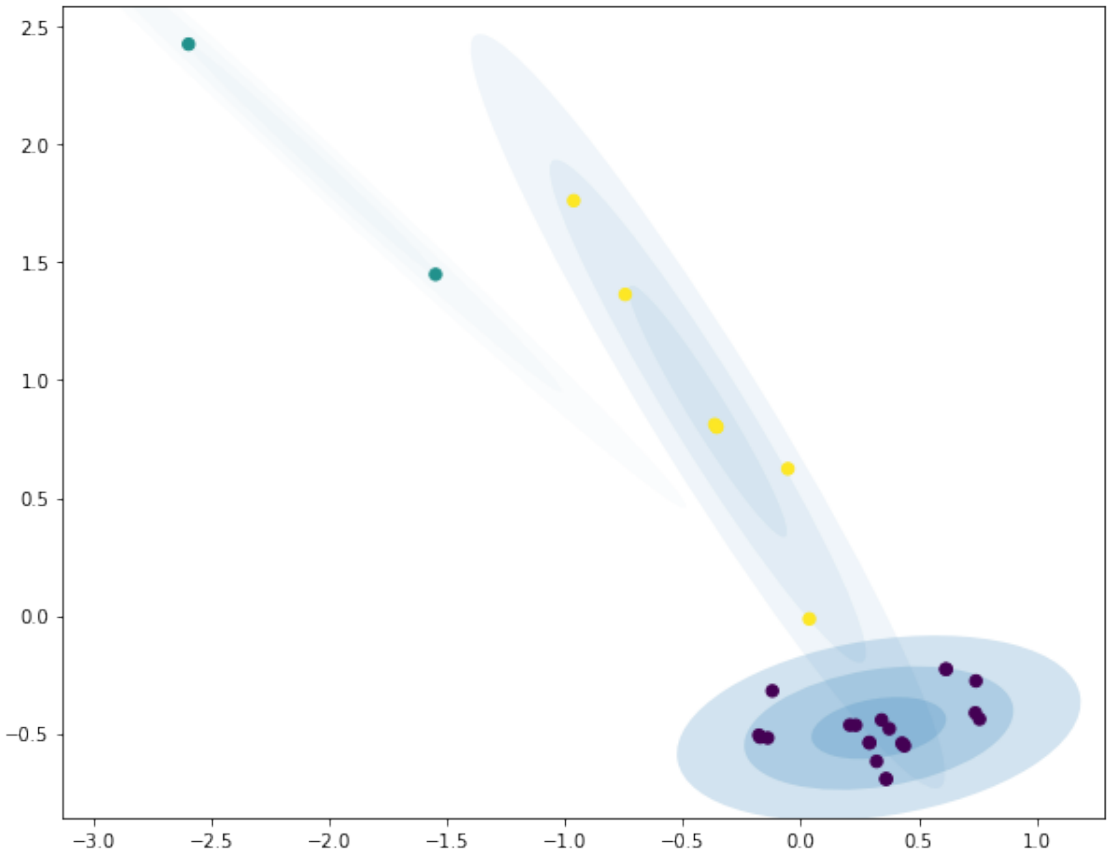

Figure 6: Ellipses based on the 3 Gaussian Mixture Model output where we apply the transformation

We take the same data and transform it, we recognize that these transformed clusters are non-circular, and thus circular clusters would be a poor fit.

we can use the Gaussian Mixture Model approach to fit our stretched dataset; allowing for a full covariance the model will fit even very oblong, stretched-out clusters:

The first one we consider 3 clusters :

The second one we consider 4 clusters :

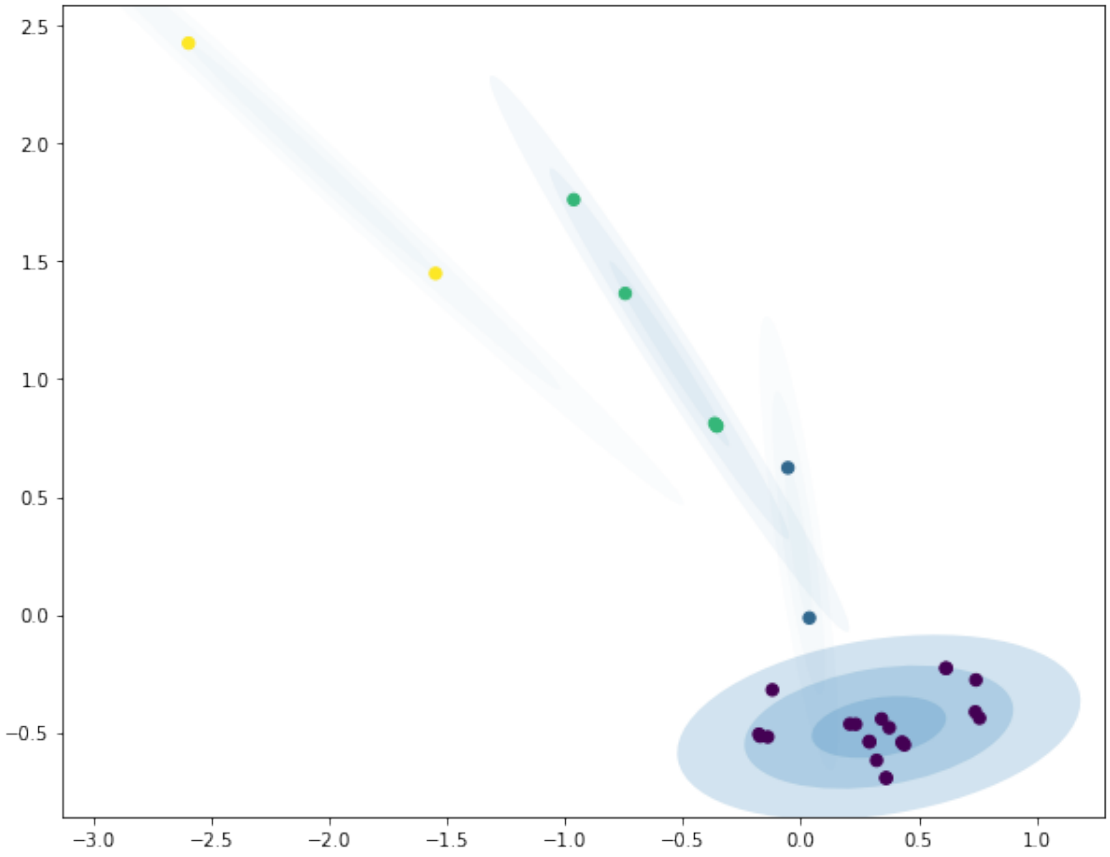

Figure 7: Ellipses based on the 4 Gaussian Mixture Model output where we apply the transformation

We can see in the graph that if we consider just 2 mixture model , we have collision with the two ellipses but a good value for $k = 3$ from this value we can see that the ellipses are well drawn. Also for $k = 4$ it's perform very well and this can be view as over-fitting.