EM Algorithm Steps

To Implement EM on Gaussian Mixture Models:

EM-Algorithm

The EM algorithm is divided in 2 steps. E and M.

E-Step(Estimation Step):

We calculate the w matrix. It is calculated using the following formula:

$$w_{nk} = \frac{\pi_k N(x_n|\,\mu_k, \Sigma_k)}{\sum_j \pi_j N(x_n|\,\mu_j, \Sigma_j)}$$

π : pi list contains the fraction of the dataset for every cluster

Multivariate Normal formula:

$$f(x;\,\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^d * det(\Sigma)}} * e^{-\frac{1}{2} * ((x-\mu)^T inv(\Sigma)(x-\mu))}$$

d : dimension

μ: distance means

Σ: covariance matrix

M-Step(Maximization Step):

We will step-by-step set the value for the mean vectors and covariance matrices to describe with them the clusters. To do that we will use the following formulas:

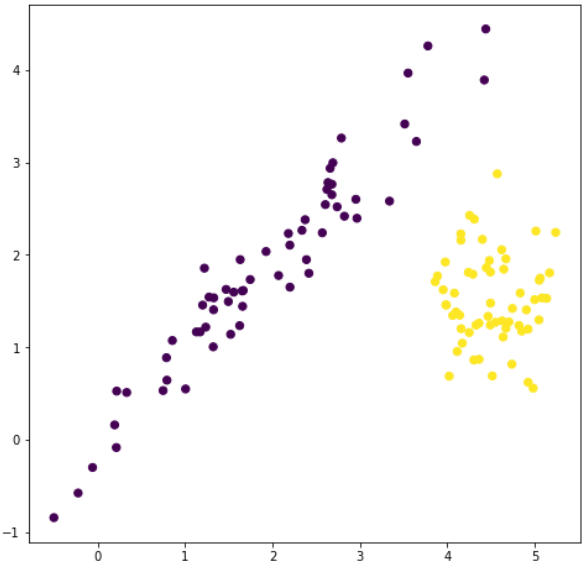
$$\mu_{\boldsymbol{k}} = \frac{\mathbf{1}}{\boldsymbol{N_k}} \sum_{\mathbf{n=1}}^{\mathrm{N}} w_{nk} \boldsymbol{x_k}$$

$$\Sigma_{\boldsymbol{k}} = \frac{\mathbf{1}}{\boldsymbol{N_k}} \sum_{\mathbf{n=1}}^{\mathrm{N}} w_{nk} (\boldsymbol{x_n} - \mu_{\boldsymbol{k}})(\boldsymbol{x_n} - \mu_{\boldsymbol{k}})^T$$

$$\pi_{\boldsymbol{k}} = \frac{\boldsymbol{N_k}}{\boldsymbol{N}}$$

$\boldsymbol{N_k}$ **:** N list, in which each element is basically the sum of the correspondent column in the w matrix.

Plot EM for K = 2



Plot EM for K = 3