Machine Learning
Exercise 6

Marc Toussaint
Machine Learning & Robotics lab, U Stuttgart
Universitätsstraße 38, 70569 Stuttgart, Germany

June 5, 2018

1 Clustering the Yale face database

On the webpage find and download the Yale face database http://ipvs.informatik.uni-stuttgart.de/mlr/marc/teaching/data/yalefaces_cropBackground.tgz. The file contains gif images of 136 faces.

We'll cluster the faces using k-means in $K = 4$ clusters.

a) Compute a k-means clustering starting with random initializations of the centers. Repeat k-means clustering 10 times. For each run, report on the clustering error $\min \sum_n \sum_k r_{nk} \|x_n - \mu_k\|^2$ and pick the best clustering. Display the center faces μ_k and perhaps some samples for each cluster.

b) Repeat the above for various K and plot the clustering error over K.

c) Repeat the above on the first 20 principal components of the data. Discussion in the tutorial: Is PCA the best way to reduce dimensionality as a precursor to k-means clustering? What would be the 'ideal' way to reduce dimensionality as precursor to k-means clustering?

2 Mixture of Gaussians

Download the data set `mixture.txt` from the course webpage, containing $n = 300$ 2-dimensional points. Load it in a data matrix $\boldsymbol{X} \in \mathbb{R}^{n \times 2}$.

a) Implement the EM-algorithm for a Gaussian Mixture on this data set. Choose $K = 3$ and the prior $\pi_k = 1/K$. Initialize by choosing the three means μ_k to be different randomly selected data points x_i (i random in $\{1, .., n\}$) and the covariances $\Sigma_k = \mathbf{I}$ (a more robust choice would be the covariance of the whole data). Iterate EM starting with the first E-step based on these initializations. Repeat with random restarts—how often does it converge to the optimum?

b) Do exactly the same, but this time initialize the posterior γ_{ik} randomly (i.e., assign each point to a random cluster: for each point x_i select $k' = rand(1 : K)$ and set $\gamma_{ik} = [k = k']$); then start EM with the first M-step. Is this better or worse than the previous way of initialization?