Theoretical Neuroscience II
Computational Neuroscience II
K-means
(10 points)

1 K-means Clustering (6P)

Compute a k-means clustering for two classes in an unsupervised fashion. Given are the following two-dimensional 14 sample data points:

$$x = \left[\binom{-2}{4}, \binom{-1}{-1}, \binom{-1}{1}, \binom{0}{0}, \binom{1}{-1}, \binom{1}{0}, \binom{1}{1}, \binom{1}{2}, \binom{2}{1}, \binom{2}{2}, \binom{2}{3}, \binom{3}{2}, \binom{3}{4}, \binom{6}{0} \right]$$

Choose an appropriate way to visualize your results.
K-means Algorithm:

1. compute the global mean of all data points
2. arrange k means randomly around the global mean
3. start iterating
   a) assign all sample points to one of the two classes, where

   $$c^* = argmin \left| \underline{x}_i - \underline{\mu}_c \right|$$

   so that every data point is assigned to one class.
   b) calculate the new mean values for each class c

   $$\underline{\mu}_c = \frac{1}{N_c} \sum_{i=1}^{N_c} \underline{x}_i$$

   c) repeat step 3a) and 3b) until the chosen criterion of termination is satisfied (e.g. no change in class assignment)

2 Comprehension Questions (4P)

What kind of problems might occur using k-means clustering? Explain the difference between discriminative clustering (k-means) and maximum likelihood clustering (using a pdf). Mention some preconditions for using maximum likelihood (ML). How would the results of task 1 look like using ML clustering?

Theoretical Neuroscience II
Computational Neuroscience II
Expectation Maximization
(10 Points)

1 Observations with two causes

An observer is asked to locate a faint light that appears at varying positions along the horizontal midline of a display. After each appearance, the observer memorizes the location and waits for a sound signal, which prompts him or her to indicate the remembered location with a cursor. In many trials, however, the observer fails to perceive the faint light and therefore cannot remember its location. In these trials, the observer is asked to guess the location. The results are provided in terms of a list of *error values*, i.e., the positive or negative diverences between true and reported locations. In this situation, the distribution of results will evidently reflect two causes. The first cause is the recollection of a perceived location. We expect that this cause generates a comparatively narrow distribution of error values around the true location. The second cause is the guessing of an unperceived location. It's error values are expected to scatter much more widely around the true location. As we do not know which of the two causes generated each error value, we use expectation maximization to estimate both distributions, that is, both the *perceived* and the *guessing* distribution. somewhere on the horizontal midline of a display.

2 Expectation Maximization

We introduce "expectation maximization" as a method for adjusting a generative model G. Our generative model includes two causes, each producing Gaussian-distributed observations. In this case, G comprises the means $\boldsymbol{g}_{A,B}$ and variances $\Sigma_{A,B}$ of the conditional distributions $p[u|v;G]$ and the prior probabilities $p[A|G] = \gamma_A$ and $p[B|G] = \gamma_B$.

The EM algorithm consists of two alternating steps, the E (expectation) step of inferring "probable causes" from the classification distribution $P[v|\boldsymbol{u};G]$, and the M (maximization) step of computing parameters G from weighted averages over observations $\boldsymbol{u}$.

$$P[A|\boldsymbol{u};G] \qquad P[B|\boldsymbol{u};G] \qquad\qquad G = \{\gamma_A, \boldsymbol{g}_A, \boldsymbol{g}_B, \Sigma_A, \Sigma_B\}$$

Expectation step

$$P[v|\boldsymbol{u};G] = \frac{p[\boldsymbol{u}|v;G]\,P[v;G]}{p[\boldsymbol{u};G]}$$

Maximization step

$$\gamma_v = \langle P[v|\boldsymbol{u};G]\rangle_v \qquad \boldsymbol{g}_v = \frac{\langle \boldsymbol{u}\,P[v|\boldsymbol{u};G]\rangle_v}{\gamma_v} \qquad \Sigma_v = \frac{\left\langle (\boldsymbol{u}-\boldsymbol{g}_v)^2\,P[v|\boldsymbol{u};G]\right\rangle_v}{\gamma_v}$$

3 Assigments

You are provided with the following Matlab files

- *sample.mat* containing the position estimates in variable *mysample*
- *maxexp2.m* to iterate betwween the expectation- and maximization steps
- *goforit.m* the main program (calls *maxexp2.m* and shows results)

In addition, there are the incomplete functions *expect2.m* and *maximize2.m* that are intended to carry out the expectation- and maximization steps.
Complete the provided functions *expect2.m* and *maximize2.m* according to the formulas and estimate the parameters of the two distributions underlying your observations in *mysample*. Report the means and standard deviations of the received distributions.