#homework

- A: EM for mixture of Gaussians
  - Read the description of EM for mixture of Gaussians, and run the code and make sure you understand the results, in this vignette:
    - https://stephens999.github.io/fiveMinuteStats/intro_to_em.html
  - The vignette includes code implementing an EM algorithm to estimate the mixture proportions (pi), assuming the means and variances are known. Following the example here, derive the EM updates for estimating the means (\(\mu\)) and the variances (\(\sigma^2\)) of the mixture components. Implement these updates in a function. Your code should allow the number of components, \(K\), and the intial values for \((\pi,\mu,\sigma^2)\) to be specified.
  - Apply your EM algorithm to the simulated data in the example with \(K=2\) components, and i) demonstrate that the log-likelihood increases every iteration; ii) check how the final estimates compare with the true values used to simulate the data.
  - The EM algorithm is a "hill-climbing" algorithm, that simply increases the log-likelihood every iteration. Such hill-climbing algorithms often get stuck in local optima, and so the final solution can depend on the initial values used. Experiment with running your EM algorithm from different starting points, and see if you can illustrate this behavior. (Sometimes this behavior is most obvious when the number of components is wrong, so try applying your EM algorithm to the simulated data, but setting \(K=3\) or 4.)
  - Apply your EM algorithm to fit a mixture of two Gaussians to second singular vector from SVD applied to the zipcode data on 2s and 3s as produced at: https://github.com/stephens999/stat34800/blob/master/analysis/svd_zip.Rmd . (If you are not familiar with SVD, do not worry for now - use the code in this example). Run your algorithm from multiple initial values and select the solution with the highest log-likelihood. Plot your fitted mixture density on top of the histogram. If you classified the samples into 2s and 3s using this mixture what would be the error rate? (Note that you do not need to use cross-validation here to prevent over-fitting because the training labels were not used during the fitting procedure.)
- B: EM for grade of membership model:
  - derive the updates for an EM-based algorithm for estimating the parameters P and Q in the "topic model" from word count data \(X\).
    - Here \(Q\) is an I by K non-negative matrix with \(\sum_k q_{ik}=1\)
    - and \(P\) is a \(K\) by \(J\) non-negative matrix with \(\sum_j p_{kj}=1\).
  - Hint 1: consider the following cases separately:
    - 1. an EM algorithm to estimate P when Q is known
    - 2. an EM algorithm to estimate Q when P is known.

    - An EM algorithm to jointly estimate P and Q can then just iterate between those steps.
- Hint 2: in the notation from the lecture notes (under #prep for this lecture) the complete data likelihood is given by:
    - $\Pr(X,Z|P,Q) = \prod_i \prod_r \prod_k \prod_j (q_{ik}p_{kj})^{z_{irk} x_{irj}}$
    - And $\Pr(Z|Q) = \prod_i \prod_r \prod_k q_{ik}^{z_{irk}}$
- Note: you are not required to implement the algorithm, only derive the updates.