

COGSCI 118B: Homework Assignment 3,
due: Thursday Oct 31st at the beginning of class
Note this homework is 2 pages

- (1) (6 points) (EM for a 2D Gaussian Model) In this problem, you will follow the steps from the class example taken from Example 2 on page 126 of Duda, Hart, and Stork. We will begin with four data points

$$D = \{x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}\} = \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ * \end{pmatrix} \right\}$$

Using the initial guess for θ^0 from the Duda-Hart-Stork example, compute the first improved estimate of μ_2 .

- (2) (6 points) (Data marginal distribution for Gaussian mixture model) Consider a Gaussian mixture model in which the marginal distribution $p(z)$ for the latent variable z is given by Bishop equation (9.10), and the conditional distribution $p(x|z)$ for the observed variable x is given by Bishop equation (9.11). Show that the marginal distribution $p(x)$, obtained by summing $p(z)p(x|z)$ over all possible values of z , is a Gaussian mixture of the form given in Bishop equation (9.7).

- (3) (6 points) (Mixture distribution mean and covariance) Consider a mixture distribution of the form

$$p(x) = \sum_{k=1}^K \pi_k p(x|k)$$

where you may assume the elements of x are discrete. Denote the mean and covariance of $p(x|k)$ by μ_k and Σ_k , respectively. Show that the mean and covariance of the mixture distribution are given by Bishop equations (9.49) and (9.50) respectively.

- (4) (6 points) (Data marginal distribution for Bernoulli mixture model) Consider the joint distribution of latent and observed variables for the Bernoulli distribution obtained by forming the product of $p(x|z, \mu)$ (given by Bishop equation (9.52)) and $p(z|\pi)$ (given by Bishop equation (9.53)). Show that if we marginalize this joint distribution with respect to z , then we obtain Bishop equation (9.47).
- (5) (6 points) (Bernoulli mixture model M step) Show that if we maximize the expected complete-data log likelihood function (Bishop equation (9.55)) for a mixture of Bernoulli distributions with respect to μ_k , we obtain the M step equation shown in Bishop equation (9.59).
- (6) (4 points) (Useful linear algebra tricks) If \mathbf{v} is an eigenvector of the matrix $\mathbf{A}\mathbf{A}'$ (\mathbf{A} multiplied by its transpose) with eigenvalue λ , show that $\mathbf{A}'\mathbf{v}$ is an eigenvector of $\mathbf{A}'\mathbf{A}$. Remember to obey the rules for *matrix* multiplication.

For 7 thru 9

Download spectclust.m eigsort.m and HW3.mat from the class website/ zipfile.

In matlab type

```
load HW3
```

HW3.mat contains an array of data named allpts. Each 2-D datapoint is given by a row in allpts. There are 50 datapoints/rows in the array. The data consist of two concentric rings of datapoints. Datapoints in the inner ring occupy the first 20 rows of allpts.

- (7) (6 points) Run the code in spectclust.m

Describe each sub-figure in the plot and say what it shows. (e.g. Were the data clustered with spectral clustering as you would wish? How does the k-means clustering compare? What do the rest of the subplots show?)

- (8) (4 points) Now try changing sigsq to different values. What does sigsq represent? How does this effect the different stages? What happens when sigsq is too small? What happens when sigsq is too large?

(9) (2 points) For this question, you will generate more two-ring datasets and run the code with your favorite value of `sigsq`. For this part you will have to comment and uncomment the noted lines in the code to create the data in the program (a new dataset with each run). How well does your `sigsq` work for different random datasets generated with the same code? Can you identify why some datasets don't work well? (be sure to run the code at least 50 times).

(10) **ungraded but good to do - we will provide solutions** Computing Eigenvectors

1. Enter the following matrix in to Matlab

$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 1 \\ 2 & 4 & 5 \end{bmatrix}$

2. Create a random 3-Dimensional vector x of unit length.

3. Compute $x_{new} = (A * x_{old}) / ||A * x_{old}||$

4. Repeat step 3 for 5 times with appropriate substitution of $x_{old} = x_{new}$. (Print out or write down your Matlab code as well as the values for x at each iteration)

5. Compute the eigenvalues and eigenvectors of A using Matlab's `eig` function and write them down.

6. How are your answers in 4. and 5. related?

7. What would happen if you repeated part (G) with the matrix A replaced by a 3x3 identity matrix (you do not need to do it if you know what will happen)? Why are the results so different?