

**Gaussian Mixture Models (GMMs)***(find and work with a partner)*

1. Write down the initialization for the three sets of variables we wish to find in GMMs:

- Cluster weights  $\pi_k =$
- Cluster means  $\vec{\mu}_k =$
- Cluster variances  $\sigma_k =$

2. At a certain point in the EM algorithm, say we have the following weight matrix for individual training examples at the end of the E-step:

$$W = \begin{bmatrix} 0.2 & & 0.3 \\ 0.7 & 0.1 & \\ & 0.4 & 0.5 \\ 0.8 & & 0.1 \end{bmatrix}$$

What is  $n$  (the number of training examples)? What is  $K$  (the number of clusters)?

3. Fill in the missing entries in the matrix above. Should the rows or columns sum to 1?

4. During the M-step we use  $W$  to update the parameters. First use  $W$  to update the  $\pi_k$  values, using the formulas:

$$M_k = \sum_{i=1}^n w_{ik} \quad \text{and} \quad \pi_k = \frac{M_k}{n}$$

5. Now say that our points are just in 1D and  $x_1 = 5$ ,  $x_2 = 10$ ,  $x_3 = 2$ , and  $x_4 = 8$ . Update the cluster means using the formula:

$$\mu_k = \frac{1}{M_k} \sum_{i=1}^n w_{ik} x_i$$

6. Finally, draw these four points on a number line - do the results make sense?