EE 660 # Homework 13 **Posted:** 11/19/2019

Jenkins **Due:** Tues., 11/26/2019, 2:00 PM

Reading - Unsupervised Learning (USL)

(no reading problems to turn in)

Sources

- **[Xu and Wunsch]** Rui Xu and Donald Wunsch II, "Survey of Clustering Algorithms", *IEEE Trans. Neural Networks*, Vol. 16, No. 3 (May 2005). Available from USC Library (electronic journals), or from the link posted on D2L with this assignment.
- **[Murphy]** Selected sections from Ch. 25

Reading

- **Introduction:** Xu and Wunsch Section I.
- **EM:** Xu and Wunsch Section II D (p 653), to end of p. 653.
- **Introduction, similarity measures, hierarchical graphical techniques:**

  Murphy 25.1.0, 25.1.1, 25.5.0-25.5.3.
- **Finish graphical techniques, how to choose *K* (number of clusters):**

  Xu and Wunsch Section II M (starting on page 664).
- **Optional reading** - you're not responsible for this, but if you'd like to delve further into unsupervised learning on your own, these are some suggested places to start:
  - Browse other sections of Xu and Wunsch, and other sections of Murphy Ch. 25.

Homework Problem – Semi-Supervised Learning (SSL)

1. Consider a 2-class semi-supervised learning problem in which there are l labeled samples and $u = 1$ unlabeled sample. (For example, think of being given l labeled samples, and then acquiring unlabeled samples one at a time.) There is 1 feature, and each class is modeled as a Gaussian:

$$p\left(x \middle| y = c, \underline{\theta}\right) = N\left(x \middle| \mu_c, \sigma_c{}^2\right), \quad c = 1,2$$

In the parts below, you will use EM to estimate the means μ_1 and μ_2 . You may assume the priors and variances are given constants. Generally the subscripts h and i will indicate unlabeled and labeled samples, respectively.

a) Consider the t^{th} iteration of EM. Derive the E step in terms of given quantities: that is, starting from

$$p\left(\mathcal{H}\middle|\mathcal{D},\underline{\theta}^{(t)}\right)=p\left(y_h=c_h\middle|x_h,\underline{\theta}^{(t)}\right)=\gamma_{hc_h}^{(t)}\ ,\quad c_h=1,2$$

show that:

$$\gamma_{hc_h}^{(t)}=\frac{\pi_{c_h}}{\alpha_h^{(t)}\sqrt{2\pi\sigma_{c_h}^{\ 2}}}\exp\left\{-\frac{\left(x_h-\mu_{c_h}^{(t)}\right)^2}{2\sigma_{c_h}^{\ 2}}\right\}$$

in which $\pi_{c_h}\triangleq p\left(y_h=c_h\middle|\underline{\theta}^{(t)}\right)=p\left(y_h=c_h\right)$. Also, find $\alpha_h^{(t)}$.

In parts (b)-(d), you will derive the M step formulas, also for the t^{th} iteration of EM.

b) First, show that

$$p\left(\mathcal{D},\mathcal{H}\middle|\underline{\theta}\right)=p\left(x_h\middle|y_h=c_h,\underline{\theta}\right)\pi_{c_h}\prod_{i=1}^{l}p\left(x_i\middle|y_i=c_i,\underline{\theta}\right)\pi_{c_i}$$

in which $\pi_{c_i}=p\left(y_i=c_i\middle|\underline{\theta}\right)=p\left(y_i=c_i\right)$ and similarly for π_{c_h}.

c) Take $\ln p\left(\mathcal{D},\mathcal{H}\middle|\underline{\theta}\right)$ from your result of (b), plug in for the normal densities, and drop any additive terms that are constants of $\underline{\theta}$. Then plug in to the M equation:

$$\begin{aligned}\underline{\theta}^{(t+1)}&=\underset{\underline{\theta}}{\arg\max}\ \mathsf{E}_{\mathcal{H}|\mathcal{D},\underline{\theta}^{(t)}}\left\{\ln p\left(\mathcal{D},\mathcal{H}\middle|\underline{\theta}\right)\right\}\\&=\underset{\underline{\theta}}{\arg\max}\ \sum_{c_h=1}^{2}\gamma_{hc_h}^{(t)}\ \ln p\left(\mathcal{D},\mathcal{H}\middle|\underline{\theta}\right)\end{aligned}$$

and simplify to get:

$$\underline{\theta}^{(t+1)}=\underset{\underline{\theta}}{\arg\max}\ \left\{\sum_{c_h=1}^{2}\gamma_{hc_h}^{(t)}\left[-\frac{\left(x_h-\mu_{c_h}\right)^2}{\sigma_{c_h}^{\ 2}}\right]+\sum_{i=1}^{l}\left[-\frac{\left(x_i-\mu_{c_i}\right)^2}{\sigma_{c_i}^{\ 2}}\right]\right\}$$

in which a constant multiplicative factor of $\frac{1}{2\pi}$ has been dropped. (**Hint:** you may find it useful to use $\gamma_{h1}^{(t)} + \gamma_{h2}^{(t)} = 1$.)

d) Re-write your result of part (c) to express it in terms of $\mu_1, \mu_2\ \sigma_1{}^2, \sigma_2{}^2$. (**Hint**: you might find it useful to use the indicator function.) Then solve for $\underline{\theta}^{(t+1)} = \left[\mu_1{}^{(t+1)}, \mu_2{}^{(t+1)}\right]^T$. (**Hint:** find the argmax by taking $\frac{\partial}{\partial \mu_1}$ and setting equal to 0; similarly for μ_2.) Let $l_1 =$ number of labeled samples with label $c_i = 1$, and $l_2 =$umber of labeled samples with label $c_i = 2$. (Note that $\gamma_{hc_h}^{(t)}$ is constant of μ_1 and μ_2 because it used the (constant) estimates $\mu_1{}^{(t)}$ and $\mu_2{}^{(t)}$ from the E step.)

e) Given: $\pi_1 = \pi_2 = 0.5$, $\sigma_1{}^2 = \sigma_2{}^2 = 1$; data as follows:

$$\text{labeled data } \left\{(x_i, y_i)\right\}_{i=1}^{l} = \left\{(1,1),(2,1),(4,2)\right\}; \quad \text{unlabeled sample } x_h = 3.$$

Suppose the values for $\underline{\theta}$ at the beginning of the t^{th} iteration of EM are: $\mu_1{}^{(t)} = 1.5$, $\mu_2{}^{(t)} = 4.0$. Note that you may solve this part (e) by hand, or use a computer to assist you (your choice).

(i) Calculate the responsibilities $\gamma_{h1}^{(t)}$ and $\gamma_{h2}^{(t)}$ from the E step (using part (a));

(ii) Calculate the new mean estimates $\mu_1{}^{(t+1)}$ and $\mu_2{}^{(t+1)}$ from the M step (using part (d) result).