Activity 15 – Expectation Maximization

Introduction

Bayes rule states that

$$p(A|B)=\frac{p(B|A)P(A)}{P(B)} \tag{1}$$

Where p(A|B) is the probability distribution function (pdf) of A occuring given B, p(B|A) is the pdf of B occuring given A, P(A) is the prior probability of A, and P(B) is the prior probability of B.

The Expectation Maximization algorithm allows the estimation of the distribution parameters of an ensemble. The pdf of an ensemble **x** given a set of parameters Θ may be expressed as sum of weighted pdfs:

$$p(\boldsymbol{x}|\Theta)=\sum_{l=1}^{M} P_l\, p_l(\boldsymbol{x}|\boldsymbol{\theta_l}) \tag{2}$$

where P_l is the prior probability of the l^{th} pdf, p_l , is the probability of observing **x** given the parameter θ_l and M is the number of component pdf's. It is useful to define the probability of an observation belonging to the l^{th} distribution given the observation itself and the parameters. This is defined as

$$P(l|x_i,\boldsymbol{\Theta})=\frac{P_l\, p_l(x_i|\theta_l)}{p(x_i|\boldsymbol{\Theta})}=\frac{P_l\, p_l(x_i|\theta_l)}{\sum_{l=1}^{M} P_l\, p_l(x_i|\theta_l)} \tag{3}$$

If we assume a d-dimensional Gaussian distribution , the component pdf is

$$p_l(\boldsymbol{x}|\boldsymbol{\mu_l},\boldsymbol{\Sigma_l})=\frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma_l}|^{1/2}}\exp\{-1/2(\boldsymbol{x}-\boldsymbol{\mu_l})^T\boldsymbol{\Sigma_l^{-1}}(\boldsymbol{x}-\boldsymbol{\mu_l})\} \tag{4}$$

where $\boldsymbol{\mu_l}$ is the l^{th} mean and $\boldsymbol{\Sigma_l}$ is the l^{th} covariance matrix. Using EM, the pdf parameters of a distribution can be estimated using the following update rules. N is the number of datapoints. The solution is guaranteed to converge.

UPDATE RULES:

$$P_l^{new} = \frac{1}{N} \sum_{i=1}^{N} P(l|\mathbf{x}_i, \mathbf{\Theta}^g) \qquad \boldsymbol{\mu}_l^{new} = \frac{\sum_{i=1}^{N} \mathbf{x}_i P(l|\mathbf{x}_i, \mathbf{\Theta}^g)}{\sum_{i=1}^{N} P(l|\mathbf{x}_i, \mathbf{\Theta}^g)}$$

$$\mathbf{\Sigma}_l^{new} = \frac{\sum_{i=1}^{N} P(l|\mathbf{x}_i, \mathbf{\Theta}^g)(\mathbf{x}_i - \boldsymbol{\mu}_l^{new})(\mathbf{x}_i - \boldsymbol{\mu}_l^{new})^T}{\sum_{i=1}^{N} P(l|\mathbf{x}_i, \mathbf{\Theta}^g)} \qquad (4)$$

where $\mathbf{\Theta}^g$ is the old guessed parameters $(\boldsymbol{\mu}_l, \mathbf{\Sigma}_l)$.

Procedure

1. Using your fruits data implement the EM algorithm. Let all prior probabilities be initially equal (1/M) and let all covariance matrices be equal. Let the means be arbitrary points as well so long as they are within the distribution. M is the number of fruit classes in your ensemble but you can test what happens if you set it to greater than M or less than M so long as the minimum is 2.
2. Alternatively, instead of using the E-step where you calculate the log likelihood and stop if it is maximized, you can define a stopping parameter to be equal to the difference between the new and the old guessed parameters. Execute the program such that it will halt when the difference threshold is reached.
3. Overlay a plot of the estimated pdf on your fruit feature space.