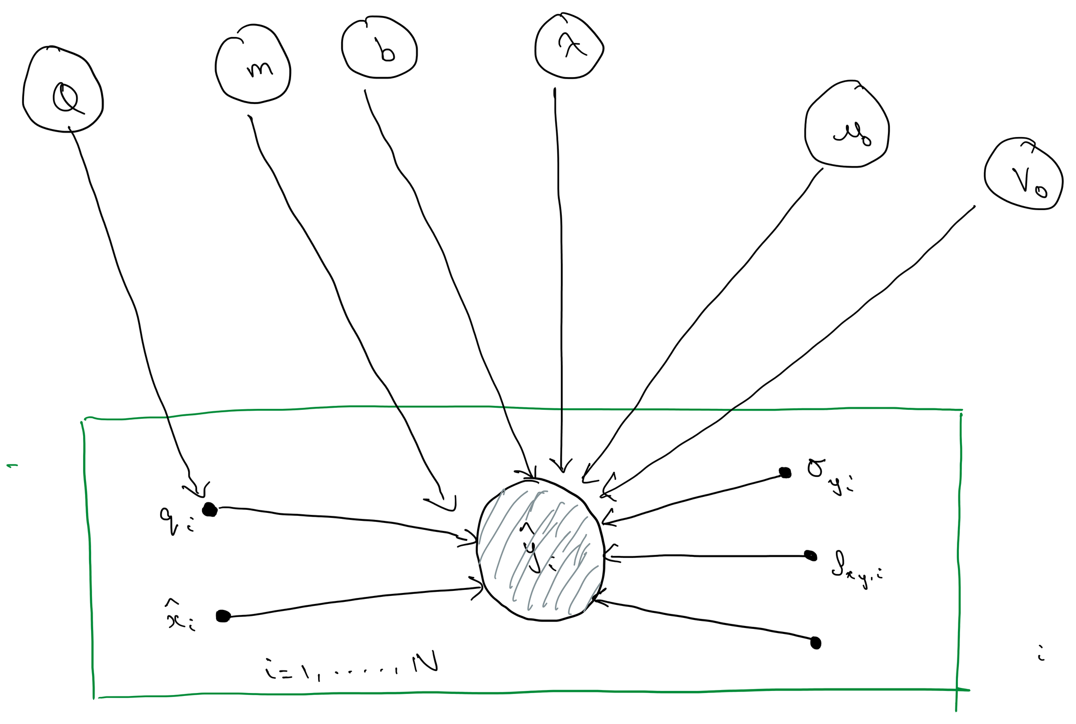
Data Analysis and Machine Learning Problem Set 3

Tuesday, 15 March 2022 5:15 pm

Question 1:



Question 2:

a. The model is a Gaussian mixture model comprising of K=5 one-dimensional Gaussians of differing means and standard deviations. The model parameters, denoted by θ, are such that $\theta = \{\mu, \sigma, \pi\}$ where μ and σ represent the mean and standard deviations of each of the components and π represents the mixing proportion (i.e. the extent to which a given component contributes to the model (the sum of all of the mixing proportions must be equal to 1). Each of the different model parameters has the following priors associated with it:

- $\pi_k \sim$ Multinomial (K) (due to the additional constraint that $\sum_k \pi_k = 1$ along with the fact that $\pi_k \sim U(0,1)$)
- $\sigma \sim U(0, \infty)$ (standard deviations must be non-negative but are otherwise uniformly distributed $\Rightarrow$ improper uniform prior.
- $\mu \sim U(-\infty, \infty)$ (improper uniform prior on the mean)

The log-likelihood function is defined as:

$$L = \prod_{i=1}^{N} \sum_{k=1}^{K} \pi_k \cdot \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(\frac{-(x_i-\mu_k)^2}{2\sigma_k^2}\right) \quad \text{(likelihood function).}$$

$$\therefore \log(\mathcal{L}) = \log\left[\prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k \cdot \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(\frac{-(x_i-\mu_k)^2}{2\sigma_k^2}\right)\right]$$

$$= \sum_{i=1}^{n} \log\left(\sum_{k=1}^{K} \pi_k \cdot \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(\frac{-(x_i-\mu_k)^2}{2\sigma_k^2}\right)\right)$$

b. In order to determine the component that a particular data point was drawn from, one must compute the posterior probability of membership. The probability of the i^{th} data point belonging to the k-th component is given by:

$$p(M_k|x_i) = \frac{p(x_i|M_k)\,p(M_k)}{\sum_{j=1}^{K} p(x_i|M_j)\,p(M_j)} \quad \left(\text{assuming that } \underbrace{p(M_k) = p(M_j)}_{\text{priors}} \text{ for each component}\right)$$

Since the data in the problem is normally distributed, it is evident that $p(x_i|M_k)$ and $p(x_i|M_j)$ have the following form:

$$p(x_i|M_k) = \frac{\pi_k}{\sqrt{2\pi\sigma_k^2}} \exp\left(\frac{-(x_i-\mu_k)^2}{2\sigma_k^2}\right)$$

(where μ_k and σ_k are the respective means and standard deviations of the k-th component)

Hence:

$$p(M_k|x_i) = \frac{\overbrace{\frac{\pi_k}{\sqrt{2\pi\sigma_k^2}} \exp\left(\frac{-(x_i-\mu_k)^2}{2\sigma_k^2}\right)}^{\text{the } k^{th} \text{ component of the mixture model } (\pi_k = \text{mixing proportion})}}{\underbrace{\sum_{j=1}^{K} \frac{\pi_j}{\sqrt{2\pi\sigma_j^2}} \exp\left(\frac{-(x_i-\mu_j)^2}{2\sigma_j^2}\right)}_{\text{sum over the K=5 different components}}} = \frac{\pi_k \sigma_k^{-2} \exp\left(\frac{-(x_i-\mu_k)^2}{2\sigma_k^2}\right)}{\sum_{j=1}^{K} \pi_j \sigma_j^{-2} \exp\left(\frac{-(x_i-\mu_j)^2}{2\sigma_j^2}\right)}$$

The above is the expectation step. This produces a 1000 x 5 (N x K) matrix of membership probabilities which can then be utilised in the maximisation step.

In the maximisation step, one can update the model parameters given the N x K matrix of membership probabilities.

The mixing proportions of each component can be updated noting that:

$$N_k = \sum^{K} \omega_{nk} \quad \Big\} \text{ provides the effective no. of data points specified by each mixture component. } (1 \times K \text{ matrix})$$

Hence the new mixing proportions can be updated by dividing each element of the above matrix N_k by the number of data points:

$$\omega_k^{(new)} = \frac{N_k}{N}$$

Similarly, the means can be updated as follows:

$$\mu_k^{(new)} = \frac{1}{N_k} \sum_{n=1}^{N} \omega_{nk}\, \underset{\sim}{x} \qquad \left(\underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}\right)$$

The standard deviations can also be updated as follows (since the data points x_i are one-dimensional)

$$\sigma_k^{(new)} = \sqrt{\frac{1}{N_k} \sum_{n=1}^{N} \omega_{nk} (x_n - \mu_k^{(new)})^2}$$

Question 3:

b. As the data are generated by drawing random samples and then adding one of 3 trends to a select few thereof (corresponding to +1°/century and -1°C/century and unaffected) one can conclude that a suitable generative model for this data could be a Gaussian mixture model consisting of 3 components (K=3). As such the One can assume a uniform prior on the means and standard deviations of the mixture components. Since the gradients of the three lines that could have generated the dataset have values of -0.01, 0 and 0.01 respectively, the means of the 3 Gaussian components can be assumed to be equal to these values. All of the priors on the parameters can be assumed to be the same as in Question 2. The log likelihood function will also assume the same form as that in Question 2.