Homework 03 B - 50 points

October 2019

Provide a PDF file entitled hw03.pdf that answers all the questions below and shows the figures generated by your code. Submit the PDF via Canvas and include it in your Github repository.

1 PCA

Given a set of data $X = \begin{bmatrix} 2 & 3 & 3 & 4 & 5 & 7 \\ 2 & 4 & 5 & 5 & 6 & 8 \end{bmatrix}$

1. Calculate the eigenvalues of $X^T X$ **by code or calculators**. **(Do not subtract mean or normalize data)**
2. Calculate the eigenvalues and eigenvectors of XX^T. Solve this by hand AND using code you implement to verify your solution. **(Do not subtract mean or normalize data)** (Note that all eigenvectors need to be normalized.)
3. Are the non-zero eigenvalues of $X^T X$ and those of XX^T the same? Why or why not?
4. For dataset X, there are 6 samples with 2 dimensions. Reduce X dimensions to 1. Print/write the reduced 1-dimension data. It should be a 1×6 vector. Solve this by hand AND using code you implement to verify your solution. **(Please subtract mean and normalize data.)**

2 EM

2.1 Background

Suppose we have 3 coins named A, B and C. The probabilities of each coin to be head are π, p and q respectively. Then, we do experiments as follows: We flip coin A first and decide what to do next based on the result. If the result is heads, we flip coin B; else, we flip coin C. If the result of the B/C coin flip is heads, we record 1; else, we record 0. We do this experiment 10 times and get the following outcomes:

$$1, 1, 0, 1, 0, 0, 1, 0, 1, 1$$

Assume that we can only know the results of the experiments rather than the process (we do not know which coins were flipped). Use EM algorithm to estimate the parameters of the three coins A, B and C, namely π, p and q.

2.2 Introduction

The model can be written as:

$$\begin{aligned} P(y|\theta) &= \sum\nolimits_z P(y,z|\theta) = \sum\nolimits_z P(z|\theta)P(y|z,\theta) \\ &= \pi p^y (1-p)^{(1-y)} + (1-\pi) q^y (1-q)^{1-y} \end{aligned} \tag{1}$$

where y is observable variable, representing that the result of flipping coins is 1 or 0; z is latent variable which represents the results of flipping coin A. $\theta = (\pi, p, q)$ are the parameters.

The observable data are written as $Y = (Y_1, Y_2, \ldots, Y_n), n = 1, 2, \ldots, 10$ and the latent data are written as $Z = (Z_1, Z_2, \ldots, Z_n), n = 1, 2, \ldots, 10$. Thus, the likelihood of the observable data is

$$P(Y|\theta) = \prod_{j=1}^{10} \left[\pi p^{y_j} (1-p)^{(1-y_j)} + (1-\pi) q^{y_j} (1-q)^{1-y_j} \right] \tag{2}$$

To estimate the parameters $\theta = (\pi, p, q)$, we can use MLE with Eq.2 and get Eq.3:

$$\hat{\theta} = arg\max_{\theta} log P(Y|\theta) \tag{3}$$

2.3 Questions

- Question 2.1 Can Eq.3 be solved directly?

2.3.1 E step

To solve Eq.3 by EM algorithms, we need to initialize the parameters $\theta^{(0)} = (\pi^{(0)}, p^{(0)}, q^{(0)})$. By updating $\theta^{(i)}$, the parameters will converge.

The E step is to calculate the probability that y_j is from flipping coin B given the parameters at the i^{th} iteration $(\pi^{(i)}, p^{(i)}, q^{(i)})$, where i means the i^{th} iteration and j means the j^{th} sample. The equation is given here as Eq.4.

$$\mu_j^{(i+1)} = \frac{\pi^{(i)} (p^{(i)})^{y_j} (1-p^{(i)})^{1-y_j}}{\pi^{(i)} (p^{(i)})^{y_j} (1-p^{(i)})^{1-y_j} + (1-\pi^{(i)}) (q^{(i)})^{y_j} (1-q^{(i)})^{1-y_j}} \tag{4}$$

2.3.2 M step

Then update the new estimated parameters by Eq.5, Eq.6 and Eq.7.

$$\pi^{(i+1)} = \frac{1}{10}\sum_{j=1}^{10}\mu_j^{(i+1)} \tag{5}$$

$$p^{(i+1)} = \frac{\sum_{j=1}^{10}\mu_j^{(i+1)}y_j}{\sum_{j=1}^{10}\mu_j^{(i+1)}} \tag{6}$$

$$q^{(i+1)} = \frac{\sum_{j=1}^{10}(1-\mu_j^{(i+1)})y_j}{\sum_{j=1}^{10}(1-\mu_j^{(i+1)})} \tag{7}$$

- Question 2.2 Calculate the $\mu^{(1)}$ given initialization value $\theta^{(0)} = (\pi^{(0)}, p^{(0)}, q^{(0)}) = (0.5, 0.5, 0.5)$
- Question 2.3 Calculate the $\theta^{(1)}$ given initialization value $\theta^{(0)} = (\pi^{(0)}, p^{(0)}, q^{(0)}) = (0.5, 0.5, 0.5)$ and $\mu^{(1)}$ calculated in Question 2.2.
- Question 2.4 Calculate the $\mu^{(2)}$ given initialization value $\theta^{(0)} = (\pi^{(0)}, p^{(0)}, q^{(0)}) = (0.5, 0.5, 0.5)$ and other parameters calculated in Question 2.2 and 2.3.
- Question 2.5 Calculate the $\theta^{(2)}$ given initialization value $\theta^{(0)} = (\pi^{(0)}, p^{(0)}, q^{(0)}) = (0.5, 0.5, 0.5)$ and other parameters calculated in Question 2.2, 2.3 and 2.4.
- Question 2.6 Calculate the $\theta^{(2)}$ given initialization value $\theta^{(0)} = (\pi^{(0)}, p^{(0)}, q^{(0)}) = (0.4, 0.6, 0.7)$. Is $\theta^{(2)}$ the same as that in Question 2.5?
- Question 2.7 Write code to implement the EM algorithm based on Eq.4-7. Iterate 20 times and print $\mu^{(i)}$ and $\theta^{(i)} = (\pi^{(i)}, p^{(i)}, q^{(i)})$ at i^{th} iteration given two different initialization $\theta^{(0)} = (\pi^{(0)}, p^{(0)}, q^{(0)}) = (0.4, 0.6, 0.7)$ and $\theta^{(0)} = (\pi^{(0)}, p^{(0)}, q^{(0)}) = (0.5, 0.5, 0.5)$. Are the results of parameters estimation the same with different initialization?

3 Supplementary Materials

We provide a brief derivation and explanation for EM questions here. From Eq. 2 and 3, we can get Eq. 8

$$\begin{aligned}\hat{\theta} &= \underset{\theta}{argmax}\, logP(Y|\theta) \\ &= \underset{\theta}{argmax}\sum_{j=1}^{10} log\left[\pi p^{y_j}(1-p)^{(1-y_j)} + (1-\pi)q^{y_j}(1-q)^{1-y_j}\right]\end{aligned} \tag{8}$$

For E step:

We define $\mu_j^{(i+1)}$ as the probability that the final result (whether 0 or 1 and that is known) is from coin B (whether from coin B or not is unknown and we estimate the probability with $\mu_j^{(i+1)}$). Thus, for each experiment result (10 in total), there should be a $\mu_j^{(i+1)}$ where j is from 1 to 10.

For M step:

Define Q function as follows:

$$Q(\theta, \theta^i) = \sum_{j=1}^{10} \left(\mu_j^{(i+1)} log \left[\pi p^{y_j} (1-p)^{(1-y_j)} \right] + (1-\mu_j^{(i+1)}) log \left[(1-\pi) q^{y_j} (1-q)^{1-y_j} \right] \right)$$

Take the derivative of Q related to π, p and q.

$$\begin{aligned}
\frac{\partial Q}{\partial \pi} &= \frac{\partial \left(\sum_{j=1}^{10} \left(\mu_j^{(i+1)} log[\pi p^{y_j} (1-p)^{(1-y_j)}] + (1-\mu_j^{(i+1)}) log[(1-\pi) q^{y_j} (1-q)^{1-y_j} \right) \right)}{\partial \pi} \\
&= \frac{\sum_{j=1}^{10} \mu_j^{(i+1)}}{\pi} + \frac{\sum_{j=1}^{10} (1-\mu_j^{(i+1)})}{\pi - 1} \\
&= \frac{10\pi - \sum_{j=1}^{10} \mu_j^{(i+1)}}{\pi(\pi-1)}
\end{aligned} \tag{9}$$

Let $\frac{\partial Q}{\partial \pi} = 0$ and we can get Eq.5.

$$\begin{aligned}
\frac{\partial Q}{\partial p} &= \frac{\partial \left(\sum_{j=1}^{10} \left(\mu_j^{(i+1)} log[\pi p^{y_j} (1-p)^{(1-y_j)}] + (1-\mu_j^{(i+1)}) log[(1-\pi) q^{y_j} (1-q)^{1-y_j} \right) \right)}{\partial p} \\
&= \sum_{j=1}^{10} \mu_j^{(i+1)} \left(\frac{y_j}{p} - \frac{1-y_j}{1-p}\right) \\
&= \sum_{j=1}^{10} \mu_j^{(i+1)} \left(\frac{y_j - p}{p(1-p)} \right)
\end{aligned} \tag{10}$$

Let $\frac{\partial Q}{\partial p} = 0$ and we can get Eq.6.

$$
\begin{aligned}
\frac{\partial Q}{\partial q} &= \frac{\partial\left(\sum_{j=1}^{10}\left(\mu_j^{(i+1)}log[\pi p^{y_j}(1-p)^{(1-y_j)}]+(1-\mu_j^{(i+1)})log[(1-\pi)q^{y_j}(1-q)^{1-y_j}\right)\right)}{\partial q} \\
&= \sum_{j=1}^{10}(1-\mu_j^{(i+1)})(\frac{y_j}{q}-\frac{1-y_j}{1-q}) \\
&= \sum_{j=1}^{10}(1-\mu_j^{(i+1)})(\frac{y_j-q}{q(1-q)})
\end{aligned} \tag{11}
$$

Let $\frac{\partial Q}{\partial q}=0$ and we can get Eq.7.