HW4 — ISYE6416

Ashish Dhiman — ashish.dhiman9@gatech.edu

May 7, 2023

Contents

1 Question 1: EM for Paper Review

The answers to this question is given in hand written notes below:

Q-1.

We have the following setup:

$x^{pr} \rightarrow$ rating by reviewer 'r' of paper 'p'

$\mu_p \rightarrow$ True Quality of paper p

$\gamma_r \rightarrow$ Bias (+ve) of reviewer 'r'

$$y^p \sim \mathcal{N}(\mu_p, \sigma_p^2)$$

$$z^r \sim \mathcal{N}(\gamma_r, \tau_r^2)$$

$$x^{pr} | y^p, z^r \sim \mathcal{N}(y^p + z^r, \sigma^2)$$

$\sigma^2 \rightarrow$ known

Total 'P' papers

Total 'R' reviewer

unknown params, $\theta = \{\mu_p, \sigma_p^2, \gamma_r, \tau_r^2\}$

$\max Lik(x^{pr} | \theta) \; \forall \; p, r$

part a

(i) Joint Density $p(x^{pr}, y^p, z^r) = ?$

Now we know for any 2 ~~RV~~ Normal RV's X, Y

$$X \sim \mathcal{N}(\mu_x, C_x) \;\&\; Y \sim \mathcal{N}(\mu_y, C_y)$$

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} C_x & C_{xy} \\ C_{yx} & C_y \end{bmatrix}\right)$$

$\hookrightarrow$ Multivariate

Hence have

$p(\cancel{x^{pn}}, y^{b}, z^{n})$

$$\cancel{\begin{bmatrix} x^{pn} \\ y^{b} \\ z^{n} \end{bmatrix} \sim N\left(\left[\right.\right.}$$

$p(y^{b}, z^{n}, x^{bn})$

$$\begin{bmatrix} y^{b} \\ z^{n} \\ x^{bn} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_{y^b} \\ \mu_{z^n} \\ \mu_{x^{bn}} \end{bmatrix}, \begin{bmatrix} C_{y^b} & C_{yz} & C_{yx} \\ C_{zy} & C_{z^n} & C_{zx} \\ C_{xy} & C_{xz} & C_{x^{bn}} \end{bmatrix} \right)$$

Now we know y^{b}, z^{n} are independent

$\Rightarrow C_{yz} = C_{zy} = 0$

Also $x^{bn} \sim N(y^{b} + z^{n}, \sigma^{2})$

$\rightarrow x^{bn} = y^{b} + z^{n} + \epsilon \quad | \; \epsilon \sim N(0, \sigma^{2}) \quad \Rightarrow x^{bn} \sim N(\mu^{b} + \tau^{n}, \sigma_{pz}^{2} + \tau_{n} + \sigma)$

$C_{xy} = \cancel{Cov}\; C_{yx} = Cov(y^{b}, x^{bn})$

$= Cov(y^{b}, y^{b} + z^{n} + \epsilon)$

$= \sigma_{b}^{2} + 0 + 0$

$C_{zx} = C_{xz} = Cov(z^{n}, x^{bn})$

$= Cov(z^{n}, y^{b} + z^{n} + \epsilon)$

$= \cancel{0}\, \tau_{n}^{2} + 0 + 0$

Hence we have

$$p(y^p, z^n, x^{pn}) \sim \mathcal{N}\left(\begin{bmatrix} \mu_p \\ \gamma_n \\ \mu_p + \gamma_n \end{bmatrix}, \begin{bmatrix} \sigma_p^2 & 0 & \sigma_p^2 \\ 0 & \tau_n^2 & \tau_n^2 \\ \sigma_p^2 & \tau_n^2 & \sigma^2 + \sigma_p^2 + \tau_n^2 \end{bmatrix}\right)$$

Hence ~~the~~ given above are the mean & co-variance of the multivariate gaussian density

(ii) $Q_{pn}(\theta' | \theta) = E\left[\log p(y^p, z^n, x^{pn} | x^{pn}, \theta)\right]$

, where

$\theta' \rightarrow$ updated params

$\theta \rightarrow$ params from prev iteration

Now we know

$p(y^p, z^{n}, ~~x^{pn}~~ | x^{pn}, \theta)$ can be found using conditional multivariate normal i.e.

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}\right)$$

$$x_1 | x_2 \sim \mathcal{N}(\mu_{1|2}, \Sigma_{1|2})$$

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Hence in our case

$p(y^p, z^n | x^{pn}) =$

$$\sim N\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{bmatrix}\right) \quad [\text{Let}]$$

Eqn 2.1

$$\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} = \begin{bmatrix}\mu_p\\ \gamma_n\end{bmatrix} + \begin{bmatrix}\sigma_p^2\\ \tau_{n}^2\end{bmatrix} \times \frac{1}{\sigma^2 + \sigma_p^2 + \tau_n^2} \times \left(x^{pn} - \mu^p - \gamma^n\right)$$

Eqn 2.2 Similarly

$$\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{bmatrix} = \begin{bmatrix}\sigma_p^2 & 0\\ 0 & \tau_n^2\end{bmatrix} - \underset{2\times n}{\begin{bmatrix}\sigma_p^2\\ \tau_n^2\end{bmatrix}} \times \left(\sigma_p^2 + \tau_n^2 + \sigma^2\right)^{-1} \times \underset{1\times 2}{\begin{bmatrix}\sigma_p^2 & \tau_n^2\end{bmatrix}}$$

$$= \begin{bmatrix}\sigma_p^2 & 0\\ 0 & \tau_n^2\end{bmatrix} - \frac{1}{\sigma_p^2 + \tau_n^2 + \sigma^2} \times \begin{bmatrix}\sigma_p^4 & \sigma_p^2\tau_n^2\\ \sigma_p^2\tau_n^2 & \tau_n^4\end{bmatrix}$$

Hence

$$Q(\theta'|\theta) = E\left[p(y^p, z^n, x^{pn})\right]$$

Now using Baye's

$$p(y^p, z^n, x^{pn}) = p\left(x^{pn} \mid y^p, z^n\right) \times \cancel{p(y^p | z^n)}^{\,p(y^p)} \times p(z^n)$$

$$= \frac{1}{(\sqrt{2\pi})^3} \times \frac{1}{\sqrt{\sigma^2 \times \sigma_p^2 \times \tau_n^2}} \times \exp\left[-\left(\frac{(x^{pn} - y^p - z^n)^2}{2\sigma^2}\right) - \frac{(y^{pn} - \mu^p)^2}{2\sigma_p^2} - \frac{(z^n - \gamma_n)^2}{2\tau_n^2}\right]$$

Now ~~we~~ need to find Expectation of above using conditional $p(y^p, z^n | x^{pn})$

Now by linearity of expectation

$$\log P(x^{pn}, y^p, z^n) = -\frac{3}{2}\log 2\pi - \frac{1}{2}\log(\sigma^2 \times \sigma_p^2 \times \tau_n^2)$$

$$-\frac{1}{2}\left(\underbrace{\frac{(x^{pn} - y^p - z^n)^2}{\sigma^2}}_{A} + \underbrace{\frac{(y^p - \mu_p^*)^2}{\sigma_p^2}}_{B} + \underbrace{\frac{(z^{pn} - \gamma_n)^2}{\tau_n^2}}_{C}\right)$$

$$\Rightarrow E(\log p(x^{pn}, y^p, z^n) \mid x^{pn}, \theta)$$

$$= -\frac{3}{2}\log 2\pi - \frac{1}{2}\log(\sigma^2 \times \sigma_p^2 \times \tau_n^2) - \frac{1}{2}E(A + B + C)$$

~~E(A)~~ For $E(A)$

$$x^{pn} = y^p + z^n + \epsilon$$

$$x^{pn} - y^p - z^n = \epsilon \sim N(0, \sigma^2)$$

$$\frac{\epsilon}{\sigma^2} \sim N(0, 1)$$

$$E((\epsilon - 0)^2) = Var(\epsilon)$$

$$\Rightarrow E(A) = E\left(\frac{(\epsilon - 0)^2}{\sigma^2}\right) = \frac{1}{\sigma^2} \times \sigma^2 = 1$$

For $E(B)$ Use $E(x^2) = Var(x) + (E(x))^2$

$$\Rightarrow E(B) = \frac{1}{(\sigma_p)^2} \times \left(E_p(y^p - \mu_p)^2_{|x^{pn}} + Var(y^p - \mu_p^*)|x^{pn}\right)$$

$$= \frac{1}{\sigma_p^2} \times \left(\sigma_{11} + (\mu_1 - \mu_p)^2\right)$$

Similarly by symmetry

$$E(C) = \frac{1}{\tau_n^2} \times \left(\sigma_{22} + (\mu_2 - \gamma_n)^2\right)$$

$\mu_1, \mu_2, \sigma_{11}, \sigma_{22}$ are defined in eqn 1

part b:

M-step

Now we will maximise the $Q(\theta' | \theta^{*})$

$$\theta = \arg\max_{\theta'} Q(\theta' | \theta^{*})$$

$$= -\arg\min_{\theta'} \sum^{P} \sum^{R} \Big[-\tfrac{1}{2} \log \left((\sigma_p')^2 \cdot (\tau_r')^2 \cdot \cancel{\sigma^2} \right) + \frac{1}{(\sigma_p')^2} \left(\sigma_{11} + (\mu_1 - \mu_p)^2 \right) + \frac{1}{(\tau_r')^2} \left(\sigma_{22} + (\mu_2 - \gamma_r)^2 \right) \Big]$$

drop

$$= \arg\min_{\theta'} L$$

We can use partial derivative to find optimal param

$$\frac{\partial L}{\partial \mu_p'} = -2 \sum^{R} (\mu_1 - \mu_p') = 0$$

$$\Rightarrow \mu_p^{*} = \frac{1}{R} \sum^{R} \mu_1 \quad (p = 1 \ldots \text{to } P)$$

$$\frac{\partial L}{\partial \gamma_r'} = -2 \sum^{P} (\mu_2 - \gamma_r') = 0$$

$$\Rightarrow \gamma_r^{*} = \frac{1}{P} \sum^{P} \mu_2 \quad (r = 1, 2 \ldots \text{to } R)$$

$$\frac{\partial L}{\partial (\sigma_p'^2)} = \sum^{R} \frac{1}{\sigma_p^2} - \frac{1}{(\sigma_p)^4} \left(\sigma_{11} + (\mu_1 - \mu_p)^2 \right) = 0$$

$$\Rightarrow \sigma_p^{2*} = \frac{1}{R} \sum^{R} \left(\sigma_{11} + (\mu_1 - \mu_p)^2 \right)$$

By Symmetery

$$\Rightarrow \tau_r^{2*} = \frac{1}{P} \sum^{P} \left(\sigma_{22} + (\mu_2 - \gamma_r)^2 \right)$$

2 Question 2: HMM step

We want to prove:

$$\mathbb{P}(S_t = i, S_{t+1} = j | o_1, \dots, o_T) \propto \alpha_i(t)\beta_j(t+1)a_{i,j}b_{j,o_{t+1}}$$

Hence

$$\begin{aligned}
&P(S_t = i, S_{t+1} = j | o_1, \dots, o_T) \\
&= \frac{P(S_t = i, S_{t+1} = j, o_1, \dots, o_T)}{P(o_1, \dots, o_T)} \\
&\propto P(S_t = i, S_{t+1} = j, o_1, \dots, o_T) \\
&= P(o_1, \dots, o_t, S_t = i) * P(S_{t+1} = j, O_{t+1}, \dots, o_T | o_1, \dots, o_t, S_t = i) \\
&= \alpha_i(t) * P(O_{t+1}, \dots, o_T | o_1, \dots, o_t, S_t = i, S_{t+1} j) * P(S_{t+1} = j | o_1, \dots, o_t, S_t = i) \\
&= \alpha_i(t) * P(O_{t+1}, \dots, o_T | o_1, \dots, o_t, S_t = i, S_{t+1} = j) * a_{ij} \\
&= \alpha_i(t) * P(O_{t+1}, O_{t+2}, \dots, o_T | S_{t+1} = j) * a_{ij} \\
&= \alpha_i(t) * P(O_{t+2}, \dots, o_T | S_{t+1} = j) * P(O_{t+1} | S_{t+1} = j) * a_{ij} \\
&= \alpha_i(t) * \beta_j(t+1) * P(O_{t+1} | S_{t+1} = j) * a_{ij} \\
&= \alpha_i(t) * \beta_j(t+1) * b_{j,O_{t+1}} * a_{ij}
\end{aligned} \tag{1}$$

Hence Proved

3 Question 3: PCA

3.1 part a: PCA decomposition of countries

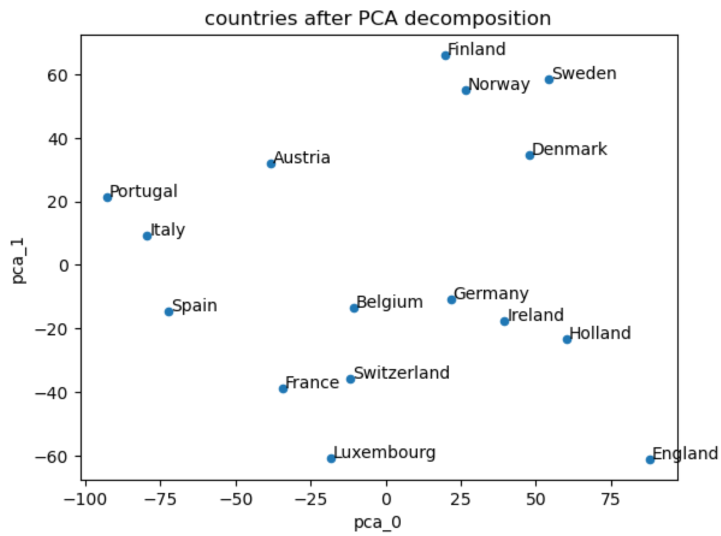

Figure 1: PCA of counties

We can see that there is some geographic pattern visible and Nordic countries cluster together. Similarly West Europe countries cluster together too.

3.2 part b: PCA decomposition of food

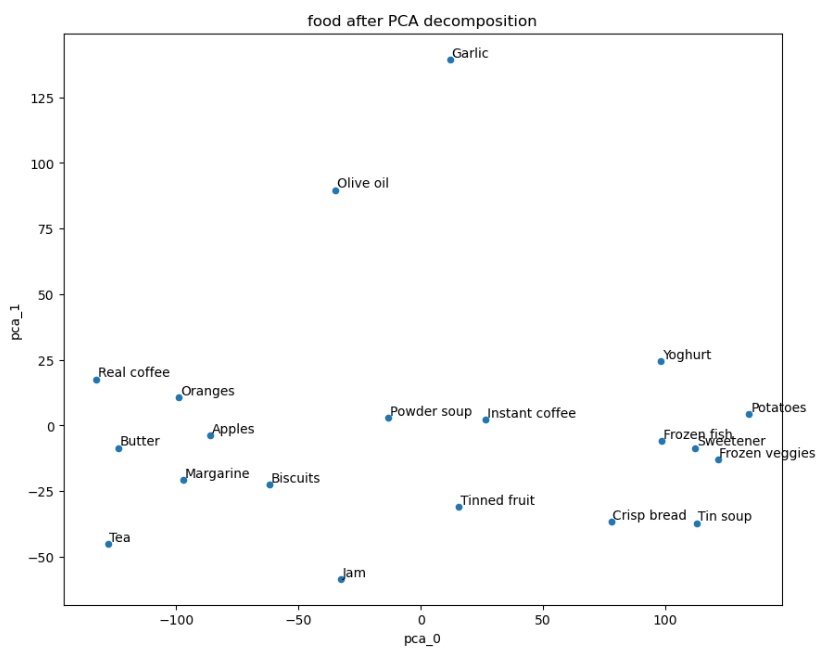

Figure 2: PCA of food

Here we see that ready to eat foods cluster on the right bottom, while ingredients are more towards the left.

4 Question 4: PCA MSE proof

We have to prove that below optimisation gives pricipal component

$$\arg\min_{\|v\|} \sum_{i=1}^{n} \|x_i - f_v(x_i)\|^2$$

Now $f_v(x_i)$ is projection of x_i onto v therefore:

$$\begin{aligned}
f_v(x_i) &= < \vec{x_i}, \vec{v} > \vec{v} \\
\text{Hence}: \|x_i - f_v(x_i)\|^2 &= \|x_i - \vec{x_i}, \vec{v} > \vec{v}|^2 \\
&= \|x_i\|^2 + < \vec{x_i}, \vec{v} >^2 \|\vec{v}\|^2 - 2(< \vec{x_i}, \vec{v} >^2) \\
&= \|x_i\|^2 - (< \vec{x_i}, \vec{v} >^2) \quad \text{(v is unitary)} \\
\implies \sum_{i=1}^{n} \|x_i - f_v(x_i)\|^2 &= \sum_{i=1}^{n} (\|x_i\|^2 - (< \vec{x_i}, \vec{v} >^2)) \\
\text{Now } x_i \text{ does'nt impact optimisation} \\
\arg\min_{\|v\|} \sum_{i=1}^{n} \|x_i - f_v(x_i)\|^2 \\
= \arg\min_{\|v\|} \sum_{i=1}^{n} -(< \vec{x_i}, \vec{v} >^2) \\
= \arg\max_{\|v\|} \sum_{i=1}^{n} (< \vec{x_i}, \vec{v} >^2) \\
\arg\max_{\|v\|} \frac{1}{n} \sum_{i=1}^{n} (< \vec{x_i}, \vec{v} >^2) \\
\text{Also } E(x^2) = (E(x))^2 - Var(x) \\
\implies \arg\max_{\|v\|} \sum_{i=1}^{n} (< \vec{x_i}, \vec{v} >^2) \\
= \arg\max_{\|v\|} (E(< \vec{x_i}, \vec{v} >)^2 + var(< \vec{x_i}, \vec{v} >) \\
= \arg\max_{\|v\|} \sum_{i=1}^{n} var(< \vec{x_i}, \vec{v} >) \implies v \text{ is the first principal component}
\end{aligned} \tag{2}$$

Hence we arrive at the regular form of PCA.

5 Question 5: Reccomendation system

We are only reccommending movies which have not yet been watched by them. So users who have watched all movies have 0 reccomendations.

Results are available in attached excel.

6 Question 6: Bootstrap

6.1 part a: Corelation

From data we get correaltion value $\rho = 0.52306627$

6.2 part b: Non-parametric bootstrap

```
def bootstrap(x,n):
    sample_idx = np.random.choice(n,n,replace=True)
    return x[sample_idx]

B = 1000
pho_list = []
for i in range(B):
    #get bootstrap sample
    sample_i = bootstrap(x,n=15)
    #get pho
    cor_mat = np.corrcoef(sample_i.T)
    pho_list.append(cor_mat[0,1])

plt.hist(pho_list,bins=50);
plt.title("Distribution of $\hat{corr}$ from non paramteric
    bootstrap")
```

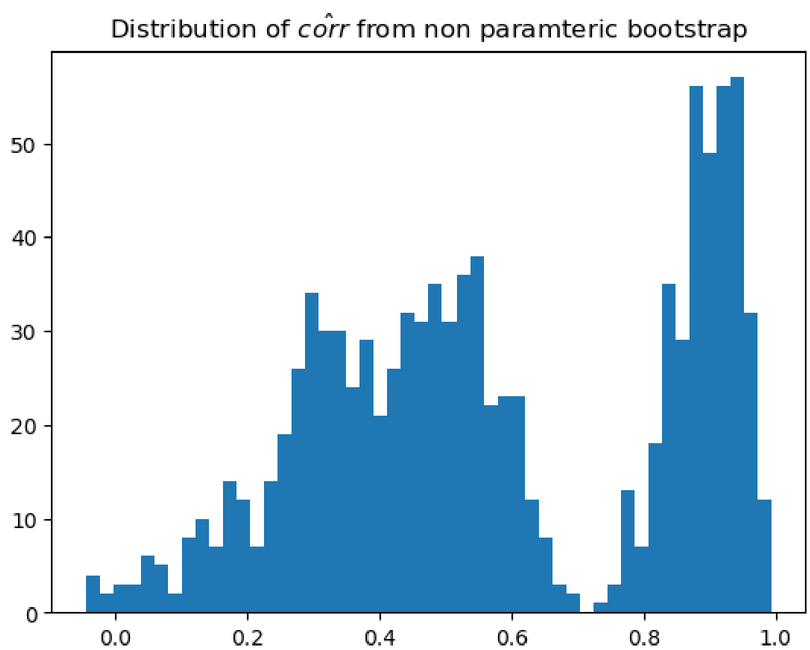

Figure 3: Bootstrap distribution of ρ

We get std. error from Bootstrap as = 0.26545402246210115
We get std. error from empirical 0.25 and 0.975 quantile as = 0.06190736, 0.96534972
We get 95% Confidence Interval =

$$(2\rho - q_{0.975}, 2\rho - q_{0.025}) = (0.08078282000000014, 0.9842251800000001)$$

6.3 part b: parametric bootstrap

```
mu_mle = x.mean(axis=0)
cov_mle = np.cov(x.T)

#paramteric bootstrap
B = 1000
pho_list = []
for i in range(B):
    #get bootstrap sample
    sample_i = np.random.multivariate_normal(mean=mu_mle,cov=
    cov_mle,size=15)
    #get pho
    cor_mat = np.corrcoef(sample_i.T)
    pho_list.append(cor_mat[0,1])

plt.hist(pho_list,bins=50);
plt.title("Distribution of $\hat{corr}$ from paramteric bootstrap")
```

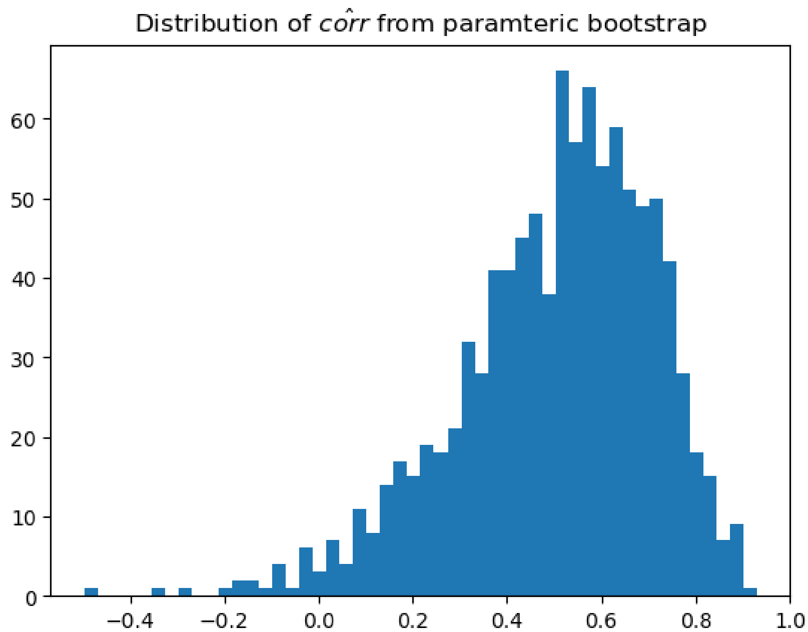

Figure 4: Bootstrap distribution of ρ

We get std. error from Bootstrap as = 0.20808987943210674
We get std. error from empirical 0.25 and 0.975 quantile as = [0.02167605, 0.83385225]
We get 95% Confidence Interval =

$$(2\rho - q_{0.975}, 2\rho - q_{0.025}) = (0.21228029000000015, 1.0244564900000002)$$

7 References:

1. **Collaborators**: Yibei, Rakesh