CAP 6610 Homework 4

Siddhant Mittal (6061-8545)

1. (a) Given:

$$\widetilde{C} = \frac{1}{n}\widetilde{\phi}\widetilde{\phi}^T$$

using $\widetilde{\phi} = (1 - \theta_1\theta_1^T)\phi$

$$= \frac{1}{n}(1 - \theta_1\theta_1^T)\phi((1 - \theta_1\theta_1^T)\phi)^T$$

$$= \frac{1}{n}(1 - \theta_1\theta_1^T)\phi\phi^T(1 - \theta_1\theta_1^T)$$

$$= \frac{1}{n}(1 - \theta_1\theta_1^T)\phi\phi^T(1 - \theta_1\theta_1^T)$$

$$= \frac{1}{n}(\phi\phi^T - \theta_1\theta_1^T\phi\phi^T - \phi\phi^T\theta_1\theta_1^T + \theta_1\theta_1^T\phi\phi^T\theta_1\theta_1^T)$$

using $\phi\phi^T\theta_1 = n\lambda_1\theta_1 => (\phi\phi^T\theta_1)^T = (n\lambda_1\theta_1)^T => \theta_1^T\phi\phi^T = n\lambda_1\theta_1^T$

$$\widetilde{C} = \frac{1}{n}(\phi\phi^T - \theta_1 n\lambda_1\theta_1^T - n\lambda_1\theta_1\theta_1^T + \theta_1 n\lambda_1\theta_1^T\theta_1\theta_1^T)$$

using $\theta_1^T\theta = 1$

$$\widetilde{C} = \frac{1}{n}\phi\phi^T - \lambda_1\theta_1\theta^T$$

using $C = \frac{1}{n}\phi\phi^T$

$$\widetilde{C} = C - \lambda_1\theta_1\theta^T$$

(b) Since $\theta_1, \theta_2, \theta_3.......\theta_k$ are the first k eigenvectors with largest eigenvalues of C, i.e the principal basis vectors, therefore

$$\lambda_1 \geq \lambda_2 \geq \lambda_3....... \geq \lambda_k$$

$$\widetilde{C}\theta_i = (\frac{1}{n}\phi\phi^T - \lambda_1\theta_1\theta^T)\theta_i$$

$$= \frac{1}{n}\phi\phi^T\theta_i - \lambda_1\theta_1\theta^T\theta_i$$

Using $\phi\phi^T\theta_i = n\lambda$

$$\lambda_i\theta_i - \lambda_1\theta_1\theta_1^T\theta_i$$

Since $\theta_1^T = 0$ for i $\neq$ 1

$$\widetilde{C}\theta_i = \lambda_i\theta_i$$

and for i = 1

$$C\theta_1 = \lambda_1\theta_1 - \lambda_1\theta_1\theta_1^T\theta_1$$
$$C\theta_1 = \lambda\theta_1 - \lambda_1\theta_1 = 0$$

Hence for i $\neq$ 1, θ_i is also a principle eigenvector of $\widetilde{C}$ with same eigenvalue λ_i. Also, θ_1 is an eigenvector of C with eigenvalue 0. In short $\widetilde{C}$ has θ_i as principle eigenvectors with eigenvalues $(0, \lambda_2, \lambda_3.....\lambda_k)$. Therefor λ_2 is the largest eigenvalue of $\widetilde{C}$ hence θ_2 is the first principle eigenvector.

(c) Below is the pseudo code

```
Def findFirstKEigenVectors(C,K,f):
        # List of lambda
        L= []
        # List of thetas = []
        T = []
        For i in range(K):
                lambda, theta = f(C)
                C = C - lambda * v * v . Transpose
                L.append(lambda)
                T.append(theta)
        Return T, L
```

2. EM algorithm where ($\Sigma_1 = \Sigma_2 = ... = \Sigma_k$)

   1. Initialization step: Initialize the means μ_c, co-variance Σ and π_c for c=1,2,....k number of clusters in our GMM model.

   2. Expectation step: Evaluate expectation ψ_{ic} where $\psi_i = E[y_i|x_i]$

$$\psi_{ic} = \frac{\pi_c N(x_i|\mu_c, \Sigma)}{\sum_{i=1}^{k} \pi_i N(x_i|\mu_c, \Sigma)}$$

   where $N(x|\mu_c, \Sigma) = det(2\pi\Sigma)^{-\frac{1}{2}} exp(-\frac{1}{2}(x - \mu_c)^T \Sigma^{-1}(x - \mu_c))$

   3. Maximization step: Re-estimate the parameters using current ψ_{ic} for each c do the following

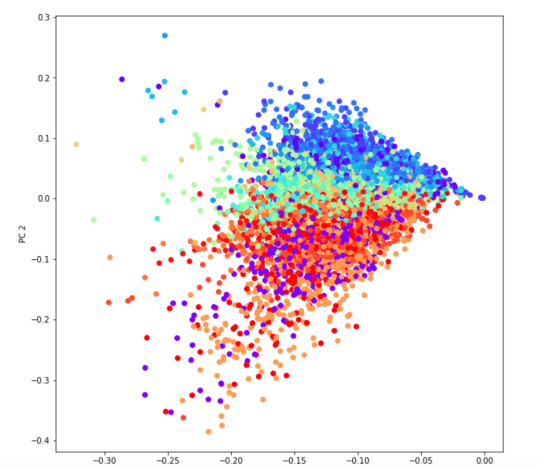
$$\mu_c{}^{new} = \frac{1}{\sum_{i=1}^{n} \psi_{ic}} \sum_{i=1}^{n} \psi_{ic} x_i$$

$$\pi_c{}^{new} = \frac{\sum_{i=1}^{n} \psi_{ic}}{n}$$

$$\Sigma^{new} = \frac{\sum_{i=1}^{n} \sum_{c=1}^{k} \psi_{ic}(x_i - \mu_c)(x_i - \mu_c)^T}{\sum_{i=1}^{n} \sum_{c=1}^{k} \psi_{ic}}$$

   4. Repeat from step 2 until converges.

3. (a)



(b)

Top 10 keys words for each clusters are shown below using my implementation of EM algorithm on GMM.

Top 10 key words for cluster: 0
disobeying

innermost

undecided

creatures

tokyo

outrageous

nas

exist

nrp

forbidden

Top 10 key words for cluster: 1
koran

suprised

rant

imminent

geoffrey

bifurcation

indicates

trading

wesleyan

chamberlain

Top 10 key words for cluster: 2
gaul

explaining

colour

moslems

ooo

italian

climbed

honor

es

door

Top 10 key words for cluster: 3
delusional

refuted

textbook

axis

papal

david

humbly

inherited

scatter

pains

Top 10 key words for cluster: 4
coincidence

fri

nrp

exist

necessity

innermost

creatures

shits

paradoxes

rusnews

Top 10 key words for cluster: 5
led

loan

bitmaps

rashid

dietary

environmental

hatching

possible

ellipses

neurons

Top 10 key words for cluster: 6
karl

quebec

stuff

symbol

mutable

kcochran

excommunicated

desire

heathers

accepts

Top 10 key words for cluster: 7
ago

advocate

clueless

specifies

dev

attest

could

excepting

replies

tiger

Top 10 key words for cluster: 8
pictured

surviving

doubtless

preying

cs

probability

manchester

misc

unpopular

replace

Top 10 key words for cluster: 9
uhhh

ends

workstation

chosen

afp

hav

dies

omnipotent

insanity

nc

Top 10 key words for cluster: 10
repressed

indonesians

consequently

cc

abdullah

talks

concluded

sociologist

merchants

travel

Top 10 key words for cluster: 11
hassles

anonymous

blanketing

vey

lasting

foes

outlaws

polemics

reported

uka

Top 10 key words for cluster: 12
dishes

backdrops

sanctions

build

chua

methodically

point

passes

findings

erroneous

Top 10 key words for cluster: 13
area

peculiar

dismissively

games

mchp

wpd

aesthetics

conveniently

adolescents

demands

Top 10 key words for cluster: 14
decline

lover

demos

reveal

chancellor

repeat

gradually

inherent

wise

shove

Top 10 key words for cluster: 15
timmbake

utoronto

frenzy

heavenly

polly

domains

contour

rushdie

saying

illiterate

Top 10 key words for cluster: 16
grating

prosecution

committs

cornflakes

shortcomings

steve

enforce

almighty

december

gross

Top 10 key words for cluster: 17
thrive

mailer

desires

viruses

capitalism

unwary

zen

spending

sim

reliability

Top 10 key words for cluster: 18
ames

revelations

marcel

factions

immersive

pop

summarized

extraordinary

king

pets

Top 10 key words for cluster: 19
replicates

subset

filled

generally

macalstr

paleontology

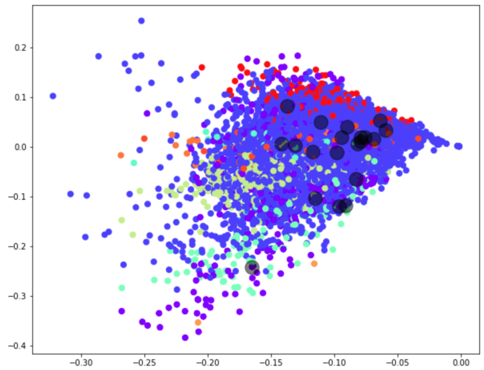
innermost

infectious

seldom

detection

Following figure shows the means of the 20 clusters on the 2D image of the data. It was hard to picture them as the dimension is 100. But I have used the 2 feature to plot the data on the 2D graph and their cluster centers means to give some visualization.



I also coded to compare with implementation of EM algorithm in the sklearn libraray to compare the two models were quite similar. Also The top 10 words in each cluster seems relate-able to each other so grouping documents with these words together in a cluster makes sense.