

ML HW4

Q2 Naive Bayes GMM

A2 Expectation-Maximization

Algorithm :-

→ The EM Algo. is an iterative method to find the M.L.E. or M.A.P. estimate for models with latent variables.

① Initialization - First we randomly initialize parameters: μ_c, Σ_c & π_c for $c=1, 2, \dots, K$ number of clusters in our GMM model

② Expectation Step - We evaluate expectation ψ_{ic} , where $\psi_i = E[y_i | x_i]$

$$\psi_{i,c} = \frac{\pi_c N(x_i | \mu_c, \Sigma_c)}{\sum_{i=1}^K \pi_i N(x_i | \mu_c, \Sigma_c)}$$

where $x_i = i^{\text{th}}$ observation &
 (Σ, μ, π) are the parameter values
 from step ①

② Maximization Step - We use the
 current $\psi_{i,c}$ for each c to re-estimate
 the parameters as follows:

$$\mu_c^{\text{new}} = \frac{1}{\sum_{i=1}^n \psi_{i,c}} \sum_{i=1}^n \psi_{i,c} x_i$$

$$\pi_c^{\text{new}} = \frac{\sum_{i=1}^n \psi_{i,c}}{n}$$

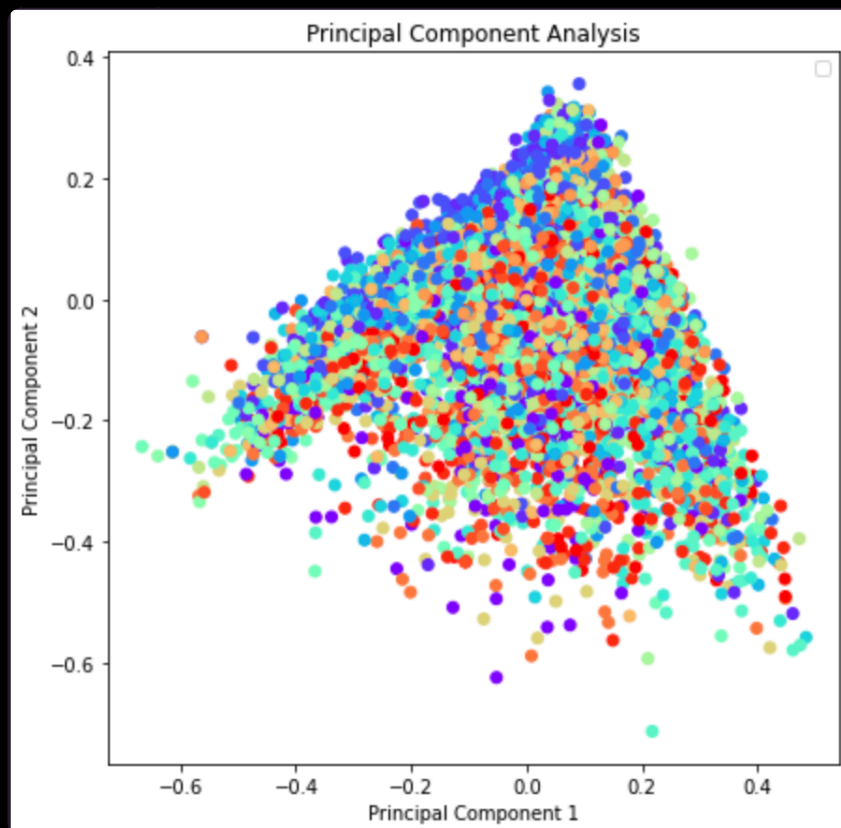
$$\Sigma_c^{\text{new}} = \text{diag} \left(\frac{1}{\sum_{i=1}^n \psi_{i,c}} \sum_{i=1}^n \psi_{i,c} (x_i - \mu_c)(x_i - \mu_c)^T \right)$$

④ Repeat from Step 2 until
converges

Q3 20 Newsgroup Revisited

A3 The code for the problem
part a & b is attached
as a ".py" file.

a



5

Top 10 words for Cluster: 1

from, last, modified, usa, addresses, last, last, religion, modified, version

Top 10 words for Cluster: 2

atheism, name, atheism, resources, atheism, modified, last, resources, last, atheism

Top 10 words for Cluster: 3

december, organizations, darwin, fish, organizations, of, usa, archive, name, religion

Top 10 words for Cluster: 4

modified, resources, resources, resources, resources, resources, resources, alt, resources, resources

Top 10 words for Cluster: 5

atheist, usa, freedom, name, archive, freedom, religion, usa, atheist, organizations

Top 10 words for Cluster: 6

from, resources, last, name, organizations, from, resources, freedom, modified, of

Top 10 words for Cluster: 7

december, december, december, modified, december, december, december, december, december, december

Top 10 words for Cluster: 8

religion, foundation, foundation, foundation, foundation, foundation, darwin, foundation, foundation, foundation

Top 10 words for Cluster: 9

atheism, atheism, atheism, atheism, atheism, atheism, atheism, name, atheism, atheism

Top 10 words for Cluster: 10

from, from, fish, from, from, fish, from, from, fish, from

Top 10 words for Cluster: 11

atheist, addresses, addresses, addresses, addresses, addresses, addresses, addresses, addresses, addresses

Top 10 words for Cluster: 12

name, organizations, version, archive, foundation, modified, addresses, version, resources, foundation

Top 10 words for Cluster: 13

freedom, freedom, last, last, last, alt, freedom, freedom, freedom, freedom

Top 10 words for Cluster: 14

usa, usa, alt, addresses, december, modified, modified, religion, organizations, atheism

Top 10 words for Cluster: 15

modified, freedom, addresses, modified, addresses, atheist, modified, modified, modified, modified

Top 10 words for Cluster: 16

atheist, last, usa, december, december, december, modified, organizations, organizations, alt

Top 10 words for Cluster: 17

of, of, of, of, of, of, of, of, of, of

Top 10 words for Cluster: 18

name, of, organizations, freedom, atheist, modified, organizations, of, atheism, alt

Top 10 words for Cluster: 19

last, last, last, last, last, last, last, last, last, last

Top 10 words for Cluster: 20

fish, freedom, version, usa, of, freedom, freedom, of, freedom, of

Q4

Given,

$$Pr(y_i = c) = \frac{1}{K}$$

$$Pr(x_i | y_i = c) \sim \text{Multi}(p_c, L_i^c)$$

$$P(x_i | y_i = c) = \frac{L_i^c!}{\prod_{j=1}^d x_{ij}!} \prod_{j=1}^d p_{c_j}^{x_{ij}}$$

(a) For M.L. formulation for estimating P_1, P_2, \dots, P_K we find the log likelihood of probability mass function

$$\Rightarrow \sum_{i=1}^n \log P(x_i, y_i; P_c)$$

where $c = 1, 2, \dots, K$

$$\Rightarrow \sum_{i=1}^n \log P(y_i = c) + \sum_{i=1}^n \log P(x_i | y_i; P_c)$$

$$\Rightarrow \sum_{i=1}^n \log(1/K) + \sum_{i=1}^n \log \left(\frac{L_i!}{\prod_{j=1}^d L_{ij}!} \prod_{j=1}^d P_{cj}^{x_{ij}} \right)$$

$$\Rightarrow -n \log(K) + \sum_{i=1}^n \sum_{j=1}^d x_{ij} \log P_{cj} + C$$

where $C =$ parameter free value

\Rightarrow This can be generalized as follows:

$$L = - \sum_{c=1}^K P\{y_i = c\} n \log K + \sum_{i=1}^n \sum_{c=1}^K \sum_{j=1}^d x_{ij} P\{y_i = c\} \log P_{cj} + C \longrightarrow \textcircled{1}$$

where $P\{y_i = c\} = 1 \text{ or } 0$

⑥ E-step:

For expectation calculation we compute the expected log likelihood under multinomial distribution.

$$\begin{aligned} E_p(\cdot, x_i, P_c) \left[\sum_{i=1}^n \log P(x_i, y_i; P_c) \right] \\ = - \sum_{i=1}^n \sum_{c=1}^K \phi_{ic} \log K + \sum_{i=1}^n \sum_{c=1}^K \sum_{j=1}^d x_{ij} \phi_{ic} \log(P_{cj}) \end{aligned}$$

where $\phi_{ic} = P(y_i = c / x_i; P_c)$ as in the GMM case.

\Rightarrow We can compute this by Bayes Rule:

$$P(y_i = c | x_i; P_c) \propto \frac{P(y_i = c) P(x_i | y_i; P_c)}{P(x_i; P_c)}$$

$$\psi_{ic} = \frac{1}{\sum_{d=1}^K \frac{1}{K} f(x_i; P_d)}$$

$\hookrightarrow \textcircled{2}$

where $f(x_i; P_c) = \frac{L!}{\prod_{j=1}^d x_{ij}!} \prod_{j=1}^d p_{cj}^{x_{ij}}$

\Rightarrow M-Step:

\rightarrow Since probability $P(c=k)$ is fixed, we will optimize it for P_c only. So

$$\frac{\partial \mathcal{L}}{\partial p_{cj}} = 0 \quad \& \quad \sum_{j=1}^d p_{cj} = 1 \quad \left[\text{this is eq. } \textcircled{1} \right]$$

$\hookrightarrow \textcircled{3}$

⇒ Using Lagrange's Multiplier

$$LLP = LL + \lambda_1 \left(1 - \sum_{j=1}^d p_{c_j}^o\right)$$

⇒ Now we set the derivative LLP to 0.

$$\rightarrow \sum_{i=1}^n \frac{x_{ij}}{p_{c_j}^o} \psi_{ic} - \lambda_1 = 0$$

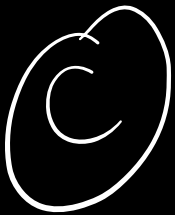
$$p_{c_j}^o = \frac{1}{\lambda_1} \sum_{i=1}^n x_{ij} \psi_{ic}$$

$$\text{Now, } \sum_{j=1}^d \frac{1}{\lambda_1} \sum_{i=1}^n x_{ij} \psi_{ic} = 1 \quad [\text{using } \textcircled{2}]$$

$$\lambda_1 = \sum_{j=1}^d \sum_{i=1}^n x_{ij} \psi_{ic}$$

$$= \sum_{i=1}^n x_i \psi_{ic} \quad (x_i = \text{total number of words})$$

$$\mu_{cj} = \frac{\sum_{i=1}^n x_{ij} \psi_{ic}}{\sum_{i=1}^n x_{ij} \psi_{ic}}$$



Top 10 key words for cluster: 0

proficient, locking, seeing, infinite, handedly, sling, reactions, ignor, workstations, maiming

Top 10 key words for cluster: 1

over, accor, pasted, dichotomy, guardian, previously, ole, accepting, art, money

Top 10 key words for cluster: 2

criteria, loses, ideals, dgraham, robertson, programmer, venture, alternatives, scheduled, wimped

Top 10 key words for cluster: 3

blacksburg, detrimental, multitude, critiques, recount, calm, loops, puma, sympathy, rit

Top 10 key words for cluster: 4

analysed, mirage, realidad, pushing, pillage, pose, eactly, incompetent, letting, guardian

Top 10 key words for cluster: 5

acquittal, fossil, giant, comet, monotheism, guardian, nextstep, executes, observing, personal

Top 10 key words for cluster: 6

observing, stopping, criteria, focusing, hair, abrax, manufacturers, remainder, broadly, oppress

Top 10 key words for cluster: 7

drink, witty, ancestors, sympathy, announce, options, priest, skeptical, ni, establishing

Top 10 key words for cluster: 8

kid, theologians, trmetu, newswire, languages, iron, antibigot, mentioning, theme, prophecies

Top 10 key words for cluster: 9

loop, sincerity, southern, broadly, rd, confess, belly, focusing, wrongs, differs

Top 10 key words for cluster: 10

eras, davidr, expert, cultures, box, plug, extremists, compaq, essentially, wrongs

Top 10 key words for cluster: 11

contains, denmark, newswire, tsang, dates, opponent, responded, evolutionism, capture, kid

Top 10 key words for cluster: 12

easter, baucaille, mic, prayer, luther, comparisons, newer, start, doctrine, window

Top 10 key words for cluster: 13

drinking, likes, warren, fishing, dismissed, unhappy, qif, helpful, measuring, bias

Top 10 key words for cluster: 14

chimps, beef, scripture, comet, chapters, contrived, resorting, airlines, christopher, compressed

Top 10 key words for cluster: 15

byte, character, backwards, stones, handle, hair, administrator, introns, eckart, checker

Top 10 key words for cluster: 16

blanketing, outlaws, vey, foes, anonymous, hassles, domain, proving, exile, landscape

Top 10 key words for cluster: 17

familiar, prophecies, interpretation, dishonesty, october, demanding, reiterate, mentioning, ida, objs

Top 10 key words for cluster: 18

bruno, domestic, ere, cwru, hardcopies, zach, social, phishnet, mish, spank

Top 10 key words for cluster: 19

gmeds, replication, afford, lands, bunk, predictably, speaking, representative, copeland, violence

Q1 Alternating optimization for PCA.

(A1) $\Phi = [\phi_1, \phi_2, \dots, \phi_n]$ &

$$Y = [y_1, y_2, \dots, y_n]$$

minimize $\|\Phi - \Theta Y\|^2$
 Θ, Y

\Rightarrow The 2 steps for alternating optimization are:

① fixing Y & finding optimal Θ

② fixing Θ & finding optimal Y

\rightarrow Initializing Θ & Y to random values & then performing 4b

steps will converge to a local minima
→ This local minima is also the global
minima due to the convex nature
of least squares. This gives us the
optimal solution.

$$\textcircled{1} \quad \nabla_y f(\theta, y) = 2(\theta^T y - \theta^T \underline{I}) \\ \Rightarrow y = (\theta^T \theta)^{-1} (\theta^T \underline{I})$$

$$\textcircled{2} \quad \nabla_{\theta} f(\theta, y) = 2(\underline{I} y y^T - \underline{I} y^T) \\ \Rightarrow \theta = (\underline{I} y y^T)^{-1} (\underline{I} y)$$