U 8941-5490

Ans to the Question no 1 (a)

Qiz - 3 Date :

Ans to the Que 1(a)

$$P(C_n | X_n) = \frac{P(X_n | C_n)\, P(C_n)}{P(X_n)}$$

$$\Rightarrow \max P(C_n | X_n) \simeq \max P(X_n | C_n)\, P(C_n)$$

where, $P(C_n | X_n) \Rightarrow$ posterior probability

$$\boxed{P(X_n | C_n)} \simeq \mathcal{N}(X_n | \mu_1, \mu_0, \Sigma_1, \Sigma_0)$$

likelihood.

$P(C_n) \simeq$ is prior class probability.

$$C_n \in \{C_1, C_2\} = \{1, 0\}$$

let's assume, $P(C_1) = \pi = \pi_1$

so, $P(C_2) = (1 - \pi) = \pi_0$

Date :

$$P(c_n = 1 \mid X_n) \cong P(X_n \mid c_1)\, P(c_1)$$

$$= \mathcal{N}(x_n \mid \mu_1, \Sigma_1)\, \pi$$

$$P(c_n = 2 \mid X_n) \cong P(X_n \mid c_2)\, P(c_2)$$

$$= \mathcal{N}(X_n \mid \mu_0, \Sigma_0)\, (1 - \pi)$$

So the likelihood of X_n under GMM.

is, $P(X_n) = \sum_{C_n \in [0, 1]^2} P(X_n, C_n)$

$$= \sum_{k=0}^{1} \underbrace{\pi_k}_{\boxed{\text{prior of class label}}} \underbrace{\mathcal{N}(x_n \mid \mu_k, \Sigma_k)}_{\text{likelihood of Gaussian class model}}$$

Assuming the instances and the class labels are independent of each other.

$$P(X) = \log \prod_{n=1}^{N} P(x_n)$$

2

Date :..............................

$$\log P(X) = \sum_{n=1}^{N} \log P(X_n)$$

$$= \sum_{n=1}^{N} \log \sum_{k=0}^{1} \pi_k \frac{\exp\{-(X_n-\mu_k)^T \Sigma_k^{-1} (X_n-\mu_k)/2\}}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \quad \text{(i)}$$

$$\cancel{= \sum_{n=1}^{N} \log \left[\sum_{k=0}^{1} \pi_k \exp\{-(X_n-\mu_k)^T \Sigma_k^{-1} (X_n-\mu_k)/2\} \right]} \quad \text{(i)}$$

~~As the denominator~~ doesn't depend on ~~data X~~.

we can re-write eq^n (i) as

$$\sum_{n=1}^{N} \log \left[\left(\frac{\pi_0 \exp\{-(X_n-\mu_0)^T \Sigma_0^{-1} (X_n-\mu_0)/2\}}{(2\pi)^{d/2} |\Sigma_0|^{1/2}} \right) + \left(\frac{\pi_1 \exp\{-(X_n-\mu_1)^T \Sigma_1^{-1} (X_n-\mu_1)/2\}}{(2\pi)^{d/2} |\Sigma_1|^{1/2}} \right) \right]$$

$$= \sum_{n=1}^{N} \log P(X_n)$$

$$\Rightarrow \sum_{n=1}^{N} \log p(x_n) = \sum_{n=1}^{N} \left[(1-\pi)\mathcal{N}(x_n|\mu_0,\Sigma_0) + \pi\mathcal{N}(x_n|\mu_1,\Sigma_1) \right]$$

from eq^n (1) we ~~get~~ can write :-

$$\log P(X) = \sum_{n=1}^{N} \log P(x_n)$$

$$= \sum_{n=1}^{N} \log \sum_{k=0}^{1} \pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)$$

$$\max_{\mu_k} \log P(X) = \frac{\partial}{\partial \mu_k} \log P(X)$$

$$\Rightarrow \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)}{\sum_{k=0}^{1} \pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)} \frac{\Sigma_k^{-1}(x_n-\mu_k)}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} = 0$$

$$\Rightarrow \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)}{\sum_{k=0}^{1} \pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)} \Sigma_k^{-1}(x_n-\mu_k) = 0 \quad -(4)$$

Date :

$$\gamma(c_{nk}) = \frac{\pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)}{\sum_{k=0}^{1} \pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)}$$

$\downarrow$

posterior probability

So, we can re-write eq^n (ii) as.

$$\max_{\mu_k} \log P(X) = \frac{\delta}{\delta \mu_k} \log P(X)$$

$$\Rightarrow \sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)}{\sum_{k=0}^{1} \pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)} \Sigma_k^{-1} (x_n - \mu_k) = 0$$

$$\Rightarrow \sum_{n=1}^{N} \gamma(c_{nk}) \Sigma_k^{-1} (x_n - \mu_k) = 0$$

$$\Rightarrow \sum_{n=1}^{N} \gamma(c_{nk}) (x_n - \mu_k) = 0$$

$$\Rightarrow \mu_k = \frac{1}{\sum_{n=1}^{N} \gamma(c_{nk})} \sum_{n=1}^{N} \gamma(c_{nk}) x_n$$

Date :

assuming, $N_k = \sum_{n=1}^{N} \gamma(C_{nk})$

where N_k is the effective # of ~~points~~ data items in class "k".

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(C_{nk}) X_n$$

Similarly we will get,

$$\max_{\Sigma_k^{-1}} \log P(X) = \frac{\delta}{\delta \Sigma_k^{-1}} \log P(X)$$

$$= \frac{\delta}{\delta \Sigma_k^{-1}} \left[\sum_{n=1}^{N} \log \sum_{k=0}^{1} \pi_k \mathcal{N}(X_n \mid \mu_k, \Sigma_k) \right] = 0$$

$$\Rightarrow \sum_{n=1}^{N} \frac{\pi_k \cancel{\mathcal{N}(X_n \mid \mu_k, \Sigma_k)}}{\sum_{k=0}^{1} \pi_k \mathcal{N}(X_n \mid \mu_k, \Sigma_k)} \frac{\delta}{\delta \Sigma_k^{-1}} \left[\frac{\exp\left\{-\frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right\}}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \right] = 0$$

iii

Date :

$$\frac{\delta \ln f}{\delta \Sigma_k^{-1}} = \frac{1}{f} \frac{\delta f}{\delta \Sigma_k^{-1}}$$

$$\Rightarrow \frac{\delta f}{\delta \Sigma_k^{-1}} = \boxed{f \cdot \frac{\delta \ln f}{\delta \Sigma_k^{-1}}}$$

$$\therefore \frac{\delta}{\delta \Sigma_k^{-1}} \left[\frac{\exp\left\{-\frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right\}}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \right]$$

$$= \exp \frac{\exp\left\{-\frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right\}}{(2\pi)^{d/2} |\Sigma_k|^{1/2}}$$

$$\cdot \frac{\delta}{\delta \Sigma_k^{-1}} \ln \left[\frac{\exp\left\{-\frac{1}{2}(x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)\right\}}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \right]$$

$$= \mathcal{N}(x_n | \mu_k, \Sigma_k) \cdot \frac{\delta}{\delta \Sigma_k^{-1}} \ln \mathcal{N}(x_n | \mu_k, \Sigma_k)$$

* As dealing with Σ_k^{-1} is harder we ~~are~~ taking the derivative w.r.t. Σ_k^{-1} to further simplify.

Date :

$$\ln \mathcal{N}(x_n \mid \mu_k, \Sigma_k)$$

$$= -\frac{d}{2} 2\pi + \frac{1}{2} \ln |\Sigma_k^{-1}| - \frac{1}{2} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)$$

~~As, calculating~~ dealing with Σ_k^{-1}

$$\frac{\delta}{\delta \Sigma^{-1}} \ln \mathcal{N}(x_n \mid \mu_k, \Sigma_k)$$

$$= \cancel{\frac{\delta}{\delta \Sigma^{-1}} \left(\frac{d}{2} 2\pi \right)} +$$

$$= 0 + \frac{1}{2} \Sigma_k - \frac{1}{2} (x_n - \mu_k)^T (x_n - \mu_k)$$

~~Now replacing this to eq^n (iii)~~.

$$\sum_{n=1}^{N} \cancel{\frac{\pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{k=0} \pi \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}} \left[\Sigma_k - (x_n - \mu_k)^T (x_n - \mu_k) \right] = 0$$

9

Date :

$$\ln \mathcal{N}(x_n \mid \mu_k, \Sigma_k)$$

$$= -\frac{d}{2} 2\pi + \frac{1}{2} \ln |\Sigma_k^{-1}| - \frac{1}{2} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k)$$

As, ~~calculating~~ dealing with Σ_k^{-1}

$$\frac{d}{d\Sigma^{-1}} \ln \mathcal{N}(x_n \mid \mu_k, \Sigma_k)$$

$$= \cancel{\frac{d}{d\Sigma^{-1}} \left(\frac{d}{2} 2\pi \right)} +$$

$$= 0 + \frac{1}{2} \Sigma_k - \frac{1}{2} (x_n - \mu_k)^T (x_n - \mu_k)$$

~~Now replacing this to eq^n (iii)~~.

$$\sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_m \mid \mu_k, \Sigma_k)}{\cancel{\frac{1}{2\pi}} \sum_{k=0} \pi \mathcal{N}(x_n \mid \mu_k, \Sigma_k)} \left[\Sigma_k - (x_n - \mu_k)^T (x_n - \mu_k) \right] = 0$$

Date :

similarly we can calculate,

$$\pi_k = \frac{N_k}{N}$$

$$\underbrace{P(C_n = 1 \mid X_n)}_{\text{posterior probability}} = \frac{\overset{\text{prior class probability}}{P(C_n = 1)}\ \overset{\text{Likelihood } X_n \text{ being of class 1}}{P(X_n \mid C_n = 1)}}{P(X_n)} = \gamma(C_{n1})$$

$$P(C_n = 0 \mid X_n) = \frac{P(C_n = 0)\ \overset{\text{Likelihood of } X_n \text{ being in class 2}}{P(X_n \mid C_n = 0)}}{P(X_n)} = \gamma(C_{n0})$$

$$\boxed{\begin{aligned} P(C_n = 1) &= \pi \\ P(C_n = 0) &= 1 - \pi \end{aligned}}$$

$P(C_n = 1) = \pi$ → prior class probability of class 1

$P(C_n = 0) = 1 - \pi$ → prior class probability of Class 2

11

Date :

similarly we can calculate,

$$\pi_k = \frac{N_k}{N}$$

prior class probability

Likelihood X_n being of class 1

$$\underbrace{P(C_n = 1 \mid X_n)}_{\text{posterior probability}} = \frac{P(C_n = 1)\, P(X_n \mid C_n = 1)}{P(X_n)} = \gamma(C_{n1})$$

Likelihood of X_n being in class 2

$$P(C_n = 0 \mid X_n) = \frac{P(C_n = 0)\, \boxed{P(X_n \mid C_n = 0)}}{P(X_n)} = \gamma(C_{n0})$$

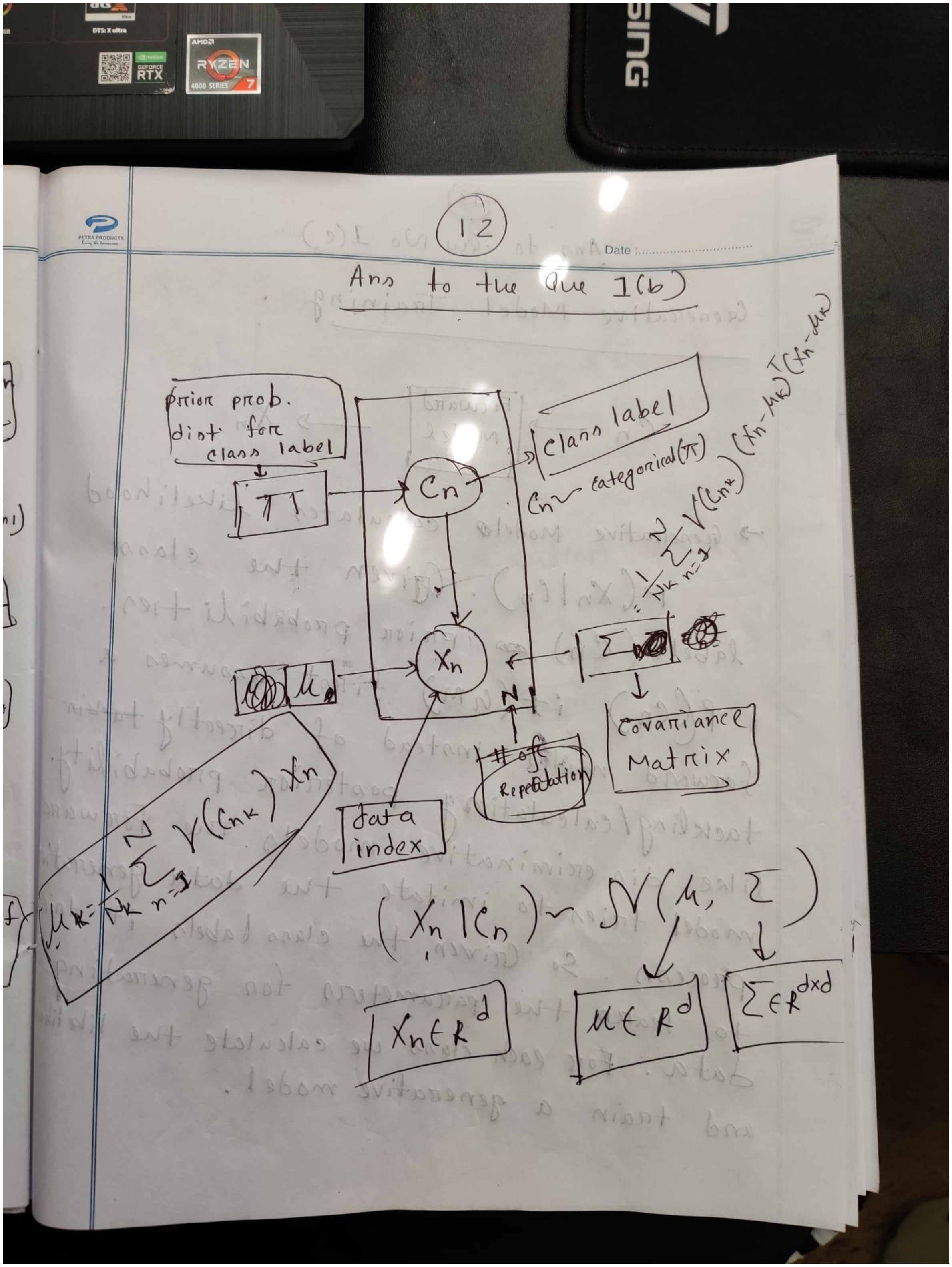
$$\boxed{\begin{aligned} P(C_n = 1) &= \pi \\ P(C_n = 0) &= 1 - \pi \end{aligned}}$$

prior class probability of class 1

prior class probability of class 2

$$P(X_n) = \sum_{k=0}^{1} P(C_n = k)\, P(X_n \mid C_n = k)$$ $\Rightarrow$ Evidence

Ans to the Question no 1 (b)



Ans to the Question no 1 (c)

Ans to Qu No 1(c)

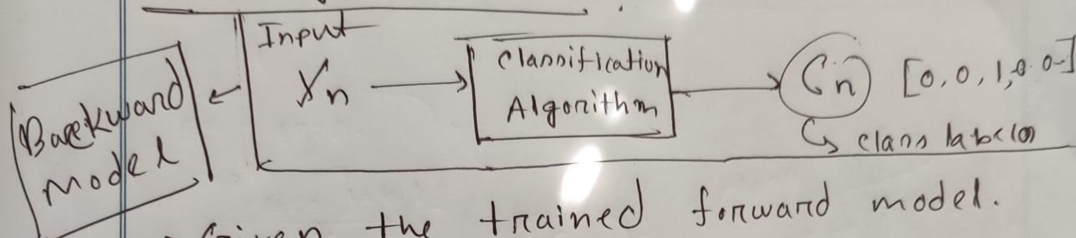
Generative Model Training :

$$C_n \longrightarrow \boxed{\text{Forward Model}} \longrightarrow X_n$$

→ Generative Models calculates likelihood $P(X_n|C_n)$. Given the class labels (C_n) prior probabilities $P(C_n)$ it (GM) first assumes a forward model instead of directly ~~tackin~~ tackling/calculating posterior probability like discriminative models. The forward model tries to imitate the data generation process. So Given the class labels it tries to learn the parameters for generating data. For each class we calculate the likelihood and train a generative model.

14

Testing with GM :



→ Given the trained forward model. in testing phase GM calculates the posterior distribution and ~~assigns~~ labels the test data ~~with~~ with the class ~~lab~~ having highest posterior probability.

$$P(C_n | X_n) = \frac{\overbrace{P(X_n | C_n)}^{\text{likelihood}} \; \overbrace{P(C_n)}^{\text{prior prob}}}{P(X_n)}$$

→ It is an indirect process. Which first calculates the likelihood and prior probability of train data. The uses those clculations to find out the posterior probability of a test data.

References for 1(a)

1. **The Matrix Cookbook - Mathematics**
2. Derivation of M-step in EM algorithm for mixture of Gaussians
3. Derivative of Gaussian mixture model with respect to covariance matrix Σk
4. EM of GMM appendix (M-Step full derivations)