POORNIMA FOUNDATION

LECTURE NOTES

Campus: PCE | Course: BTECH in CSE | Class/Section: III Yr. Section- A | Date: 9/3/21

Name of Faculty: Praveen Kumar Yadav | Name of Subject: Machine Learning | Code: 6CS4-02

Date (Prep.): 9/3/21 Date (Del.): 3-4-21 Unit No.: II Lect. No: 12

OBJECTIVE: To be written before taking the lecture (Pl. write in bullet points the main topics/concepts etc., which will be taught in this lecture)

Probablistic clustering.

IMPORTANT & RELEVANT QUESTIONS:

1. what is Gaussian distribution?

FEED BACK QUESTIONS (AFTER 20 MINUTES):

How a Gaussian distribution help in Probablistic clustering.

OUTCOME OF THE DELIVERED LECTURE: To be written after taking the lecture (Pl. write in bullet points about students' feedback on this lecture, level of understanding of this lecture by students etc.)

good

REFERENCES: Text/Ref. Book with Page No. and relevant Internet Websites:

scikit with ML, youtube video

Probabilistic clustering:- In Baysian Hierachical clustering instead of physical distance metrics such As Euclidean and Manhattan, distance, we use probabilistic distance between the data points are used.

☞ The probabilistic distance is given by bayesian posterior probability and Hence the name is, BHC.

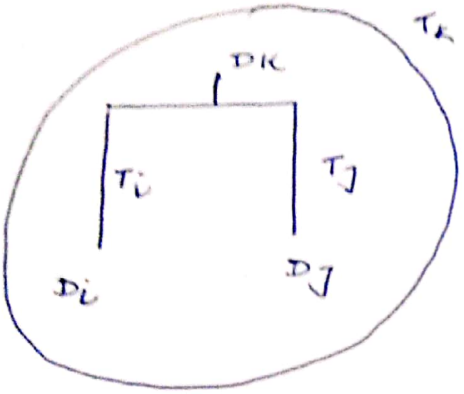
BHC Algorithm:-

Let $D = \{x^{(1)}, x^{(2)}, \ldots x^{(n)}\}$ represents the dataset of n elements.

Initially, All each data elements (x^{i}) is a single cluster, yielding a set of cluster at leaf-level, i.e $D_i \subset D$

These leaf level clusters can also be denoted as subtrees $\{T_i : 1 \ldots n\}$

(As we have 'n' subtrees initially denoting 'n' clusters and 'n' data elements)



$D_K = D_i \cup D_j$

Here subtrees are merged until a single tree is formed uniting all subtrees.

The Merging Criterion :- Here we have two Hypothesis H_1 + H_2 to merge clusters (subtrees)

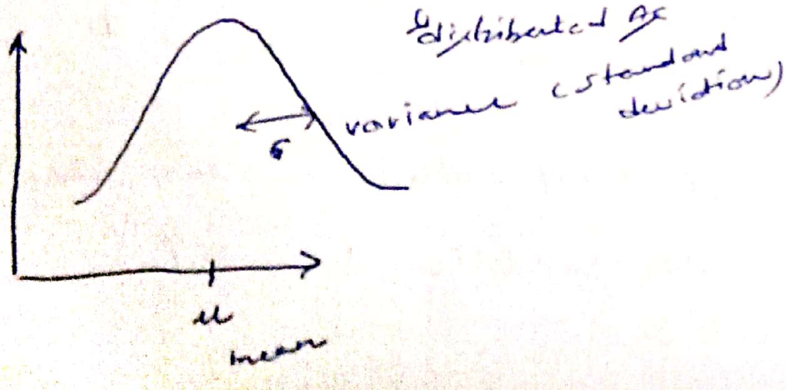
Hypothesis H1 :- All data elements belonging to the merged clusters (D_K) were independently and identically from the same probabilistic model (Gaussian Model) which has mean and variance pre. are unknown parameter

that is $P(x/\theta)$ - θ

where $\theta = (\underset{\text{mean}}{\mu}, \underset{\text{variance}}{\Sigma})$

Gaussian distribution - (Normal Dist.) Say x is random variable. If 'x' is a distributed gaussian.

$x \sim N(\mu, \sigma^2)$
distributed as



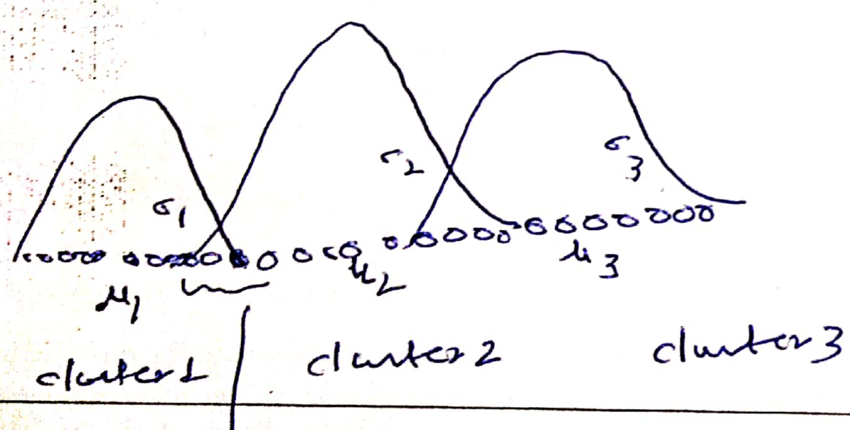
$$P(x : \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

for a given dataset: $\{x^1, x^2, \ldots x^n\}$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x^i \qquad \sigma^2 = \frac{1}{n}\sum_{i=1}^{m} (x^i - \mu)^2 \quad (2)$$

" The initiution is that all the "x" value is supposed to forming in single Gaussian distribution be in a single cluster.

↳ Like we have single Gaussian distribution we have also Gaussian Mixture models.



soft clustering of some points

$$P(D_k / H_1^k) = \int P(D_k/\theta)\, P(\theta/\beta)$$

$\theta \leftarrow \mu, \sigma^2$

$\beta \leftarrow$ Hyper parameters (features)

Hypothesis H_2 - The data subset in D_K has two or more clusters in H.

$$P(D_K / H_2^K) = P(D_i/T_i)\, P(D_J/T_J)$$

product of two subtrees (Joint prob. of the subtrees)

So we put all the two prob. in Bayes theorem.

$$P(D_K/T_K) = \underbrace{\pi_k P(D_K/H_1^K)}_{H_1} + (1-\pi_K)\underbrace{P(D_i/T_i)}_{H_2}\, P(D_J/T_J)$$

where $\pi_K = P(H_1^k)$ prior probability that all points in D_K belongs to single cluster

$\pi_2 = P(H_2^K)$