CMPE258 Spring 2023 (Part III) 1/

April 20 (Thursday).

Note 1. Final Exam.

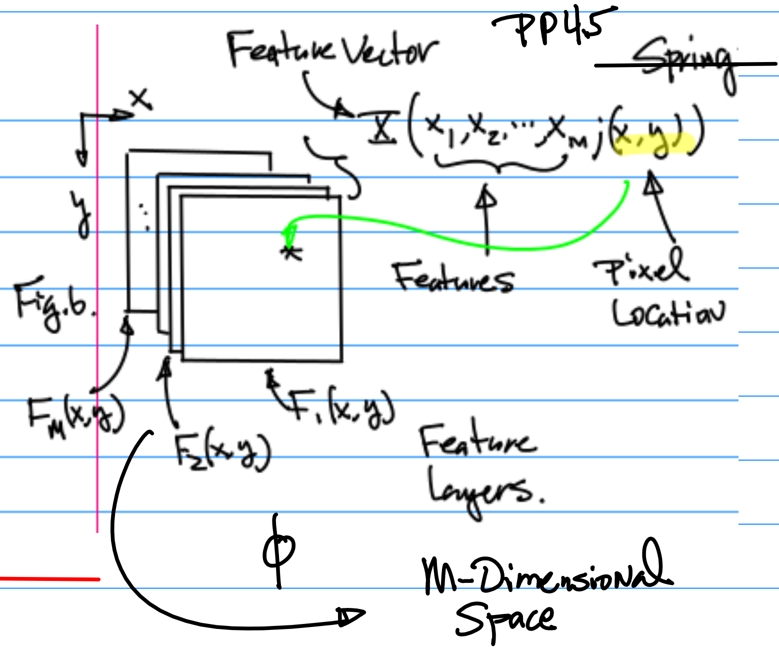
Cluster Analysis. "Mapping"

Group II Classes

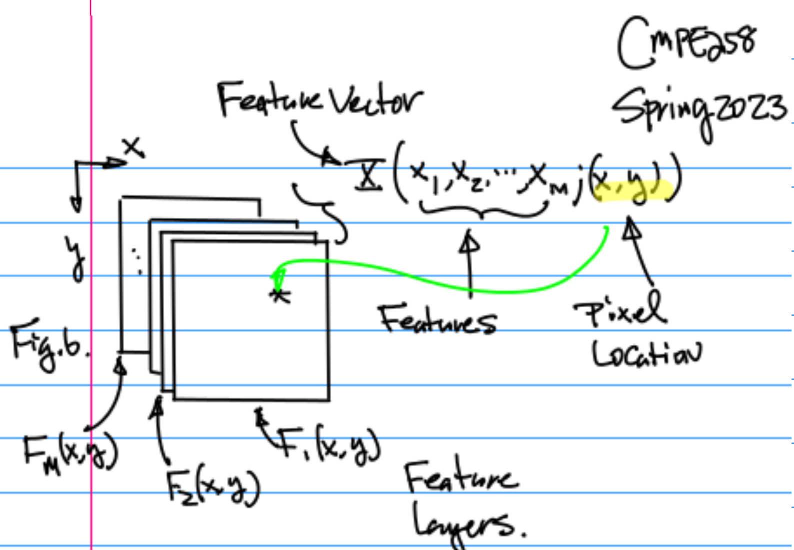
Group II classes are those classes which meet TR, T, R, TWR, MTR, TRF, MTI MTWR, TWRF, RF, TF, TRS.

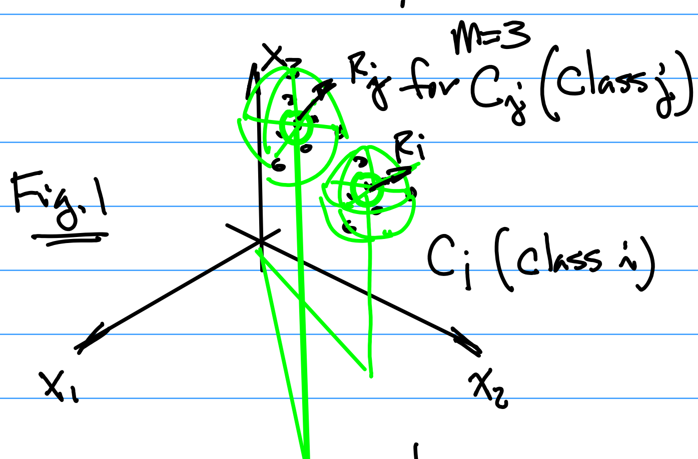
Regular Class Start Times	Final Examination Days	Final Examination
7:00 through 8:25 AM	Monday, May 22	7:15-9:30 AM
8:30 through 9:25 AM	Wednesday, May 17	7:15-9:30 AM
9:30 through 10:25 AM	Friday, May 19	9:45 AM-12:00 PM
10:30 through 11:25 AM	Tuesday, May 23	9:45 AM-12:00 PM
11:30 AM through 12:25 PM	Thursday, May 18	9:45 AM-12:00 PM
12:30 through 1:25 PM	Monday, May 22	12:15-2:30 PM
1:30 through 2:25 PM	Wednesday, May 17	12:15-2:30 PM
2:30 through 3:25 PM	Friday, May 19	2:45-5:00 PM
3:30 through 4:25 PM*	Tuesday, May 23	2:45-5:00 PM
4:30* through 5:25 PM*	Thursday, May 18	2:45-5:00 PM

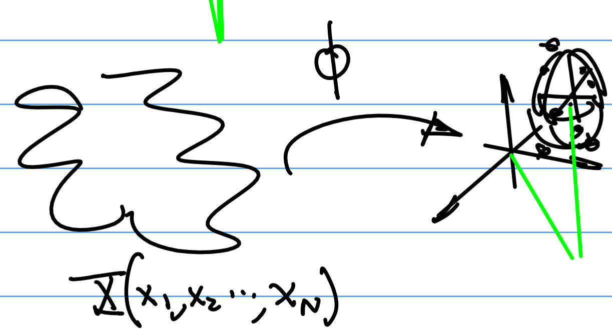
PP45 ~~Spring~~

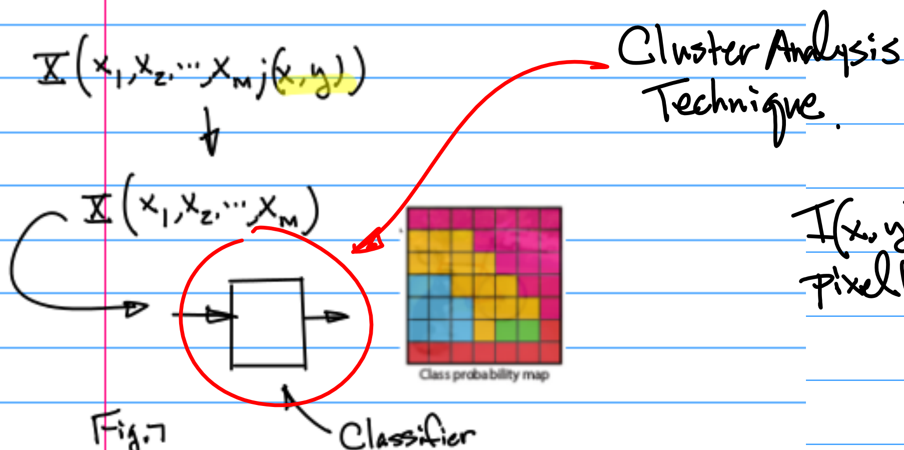


From Notes Part II, pp.45









$I(x,y)$ pixel(s) → Feature $X(x_1,x_2,\cdots,x_N)$ $\xrightarrow{\phi}$ Classification C_i

Prob(C_i)

CMPE258
Spring 2023 (III) 2/

Note: $\sum_{i=1}^{N} Prob(C_i) = 1$ (Previously M)

N: Number of Classes.

from "Heuristics" Expert Knowledge.

Consider K-mean Cluster Algorithm.

github.

- 2022S-114c-Kmean-handCalculation...
- 2022S-114c-KmeanCluster-v3-2022-... ← PPT

Example: K-mean Cluster Algorithm.

First, Notation. Note 1. Vectors

Given a set of observations ($\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$),

e.g. $X_1 = (x_{11}, x_{12}, \cdots, x_{1N})$ ...(1)

first observation

For X_i: observation i

x_{ij}: Component j for the Observation i

Note: if for d-dimensional Vector, then Eqn(1) has its $N = d$

K-mean: partition the n observations into k ($\le n$) sets $\mathbf{S} = \{S_1, S_2, \ldots, S_k\}$ ...(2)

$\{X_i \mid i = 1, 2, \cdots, N\}$

K-Classes.

S_i for Class i

Note 1. Note 2. Squared Distance

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \arg\min_{\mathbf{S}} \sum_{i=1}^{k} |S_i| \operatorname{Var} S_i \quad \ldots(3)$$

a. "Argmin" minimization

b. "S" Domain, "Scope" of the minimization

Example for $\|\vec{x} - \vec{\mu}_i\|^2$,

if $\vec{x} = (x_1, x_2)$, $\vec{\mu}_i = (\mu_{i1}, \mu_{i2})$

then

$$\|\vec{x} - \vec{\mu}_i\|^2 = \left(\sqrt{(x_1 - \mu_{i1})^2 + (x_2 - \mu_{i2})^2}\right)^2$$

$$= (x_1 - \mu_{i1})^2 + (x_2 - \mu_{i2})^2 \quad \ldots(4)$$

Note 3. "$\sum_{x \in S_i}$" Notation $\sum_{i=1}^{M} x_i$

Summation for each & every x as long as it is from the set S_i

Note 4: from Eqn (3), we have

S_i: Collection of vectors X belonging to Class i

Note 5: $\sum_{i=1}^{k}$ → to cover the Collection of all Classes.

Note 6: $\vec{\mu}_i$ ① $\vec{\mu}_i = (\mu_{i1}, \mu_{i2}, \cdots, \mu_{id})$

② Cluster for the Class i

April 25 (Tue)

Note 1. Quick update on Project Progress Report. (Next Lecture)

Example: Continuation of K-mean Cluster Algorithm.

$$S_i^{(t)} = \{x_p : \|x_p - m_i^{(t)}\|^2 \le \|x_p - m_j^{(t)}\|^2 \ \forall j, 1 \le j \le k\} \quad \ldots (1)$$

Ref

2022S-114c-Kmean-handCalculation1-c

Note 1: A set of Feature Vectors

$S_i^{(t)}$ ← Captured at Step t ; i ← Class id : ith Class

$S_i^{(t)} = \{ \vec{X}_p \quad \}$

index, $p = 1, 2, \ldots$ just like Notation i, j, or k

$S_i^{(t)} = \{ \vec{X}_p : \quad \}$

Condition

$\|\vec{X}_p - \vec{m}_i^{(t)}\|^2$

Distance (Squared) at time (t) to the Cluster of class i

$\|\vec{X}_p - \vec{m}_j^{(t)}\|^2$

... ... to the Cluster class j

" $\forall j$ " for Any j, such as

$1 \le j \le K$ covers all the different classes.

Hand Calculation Example

Given the following feature vectors use K-mean Algorithm to find the clusters.

$X_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $X_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $X_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ $X_4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$X_5 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ $X_6 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ $X_7 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ $X_8 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$

$X_9 = \begin{pmatrix} 6 \\ 6 \end{pmatrix}$ $X_{10} = \begin{pmatrix} 7 \\ 6 \end{pmatrix}$ $X_{11} = \begin{pmatrix} 8 \\ 6 \end{pmatrix}$ $X_{12} = \begin{pmatrix} 6 \\ 7 \end{pmatrix}$

$X_{13} = \begin{pmatrix} 7 \\ 7 \end{pmatrix}$ $X_{14} = \begin{pmatrix} 8 \\ 7 \end{pmatrix}$ $X_{15} = \begin{pmatrix} 9 \\ 7 \end{pmatrix}$ $X_{16} = \begin{pmatrix} 7 \\ 8 \end{pmatrix}$

$X_{17} = \begin{pmatrix} 8 \\ 8 \end{pmatrix}$ $X_{18} = \begin{pmatrix} 9 \\ 8 \end{pmatrix}$ $X_{19} = \begin{pmatrix} 8 \\ 9 \end{pmatrix}$ $X_{20} = \begin{pmatrix} 9 \\ 9 \end{pmatrix}$

Sol: Step 1. Define K=2 per Heuristics. Expert Knowledge

Note: "0" Initial step.

Let cluster

$\vec{m}_1^0 = \vec{X}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ ... (1)

$\vec{m}_2^0 = \vec{X}_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ ... (2)

Initial Arbitrary Values

"1" Class 1

And Arbitrarily assign Feature Vectors into 2 Classes.

Step 2. Use Eqn (1) To Compute the distance

$\|\vec{X}_p - \vec{m}_i^{(t)}\|^2$

and $\|\vec{X}_p - \vec{m}_j^{(t)}\|^2$

To Evaluate the Grouping of $\vec{X}_p$ to the Class i per Eqn (1).

If Eqn(1) holds good, then $\vec{X}_p$ Stays in the Class i.

O/W Re-assign $\vec{X}_p$ to the Class j.

Step 3. Update the Cluster (when New Grouping is formed)

$$m_1^{(t)}\Big|_{t=1} = \frac{1}{N}\sum_{\vec{X}_p \in S_1} \vec{X}_p$$

$\vec{X}_j$

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|}\sum_{x_j \in S_i^{(t)}} x_j \quad \cdots (2)$$

Total Number of Feature Vectors in the Class i

Step 4. Carry out the Computation with the New Cluster.
(To Decide if the grouping is final or to Continue updating Cluster Values)

Note: "Stop" if No Regrouping
O/W. Continue By Repeating the process, e.g. update Cluster Values, the Evaluate the grouping.

Step 5. Perform the Computation as Described in Step 4. which Leads to

C_1 (Class 1): $S_1 = \{\vec{X}_1, \vec{X}_2, \cdots, \vec{X}_8\}$

$S_2 = \{\vec{X}_9, \vec{X}_{10}, \cdots, \vec{X}_{20}\}$

then, update the Cluster $\vec{m}_1^{(t)}, \vec{m}_2^{(t)}$.

↓

Check. No New Grouping

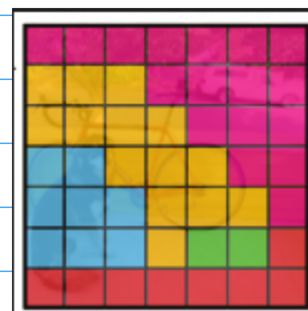
↓

$\vec{m}_1^{(t+1)}, \vec{m}_2^{(t+1)}$ are the same.

$\therefore$ Stop. (Converged).

Discussion on Probability Distribution map.

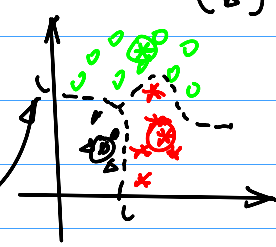
(a)



Class probability map

Feature Vector map.

(b)

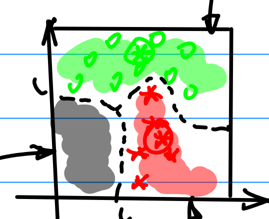


Boundary for the Classification

Regular K-mean Cluster Algorithm

C_3

(c)



C_1

C_2

Fig.1

$$\text{Prob}(C_1) \overset{?}{=} \frac{\text{Area of Black Pixels}}{\text{Area of } \Omega \text{ (Image Plane)}} \quad \cdots (3a)$$

where Area of Black Pixels can be computed.

Area of Ω(Image Plane) = Resolution of the image plane.

For Example, 448×448

Similarly, find

$$\text{Prob}(C_2) = \frac{\text{Area of Red Pixels}}{\text{Area of } \Omega(\text{Image Plane})} \quad \ldots (3b)$$

$$\text{Prob}(C_3) = \frac{\text{Area of green Pixels}}{\text{Area of } \Omega(\text{Image Plane})} \quad \ldots (3c)$$

then

$$\sum_{i=1}^{N} \Big|_{N=3} \text{Prob}(C_i) = 1 \quad \ldots (4)$$

Since

$$\sum_{i=1}^{3} \text{Prob}(C_i) = \text{Prob}(C_1) + \text{Prob}(C_2) + \text{Prob}(C_3) \quad \ldots (4\text{-}b)$$

$$= \frac{\text{Area of Black Pixels}}{\text{Area of } \Omega(\text{Image Plane})} + \frac{\text{Area of Red Pixels}}{\text{Area of } \Omega(\text{Image Plane})} + \frac{\text{Area of green Pixels}}{\text{Area of } \Omega(\text{Image Plane})}$$

$$= \frac{\text{Area of } \Omega(\text{Image Plane})}{\text{Area of } \Omega(\text{Image Plane})} = 1.$$

//