K-mean Algorithm (1)

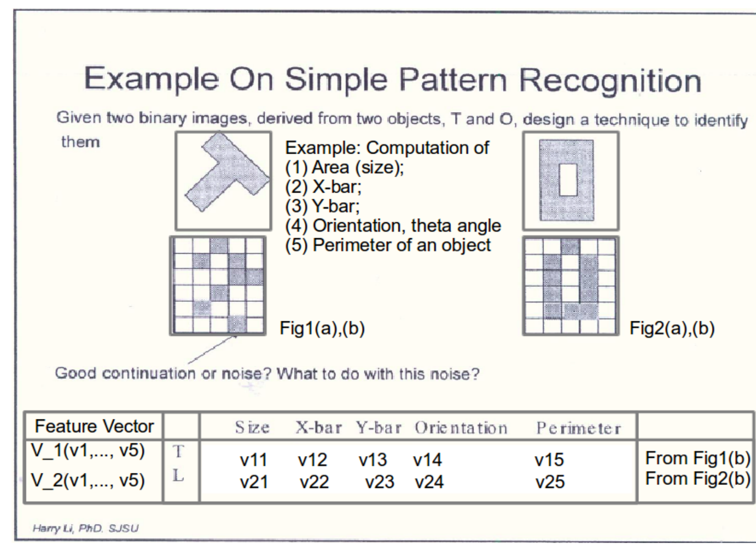
https://en.wikipedia.org/wiki/K-means_clustering

Given a set of observations ($\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_n$), where each observation is a *d*-dimensional real vector, *k*-means clustering aims to partition the *n* observations into *k* ($\leq n$) sets $\mathbf{S} = \{S_1, S_2, ..., S_k\}$ so as to minimize the within-cluster sum of squares (WCSS) (i.e. variance). Formally, the objective is to find:

$$\underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \left\| \mathbf{x} - \boldsymbol{\mu}_i \right\|^2 = \underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} |S_i| \operatorname{Var} S_i$$

Example:

Form feature vectors



Example On Simple Pattern Recognition

Given two binary images, derived from two objects, T and O, design a technique to identify them

Example: Computation of
(1) Area (size);
(2) X-bar;
(3) Y-bar;
(4) Orientation, theta angle
(5) Perimeter of an object

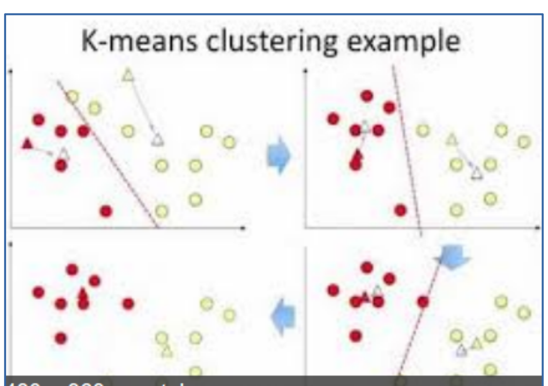
Fig1(a),(b)

Fig2(a),(b)

Good continuation or noise? What to do with this noise?

Feature Vector		Size	X-bar	Y-bar	Orientation	Perimeter	
V_1(v1,..., v5)	T	v11	v12	v13	v14	v15	From Fig1(b)
V_2(v1,..., v5)	L	v21	v22	v23	v24	v25	From Fig2(b)

Harry Li, PhD. SJSU



Cluster Seekingg

K-mean Algorithm (2)

https://en.wikipedia.org/wiki/K-means_clustering

Assignment step: Assign each observation to the cluster whose mean has the least squared Euclidean distance, this is intuitively the "nearest" mean.[7] (Mathematically, this means partitioning the observations according to the Voronoi diagram generated by the means).

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

Update step: Calculate the new means to be the centroids

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

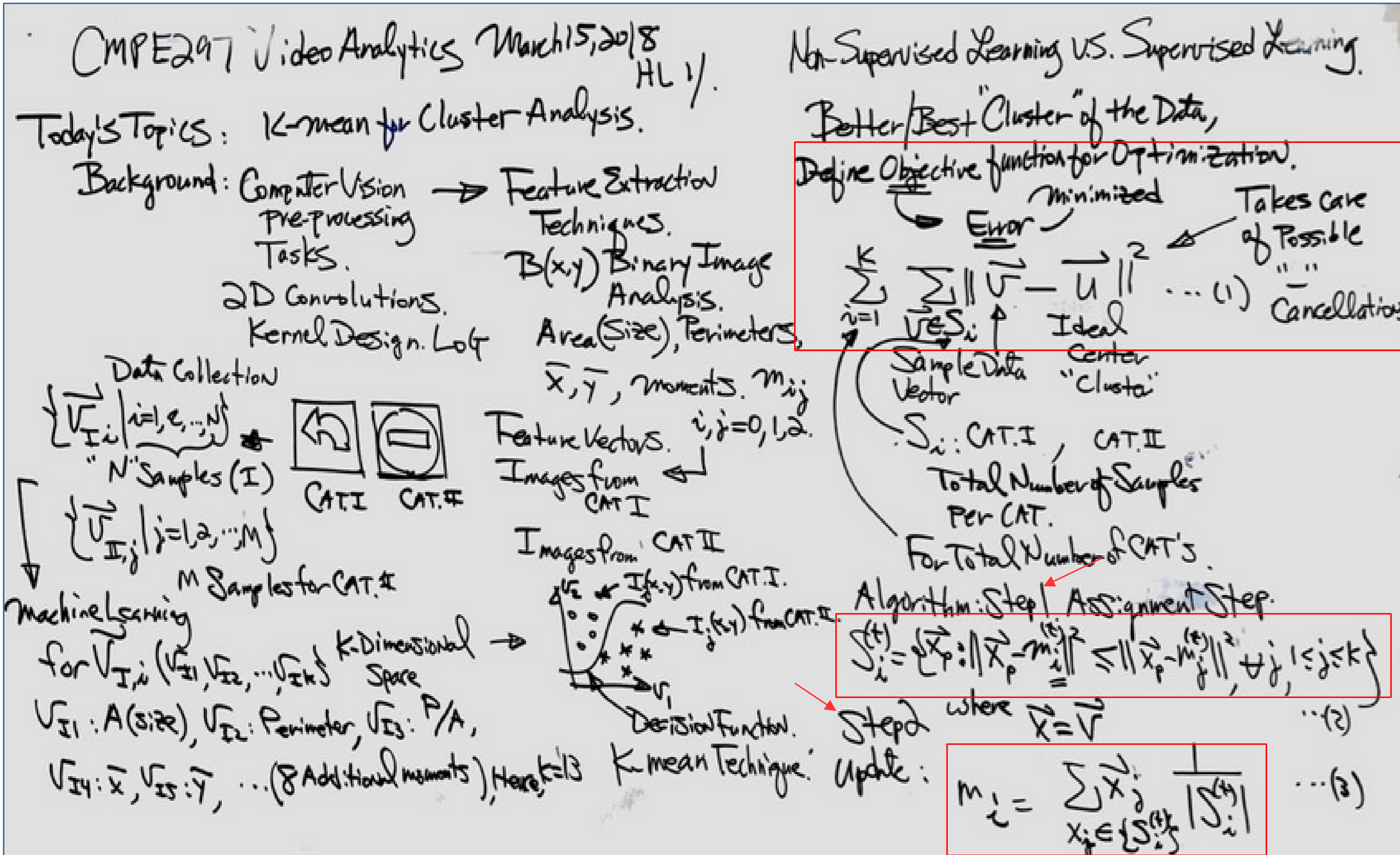
Algorithm:

1) Assum Number of Classes $= K$.
   Pick $\vec{m}_i(t)\big|_{t=1}$, $i=1,2,\dots,K$.
   Arbitrary.
   e.g. $\vec{m}_i(t) = \vec{X}_i$

2) $\|\vec{X}_\ell - \vec{m}_i(t)\| \le \|\vec{X}_\ell - \vec{m}_j(t)\|$ for $j \ne i$, $1 \le j \le K$
   Regroup $\vec{X}_\ell$, such that $\vec{X}_\ell \in \omega_i(S_i(t))$

3) Update $\vec{m}_i(t+1)$ for $i=1,2,\dots,K$, Check
   Termination $\vec{m}_i(t+1) = \vec{m}_i(t)$ Yes, Stop
   No, Goto 2

K-mean Example



Hand Calculation

CMPE 297 March 15, 2018 HL. 21.

Example: Given 2 CATs (classes), find the means, m_j, for $j=1,2$ Based ON K-mean Algorithm.

(See Handout pp1.) In Class.

$\vec{X} : \vec{V}$

Notation

Step 1. Arbitrary 2 pts as 2 classes & means $K=2$

$$m_i(t) \begin{cases} m_1(1) = \vec{X}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ m_2(1) = \vec{X}_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = (1,0)^t \end{cases}$$

mean.

Step 2. Compare Each pt with the mean from Each Class, to make sure Eqn(2) Satisfied.

$$\|\vec{X}_1 - \vec{M}_1\| \le \|\vec{X}_1 - \vec{m}_2\| \text{ holds good}$$

Since,

$$\left\| \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} - \begin{pmatrix} m_{11} \\ m_{12} \end{pmatrix} \right\| = (x_{11}-m_{11})^2 + (x_{12}-m_{12})^2 = 0^2 + 0^2 = 0$$

Similarly

$$\|\vec{X}_1 - \vec{m}_2\| = (x_{11}-m_{21})^2 + (x_{12}-m_{22})^2 = (0-1)^2 + (0-0)^2 = 1 \to \vec{X}_1 \in \omega_1 \left(S_1(t+1)\right)$$

or,

And $\vec{X}_2$, we have

$$\|\vec{X}_2 - \vec{m}_2\| \le \|\vec{X}_2 - \vec{m}_1\|$$

or. $\vec{X}_2 \in \omega_2 \left(S_2(t+1)\right)$

for $\vec{X}_3$,

$$\underbrace{\|\vec{X}_3 - \vec{m}_1\|}_{=1} \le \underbrace{\|\vec{X}_3 - \vec{m}_2\|}_{=2}$$

$\vec{X}_3 \in \omega_1 \left(S_1(1)\right)$ or.

Step 3. Update means.

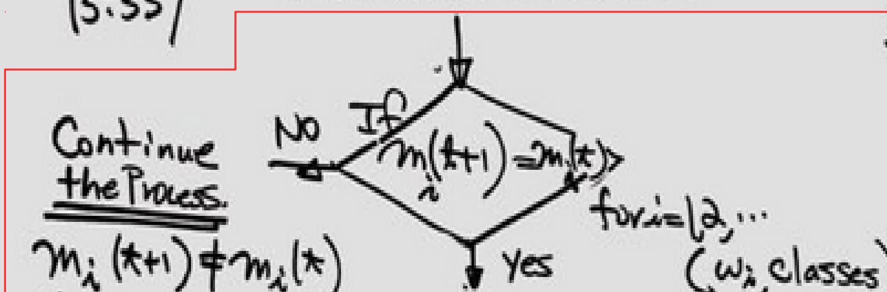
$$\vec{m}(2) = \frac{1}{|S_1(1)|} \sum_{x_j \in \{S_1(1)\}} \vec{X}_j = \frac{1}{2}\left(\vec{X}_1 + \vec{X}_3\right) = \frac{1}{2}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix}$$

K-mean

CMPE297 March 5, 2018, Harry Li 3/

$$\vec{m}_2(2) = \frac{1}{|S_2(1)|} \sum_{j \in \{S_2(1)\}} \vec{X}_j = \frac{1}{18}\left(\vec{X}_2 + \vec{X}_4 + \cdots + \vec{X}_{20}\right) = \begin{pmatrix} 5.67 \\ 5.33 \end{pmatrix}$$

Terminate the Process?



Step 4

$$\|\vec{X}_l - \vec{m}_1(2)\| \leq \|\vec{X}_l - \vec{m}_2(2)\|, \text{ for } l = 1, 2, \cdots, 8$$

and

$$\|\vec{X}_p - \vec{m}_2(2)\| \leq \|\vec{X}_p - \vec{m}_1(2)\|, \text{ for } p = 9, 10, \cdots, 20$$

Hence, $S_1(2) = \{\vec{X}_1, \vec{X}_2, \cdots, \vec{X}_8\}$, $S_2(2) = \{\vec{X}_9, \vec{X}_{10}, \cdots, \vec{X}_{20}\}$

Update mean,

$$\vec{m}_1(3) = \frac{1}{10}\left(\vec{X}_1 + \vec{X}_2 + \cdots + \vec{X}_8\right) = \begin{pmatrix} 1.25 \\ 1.33 \end{pmatrix}$$

$$\vec{m}_2(3) = \frac{1}{10}\left(\vec{X}_9 + \vec{X}_{10} + \cdots + \vec{X}_{20}\right) = \begin{pmatrix} 7.67 \\ 7.33 \end{pmatrix}$$

Check if $\vec{m}_i(t+1) \overset{?}{=} \vec{m}_i(t)$, $i = 1, 2$.

Since, not, continue,

Step 5. Check

$$\|\vec{X}_l - \vec{m}_1(3)\| \overset{?}{\leq} \|\vec{X}_l - \vec{m}_2(3)\|$$

for $l = 1, 2, \cdots, 8$.

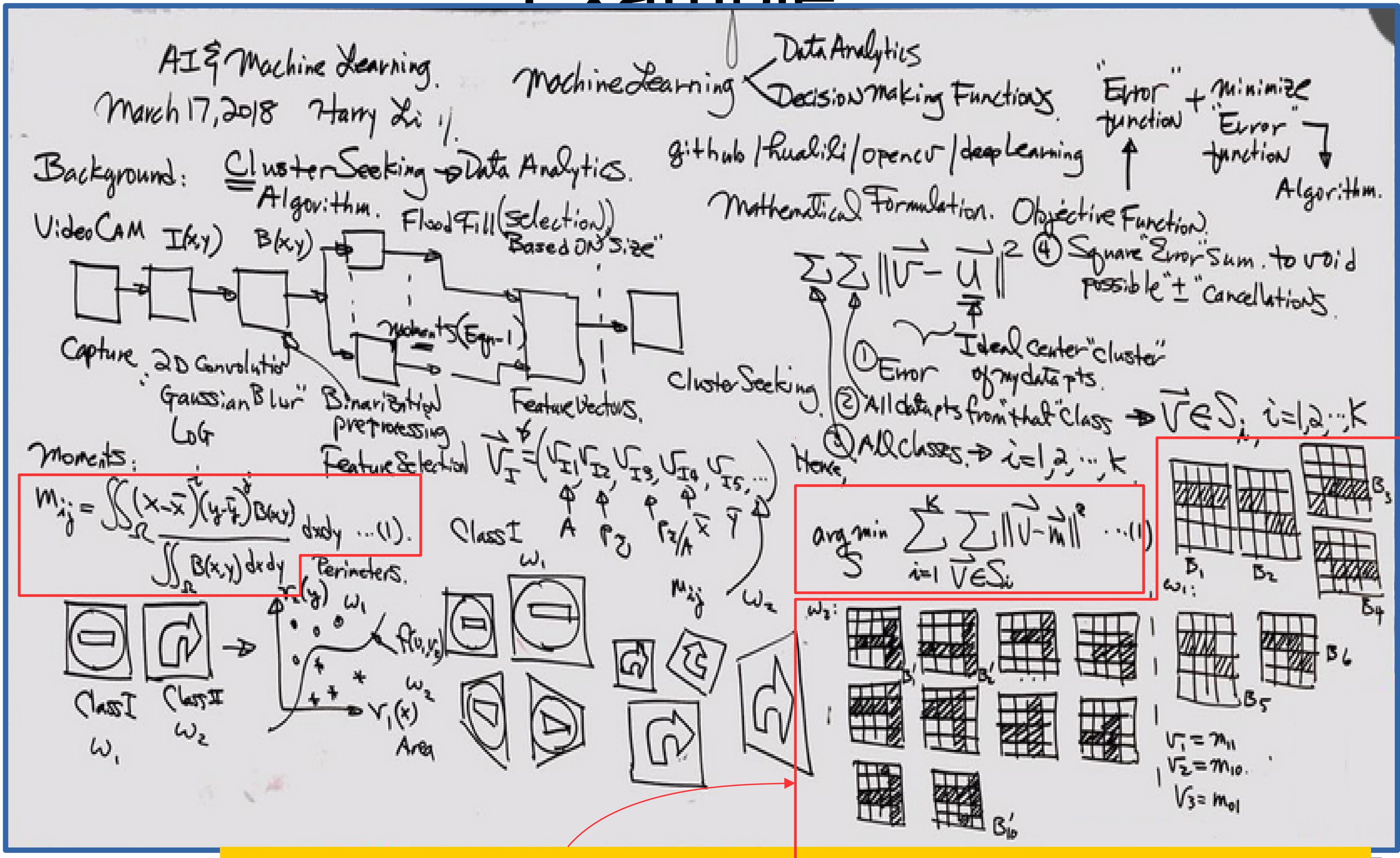
Similarly, $\|\vec{X}_p - \vec{m}_2(3)\| \overset{?}{\leq} \|\vec{X}_p - \vec{m}_1(3)\|$

Both is Satisfied, No Regrouping, therefore,

$\vec{m}_i(t+1) \overset{?}{=} \vec{m}_i(t)$, for $i = 1, 2$.

Yes! Converged, Stop.

Use kmean.py To Compute Example



Example: w1: B1(x,y), …, B6(x,y), w2: B'1(x,y), …, B'10(x,y), V = (v1, v2, v3), where v1 = m11, v2=m10, v3=m01, find m1 and m2 by using kmean.py

kmean.py Example