

K-mean Algorithm (1)

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

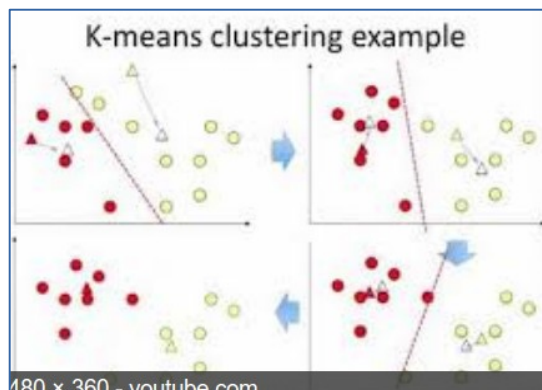
Given a set of observations ($\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$), where each observation is a d -dimensional real vector, k -means clustering aims to partition the n observations into k ($\leq n$) sets $\mathbf{S} = \{S_1, S_2, \dots, S_k\}$ so as to minimize the within-cluster sum of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

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = \arg \min_{\mathbf{S}} \sum_{i=1}^k |S_i| \text{Var } S_i$$

Example:



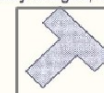
Form
feature
vectors



Cluster
Seekingg

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

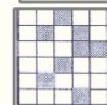
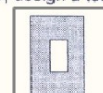


Fig1(a),(b)

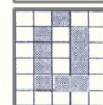


Fig2(a),(b)

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

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).

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

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) Assume Number of Classes = k.

Pick $\vec{m}_i(t) |_{t=1, i=1,2,\dots,k}$.
Arbitrary.

e.g. $\vec{m}_i(t) = \vec{x}_i$

2) $\|\vec{x}_l - \vec{m}_i(t)\| \leq \|\vec{x}_l - \vec{m}_j(t)\|$ for $j \neq i$
 $1 \leq j \leq k$
Regroup \vec{x}_l such that $\vec{x}_l \in \omega_i(S_i(t))$

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

Video Analytics March 15, 2018
HL 1/.

Today's Topics: K-mean for Cluster Analysis.

Background: Computer Vision Pre-processing Tasks.

2D Convolutions.
Kernel Design. LoG

Feature Extraction Techniques.

B(x,y) Binary Image Analysis.
Area (Size), Perimeters,
 \bar{x}, \bar{y} , Moments. m_{ij}

Feature Vectors. $i, j = 0, 1, 2$.

Images from CAT I

Images from CAT II

$I_1(x,y)$ from CAT I.

$I_2(x,y)$ from CAT II.

Decision Function.

K-mean Technique.

Non-Supervised Learning v.s. Supervised Learning.

Better/Best "Cluster" of the Data,

Define Objective function for Optimization.

minimized Error

$$\sum_{i=1}^K \sum_{\vec{v} \in S_i} \|\vec{v} - \vec{u}\|^2 \quad \dots (1)$$

Takes care of Possible "Cancellations"

\vec{v} : Sample Data Vector
 \vec{u} : Ideal Center "Cluster"
 S_i : CAT I, CAT II
Total Number of Samples Per CAT.
For Total Number of CAT's.

Algorithm: Step 1 Assignment Step.

$$S_i^{(t)} = \{p : \|X_p - m_i^{(t)}\| \leq \|X_p - m_j^{(t)}\|, \forall j, 1 \leq j \leq K\}$$

where $X = \vec{v}$

Step 2

Update:

$$m_i = \frac{\sum_{X_j \in S_i^{(t)}} X_j}{|S_i^{(t)}|} \quad \dots (3)$$

Data Collection

$\{\vec{V}_{I,i} | i=1, 2, \dots, N\}$
"N" Samples (I)



$\{\vec{V}_{I,j} | j=1, 2, \dots, M\}$
M Samples for CAT II

Machine Learning

for $\vec{V}_{I,i}$ ($V_{I1}, V_{I2}, \dots, V_{Ik}$) K-Dimensional Space

V_{I1} : A (size), V_{I2} : Perimeter, V_{I3} : P/A,

V_{I4} : \bar{x} , V_{I5} : \bar{y} , ... (8 Additional moments), Here, $k=13$

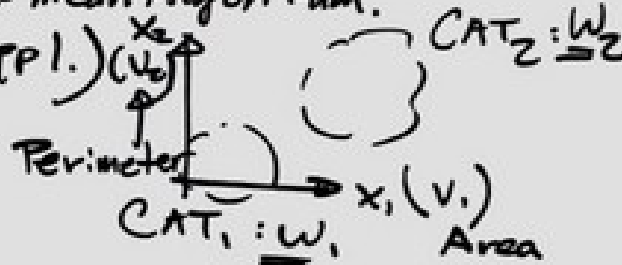
Hand Calculation

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 p1.)
In Class.

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



Step 1: Arbitrary 2 pts as 2 classes
 $m_i(t) \begin{cases} m_1(1) = \vec{X}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ m_2(1) = \vec{X}_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{cases}$ $k=2$ means

Mean.

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

Since,

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

holds good

Similarly

$$\left\| \vec{X}_1 - \vec{m}_2 \right\| = (x_{11} - m_{21})^2 + (x_{12} - m_{22})^2 = (0 - 1)^2 + (0 - 0)^2 = 1 \rightarrow \vec{X}_1 \in W_1 \text{ (or } S_1(t+1))$$

And \vec{X}_2 , we have

$$\left\| \vec{X}_2 - \vec{m}_2 \right\| \leq \left\| \vec{X}_2 - \vec{m}_1 \right\|$$

$$\vec{X}_2 \in W_2 \text{ (or } S_2(t+1))$$

for \vec{X}_3 ,

$$\left\| \vec{X}_3 - \vec{m}_1 \right\| \leq \left\| \vec{X}_3 - \vec{m}_2 \right\|$$

$$\vec{X}_3 \in W_1 \text{ (or } S_1(t+1))$$

Step 3. Update means.

$$\vec{m}(t) = \frac{1}{|S(t)|} \sum_{x_i \in S(t)} \vec{X}_i = \frac{1}{2} (\vec{X}_1 + \vec{X}_3)$$

$$= \frac{1}{2} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix}$$

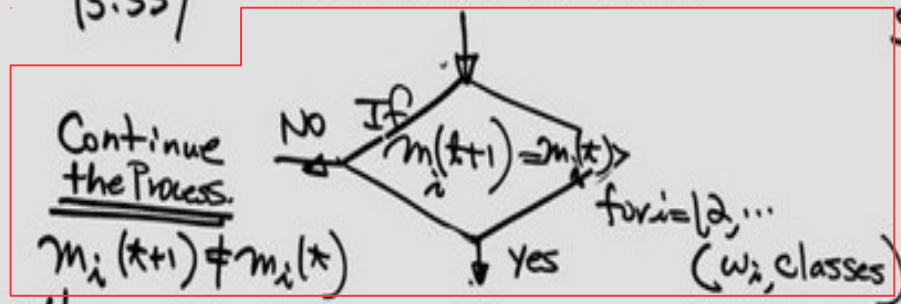
K-mean

CMPE 97 March 5, 2018, Harry Li 3/

$$\vec{m}_2(2) = \frac{1}{|S_2(1)|} \sum_{j \in S_2(1)} \vec{x}_j = \frac{1}{18} (\vec{x}_2 + \vec{x}_4 + \dots + \vec{x}_{20})$$

$$= \begin{pmatrix} 5.67 \\ 5.33 \end{pmatrix}$$

Terminate the Process?



Step 4

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

and

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

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

Update mean,

$$\vec{m}_1(3) = \frac{1}{10} (\vec{x}_1 + \vec{x}_2 + \dots + \vec{x}_8) = \begin{pmatrix} 1.25 \\ 1.33 \end{pmatrix}$$

$$\vec{m}_2(3) = \frac{1}{10} (\vec{x}_9 + \vec{x}_{10} + \dots + \vec{x}_{20}) = \begin{pmatrix} 7.67 \\ 7.33 \end{pmatrix}$$

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

Since, not, continue,

Step 5. Check

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

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

$$\text{Similarly, } \|\vec{x}_p - \vec{m}_2(3)\| \stackrel{?}{\leq} \|\vec{x}_p - \vec{m}_1(3)\|$$

Both is Satisfied, No Regrouping, therefore,

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

Yes! Converged, Stop.

Use kmean.py To Compute Example

AI & Machine Learning.
March 17, 2018 Harry Li 1/

Machine Learning \leftarrow Data Analytics
Decision Making Functions

Background: Cluster Seeking \rightarrow Data Analytics.
Algorithm. Flood Fill (Selection) Based ON "Size"

VideoCAM I(x,y) B(x,y) \rightarrow Capture \rightarrow 2D Convolution \rightarrow Gaussian Blur \rightarrow Log \rightarrow Binarization \rightarrow preprocessing \rightarrow Feature Vectors.

Moments:
$$m_{ij} = \frac{\iint_{\Omega} (x-\bar{x})^i (y-\bar{y})^j B(x,y) dx dy}{\iint_{\Omega} B(x,y) dx dy} \dots (1)$$

Perimeters. \rightarrow Class I w_1 \rightarrow Class II w_2 \rightarrow Area \rightarrow Feature Selection \rightarrow Feature Vectors.

Cluster Seeking. \rightarrow Mathematical Formulation. Objective Function.

github / hualili / opencv / deep learning

"Error" + "minimize" \rightarrow "Error" function \rightarrow Algorithm.

$\sum \sum \| \vec{V} - \vec{U} \|^2$ (4) Square Error Sum. to void possible "cancellations".

① Error of my data pts. Ideal center "cluster".
② All data pts from that "class" $\rightarrow \vec{V} \in S_i, i=1,2,\dots,k$
③ All classes $\rightarrow i=1,2,\dots,k$

Here,
$$\arg \min_S \sum_{i=1}^K \sum_{\vec{V} \in S_i} \| \vec{V} - \vec{m}_i \|^2 \dots (1)$$

$$V_1 = m_{11}$$

$$V_2 = m_{10}$$

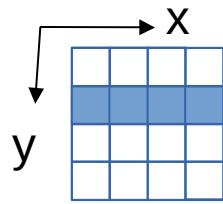
$$V_3 = m_{01}$$

Example: $w_1: B_1(x,y), \dots, B_6(x,y)$, $w_2: B'_1(x,y), \dots, B'_{10}(x,y)$, $V = (v_1, v_2, v_3)$, where $v_1 = m_{11}$, $v_2 = m_{10}$, $v_3 = m_{01}$, find m_1 and m_2 by using kmean.py

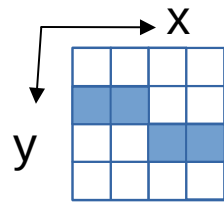
kmean.py Example



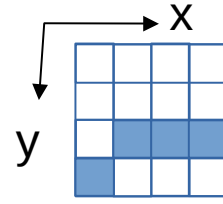
w1



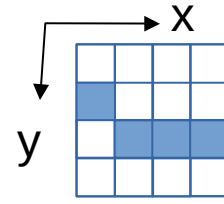
$B1(x,y)$



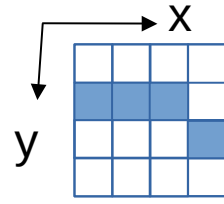
$B2(x,y)$



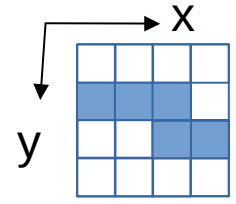
$B3(x,y)$



$B4(x,y)$



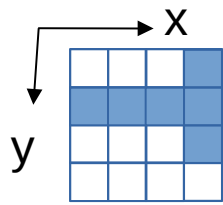
$B5(x,y)$



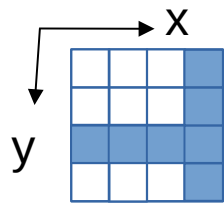
$B6(x,y)$



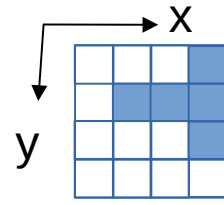
w2



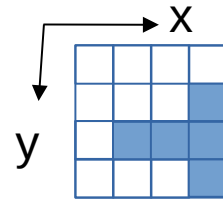
$B'1(x,y)$



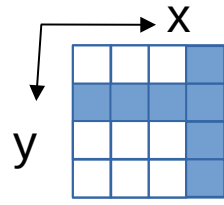
$B'2(x,y)$



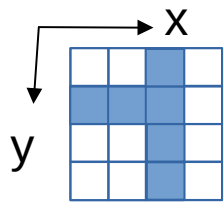
$B'3(x,y)$



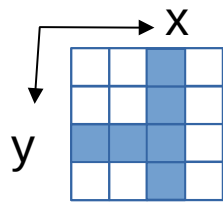
$B'4(x,y)$



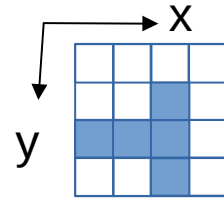
$B'5(x,y)$



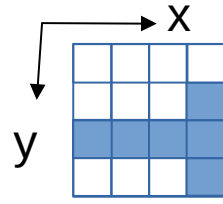
$B'6(x,y)$



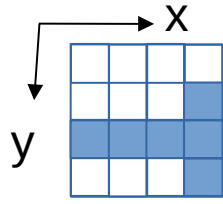
$B'7(x,y)$



$B'8(x,y)$



$B'9(x,y)$



$B'10(x,y)$

Midterm Review

