

# CS 360: Machine Learning

Prof. Sara Mathieson

Fall 2019



## Announcements

- Lab 6 graded
- Project meetings in lab this week  
(required) partners  
try to come to same lab
- Office Hours today  
12:30-1:30pm

# Outline for December 3

- K-means
- Handout 21
- Gaussian Mixture Models (GMM)

# Outline for December 3

- K-means
- Handout 21
- Gaussian Mixture Models (GMM)



# Clustering Goals

- \* learn about underlying structure in the data
- \* cluster new data

$$C_k = \{ \vec{x}_7, \vec{x}_{10}, \vec{x}_{18} \}$$

$$\vec{\mu}_k = \frac{\vec{x}_7 + \vec{x}_{10} + \vec{x}_{18}}{3}$$

Goal: minimize WCSS

(within cluster sum of squares)

find  $\{C_1, C_2, \dots, C_K\} = C$

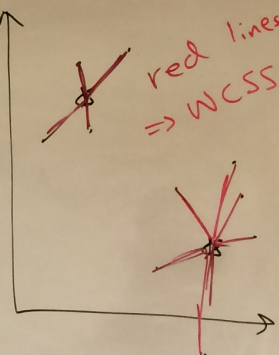
s.t.

$$\sum_{k=1}^K \sum_{\vec{x}_i \in C_k} \|\vec{x}_i - \vec{\mu}_k\|^2 = J(C) = \text{WCSS}$$

is minimized

mean of cluster  $k$

say  $K$  is fixed



NP-hard: try out all cluster combos.

# K-means (given K)

## ① Initialization

choose means for K clusters

$$\vec{\mu}_1^{(1)}, \vec{\mu}_2^{(1)} \dots \vec{\mu}_K^{(1)}$$

iterate: say iter t

① E-step (Assignment) for each training example  $\vec{x}$ , find closest mean  $\vec{\mu}_k^{(t)}$  & assign it label k.  $\Rightarrow \vec{x} \in \mathcal{C}_k^{(t)}$

b/c features  
p

$$\vec{\mu}_k^{(1)} \in \mathbb{R}^p$$

choose among training data

randomly w/o replacement

## ② M-step Update params

maximization

$$\mu_k^{(t+1)} = \frac{1}{|\mathcal{C}_k^{(t)}|} \sum_{\vec{x}_i \in \mathcal{C}_k^{(t)}} \vec{x}_i$$

size of cluster  $\mathcal{C}_k^{(t)}$

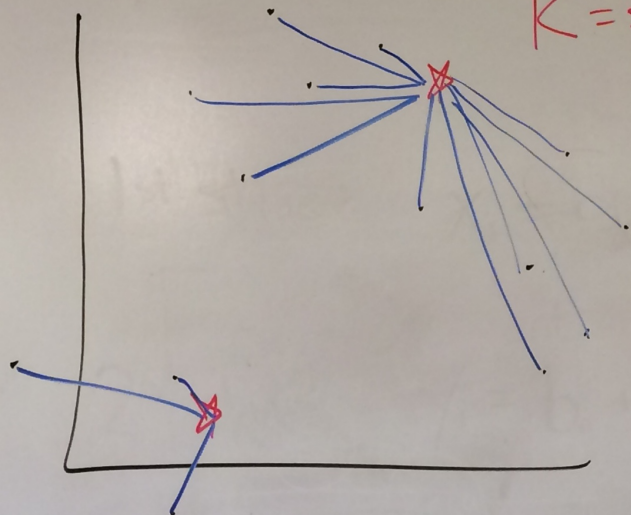
Stop: \* no cluster membership changes

\* max iterations exceeded

\* seen a configuration you've seen before (cycle)

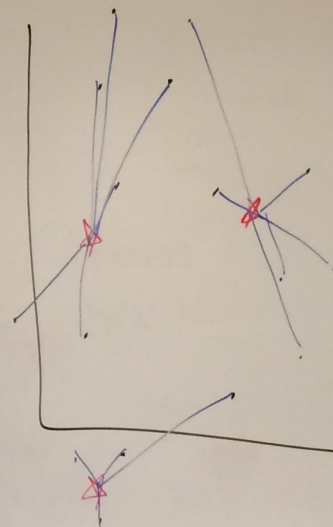
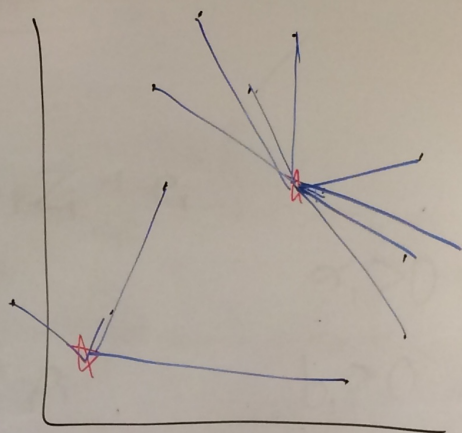


$K=3$

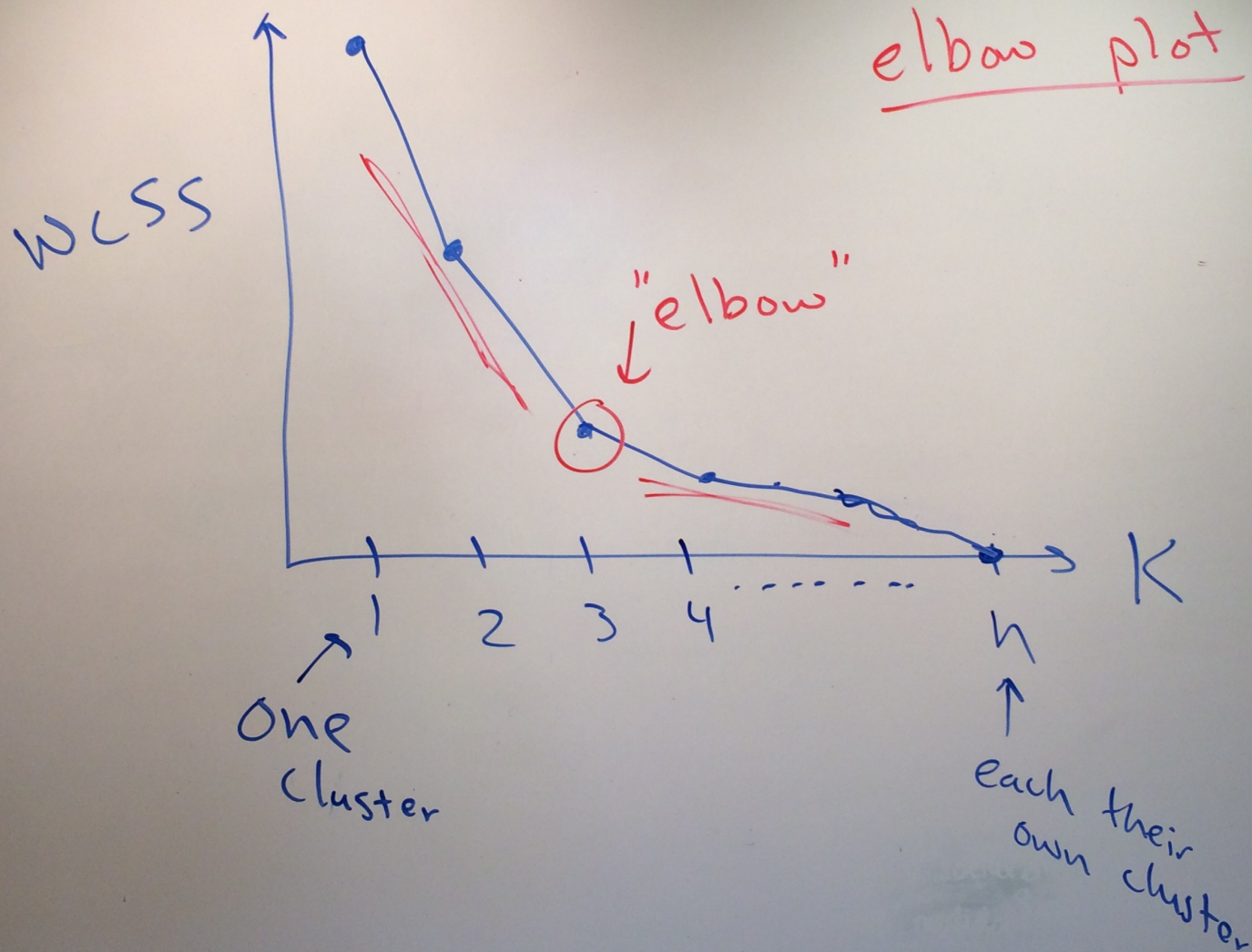


E-step

M-step



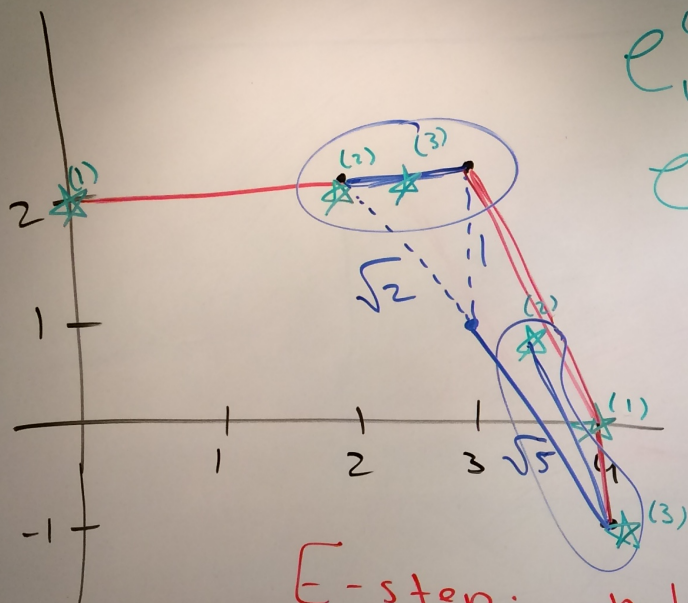
# How to choose K?



# Outline for December 3

- K-means
- Handout 21
- Gaussian Mixture Models (GMM)





$$C_1^{(1)} = \{\bar{x}_2\}$$

$$C_2^{(1)} = \{\bar{x}_1, \bar{x}_3\}$$

$$\mu_1^{(2)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\mu_2^{(2)} = \begin{bmatrix} 3.5 \\ 0.5 \end{bmatrix}$$

E-step:

$n \cdot K \cdot p$

M-step:

$np$

$O(nKpT)$

$$K=1: \quad \vec{\mu} = \begin{bmatrix} (3+2+4)/3 \\ (2+2+1)/3 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

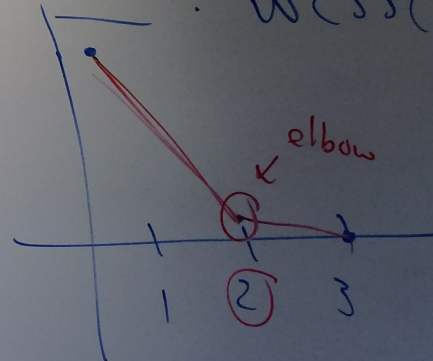
$$WCSS(1) = 8$$

K=2:

$$WCSS(2) = \frac{1}{2}$$

K=3:

$$WCSS(3) = 0$$



# Outline for December 3

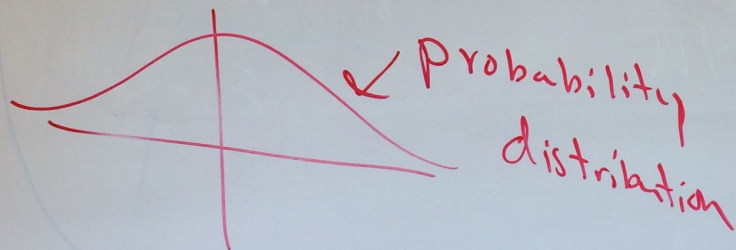
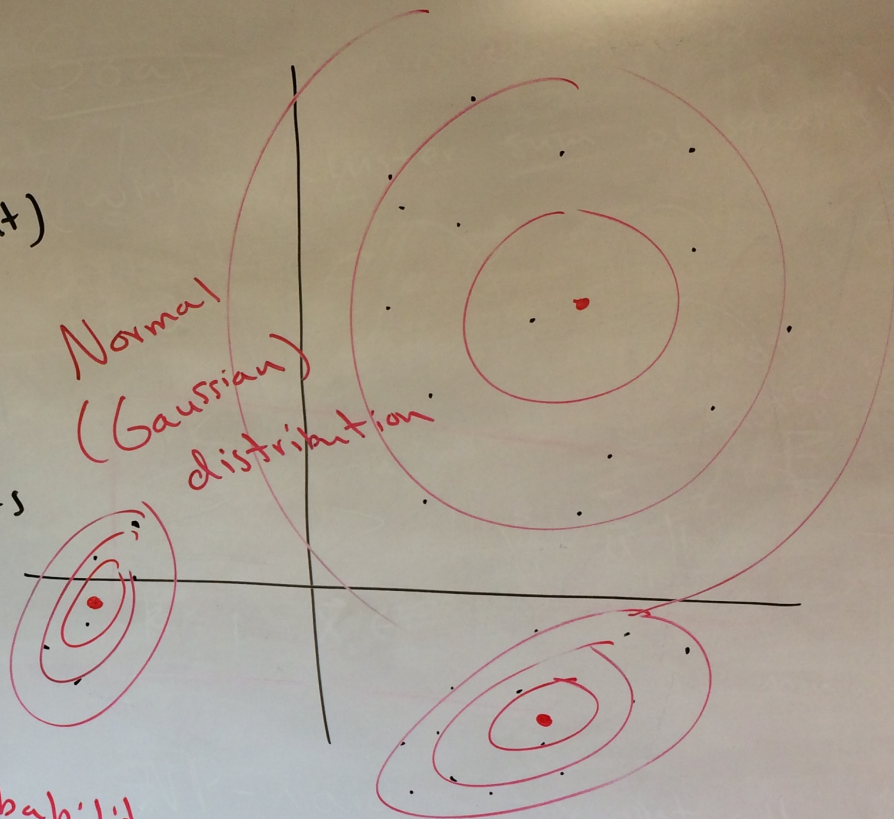
- K-means
- Handout 21
- Gaussian Mixture Models (GMM)



# Problems with K-means

\* not generative (could not create a new data point)

\* does not account for different cluster sizes and variances.





# Gaussian Mixture Models (GMM)

Likelihood!

$$p(\vec{x}) = \sum_{k=1}^K p(\vec{x}, z=k) = \sum_{k=1}^K \pi_k p(\vec{x} | z=k)$$

all data  $\rightarrow$

cluster membership  $\rightarrow$

prior prob. of class  $k$   $\rightarrow$  cluster size.

assume Normal distribution.

Goal!

$$L(X) = \prod_{i=1}^n \sum_{k=1}^K \pi_k N(\vec{x}_i; \vec{\mu}_k, \sigma_k^2)$$

mean  $\rightarrow$

variance  $\rightarrow$

BAYES  $\rightarrow$

want to maximize!

wrt  $\pi$ 's,  $\mu$ 's,  $\sigma$ 's

- $\pi_k$  = prob of class  $k$
- $\vec{\mu}_k$  = mean of cluster  $k$
- $\sigma_k^2$  = variance of cluster  $k$

Initialization

Same as K-means

$$\pi_k = \frac{1}{K} \quad \forall k$$

E-step

"soft" assignment

$w_{ik}$  = prob that  $\vec{x}_i$  came from cluster  $k$

$$w_{ik} = \underbrace{p(k|\vec{x}_i)}_{\text{posterior}} = \frac{\underbrace{p(k)}_{\text{prior}} \underbrace{p(\vec{x}_i|k)}_{\text{likelihood}}}{\underbrace{p(\vec{x}_i)}_{\text{evidence}}}$$

BAYES

$$w_{ik} = \frac{\pi_k N(\vec{x}_i; \vec{\mu}_k, \sigma_k^2)}{\sum_{k'=1}^K \pi_{k'} N(\vec{x}_i; \vec{\mu}_{k'}, \sigma_{k'}^2)}$$

$W =$

$$\begin{matrix} & \overbrace{\hspace{2cm}}^{K=3} \\ \begin{bmatrix} 0.5 & 0.2 & 0.3 \end{bmatrix} & \vec{x}_i \\ & n \times K \end{matrix}$$