## CIS 635 - Knowledge Discovery & Data Mining

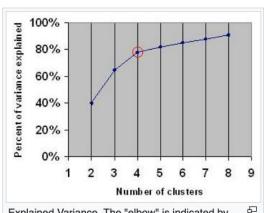
**Optimal Number of Clusters** 

## **Number of clusters**

- Silhouette score based
- Elbow method

## **Elbow method**

 Method can be used for number of clusters selections, and also for some other applications such as number of PCA components etc.



Explained Variance. The "elbow" is indicated by the red circle. The number of clusters chosen should therefore be 4.

## Silhouette score based

For each sample:

a: mean intra-cluster distance

b: mean nearest-cluster distance

Silhouette Coefficient:

$$sc = \frac{(b-a)}{max(a,b)}$$



QA