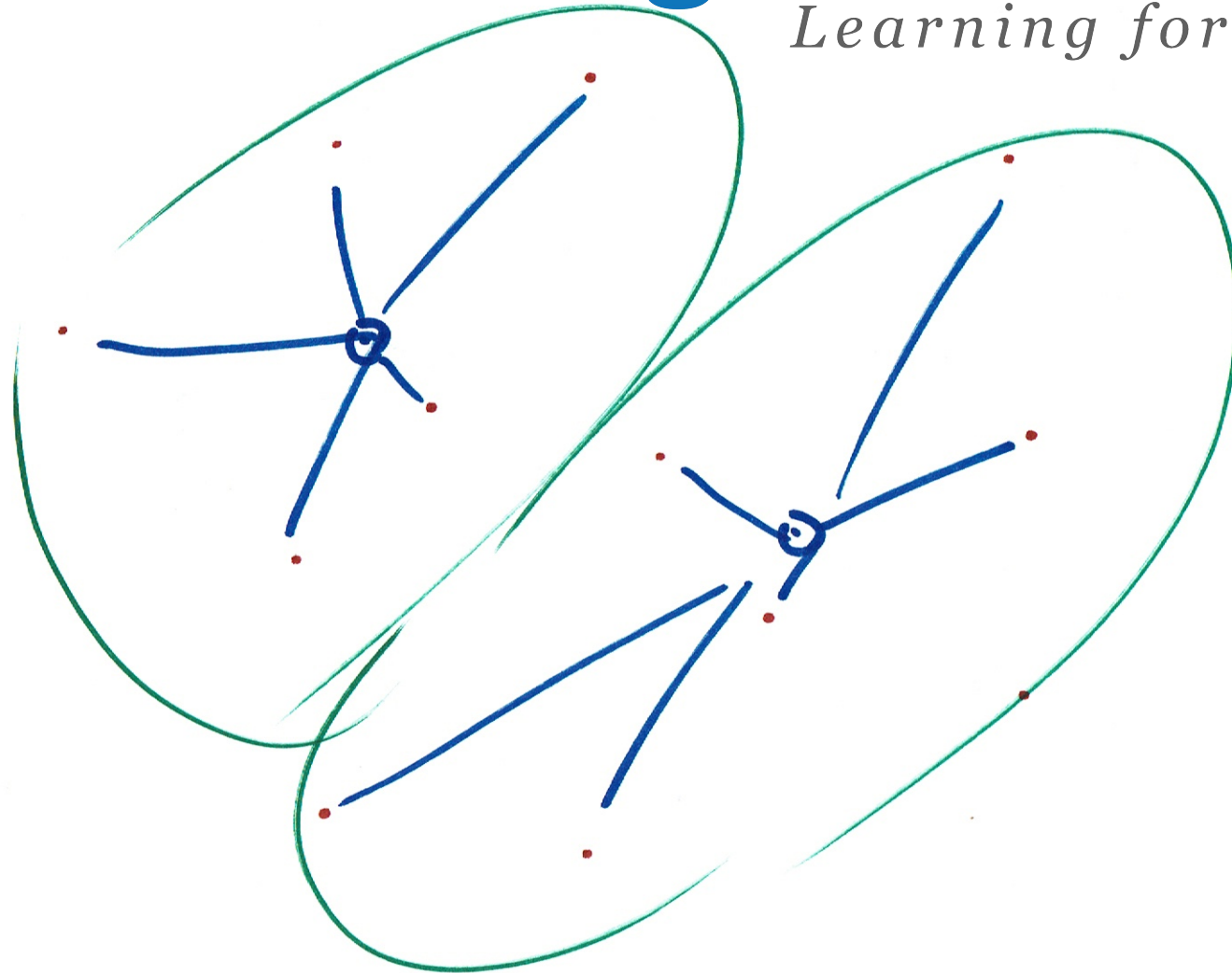


K

$$\frac{n(n+1)}{2}$$



# Centroid based: K-Means Clustering



- K-Means is probably the most used clustering technique
- Aims to partition the  $n$  observations into  $k$  clusters so as to minimize the within-cluster sum of squares (i.e. variance). }
- Computationally less expensive compared to hierarchical techniques.
- Have to pre-define  $K$ , the no of clusters

$$\frac{\sum (\bar{x} - x_i)^2}{n}$$

$K=3$

greatlearning  
*Learning for Life*

