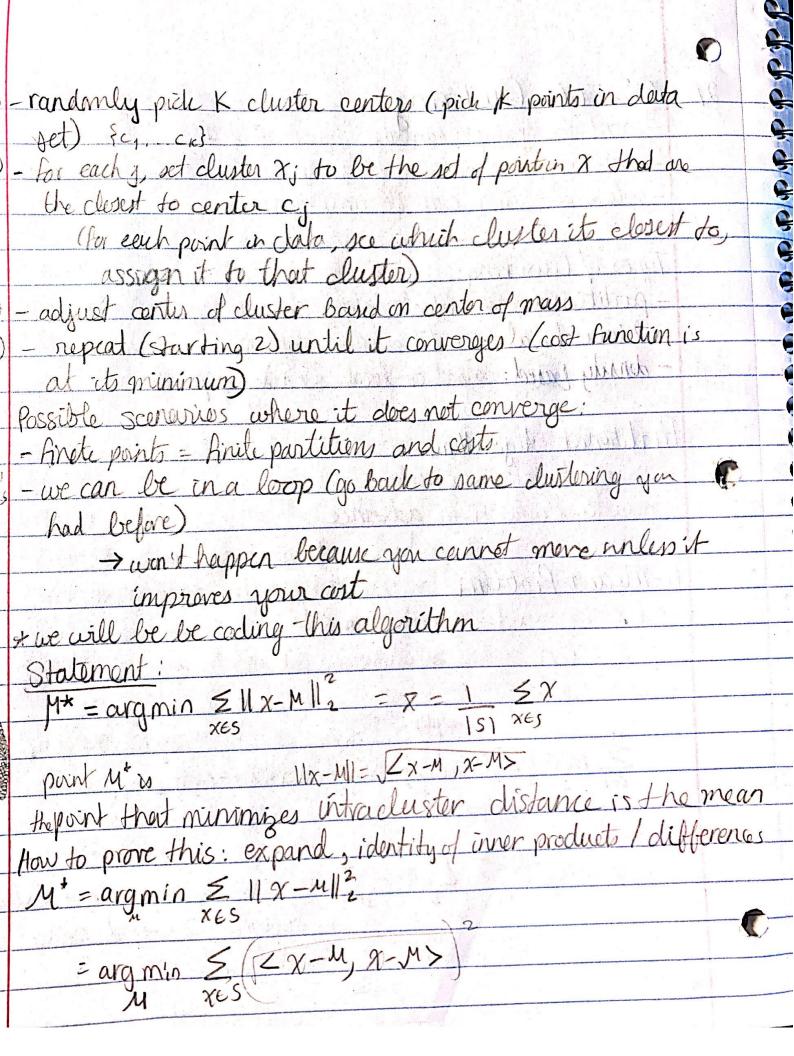
ļ	
	CS506 : Clustering
`	- want to group similar objects together
	- dissimilar objects in different clusters
	- notion of cluster can be ambiguous
1	The first the second of the se
	Types of Clustering:
	- partitional: each object belongs in exactly 1 cluster
	- hierarchical: nested clusters organized in a tree
	- density based: based on local density of points
net repr	Partitional Algorithm
	- n object into K cluster
	- need to know K in advance
7	
	K-Meens Problem
	-set x = Ex, xn3 dn points in Ra
	- Kis given
	- Cost function: center center n center n center n
	$\frac{1}{n}$ center $\frac{1}{n}$ vector $x-c_i$
	$\frac{\sum_{i=1}^{n} \min \left\{ L_{2}^{2}(x_{i}, c_{i}) \right\}}{\sum_{i=1}^{n} \min \left\ x_{i} - c_{j} \right\ ^{2}} = \sum_{i=1}^{n} \min \left\ x_{i} - c_{j} \right\ ^{2}$
,	Lydistance
	Goal 2 1 (2)
	minimize \$\leq \(\(\infty \) \(\tau \) \(
_	all points x in cluster duster
	cluster cj Caus 4



1=n=argmin n<M, u>-2 \(\alpha \) \(\chi \) \(\chi \) Hwhy did we get = argmun $n < u, u > -2n/t \ge x, u >$ = argmin 2 4,4>-22x,4> = argmin 24, 4>-2 < x, 4> + < x, x) = arg min < \(\overline{\chi} - \mu > dadded I, den't recel to subtract it because it's = $\frac{\text{argmin}}{\mu} \frac{11}{2} - \frac{11}{2}$ constant w/ respect to M *works in dimensions higher than 2, but inefficient Properties of K-means algorithm
- finds a local optimum - often converges quickly (brd not always)
- choice of initial points can have longe influence in Limitetions of K-Means - non-spherical shapes won't work + only way to determine besk # of clusters is to