

























































<section-header><equation-block><text><text><list-item><list-item><list-item>

Improving a suboptimal configuration...

Distortion =
$$\sum_{i=1}^{R} (\mathbf{x}_i - \mathbf{c}_{\text{ENCODE}(\mathbf{x}_i)})^2$$

What properties can be changed for centers c_1 , c_2 , ..., c_k when distortion is not minimized?

(1) Change encoding so that x_i is encoded by its nearest center

(2) Set each Center to the centroid of points it owns.

There's no point applying either operation twice in succession.

But it can be profitable to alternate.

...And that's K-means!

Easy to prove this procedure will terminate in a state at which neither (1) or (2) change the configuration. Why?

Copyright © 2001, Andrew W. Moore

K-means Clustering: Slide 31

















