

HUDK 4050: CORE METHODS IN EDM

K-means Gotchas

- Assumes there are clusters to find - it will find clusters regardless of whether there are any or not
- Does not work on some shapes (Like PB&J need an even spread)
- Need uniform scale (uniform variance) - larger scale will swamp smaller scale
- Doesn't work on categorical data of more than two categories (and the scale may be difficult to interpret)
- Can get stuck on local minima (need to run iterations)
- Too easy

<http://bit.ly/2do0DTP>

