

K-means: Initialization issues

- K-means is **extremely sensitive to cluster center initialization**
- Bad initialization can lead to
 - Poor convergence speed
 - Bad overall clustering
- **Safeguarding measures:**
 - Choose first center as one of the examples, second which is the farthest from the first, third which is the farthest from both, and so on.
 - **Try multiple initializations** and choose the **best result**
 - Other smarter initialization schemes (e.g., look at the **K-means++** algorithm by Arthur and Vassilvitskii)