Explanatory Notes for 6.390

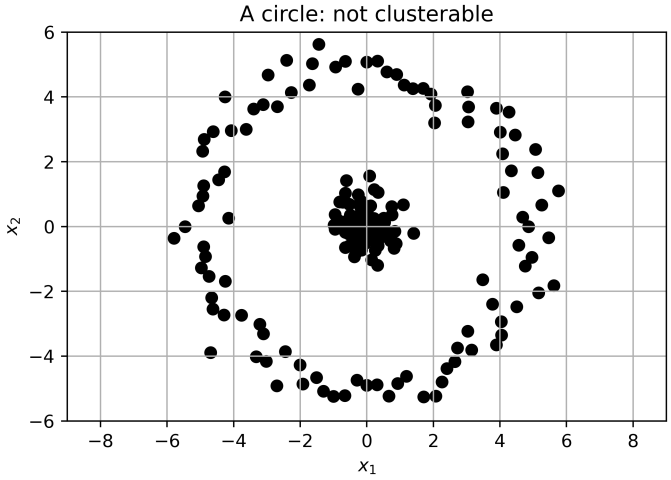
Shaunticlair Ruiz (Current TA)

Fall 2022

Weaknesses of k-means

There are some **weaknesses** to k-means clustering. Certain patterns that we can **see** aren't easily **clustered**.

We can see this with a few **examples**:



This data can't be simply clustered.

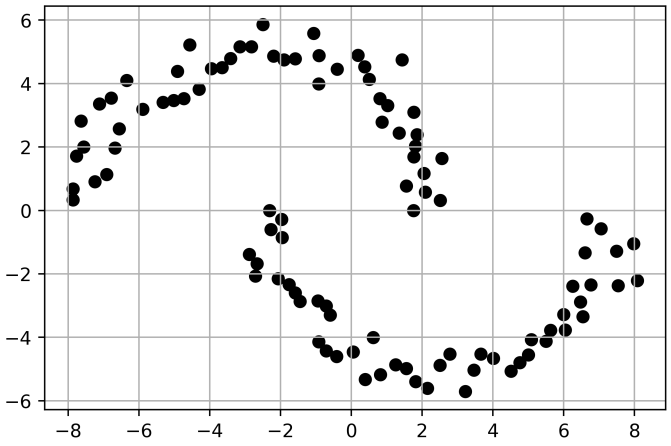
This example can't be effectively **clustered**: most people would agree that the "outer ring" should be **one** cluster, while the "inner circle" should be **another**.

But, assuming we (correctly) place one cluster mean in the **center**, there's **nowhere** we can put our other cluster mean to be **closest** to all of the **outer** points, but **not** the inner points.

We might be able to **resolve** this using a **feature** transformation. But, the problem remains.

For example, we could have a feature represent the radius! But then, we would still struggle with a ring not centered on the origin.

Another example works for clusters that aren't very centralized:



This data can't be clustered either!

The edge of one cluster is too close to the other: we can't easily create a good pair of cluster means for each semi-circle.