

1. As beta approaches infinity, the distribution tightens and the variance decreases. This means a higher weight is put on data that is closer to the means. As beta increases, the clustering becomes harder and harder. When beta is infinity, it is essentially regular K-means.
2. GMM is better at clustering non-spherical data, so a dataset like this with diagonal data would be better modeled using a GMM.
- 3.

