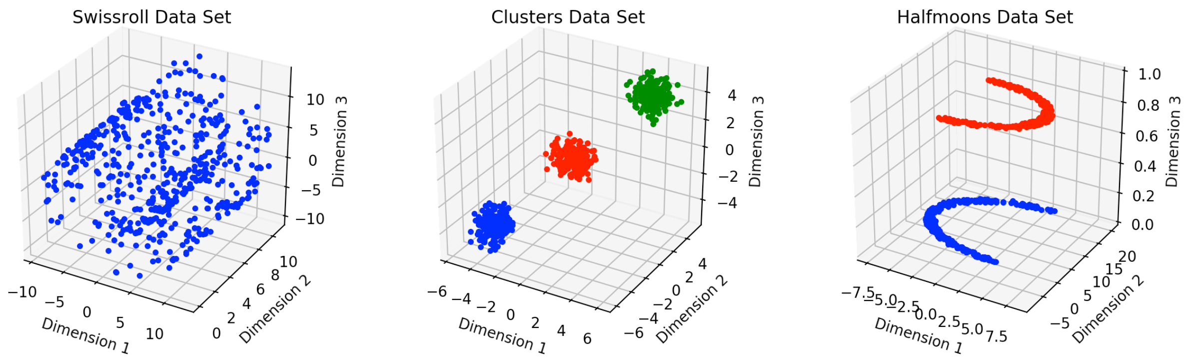
Valentino Christian and Jake Rosenzweig 2018-02-13

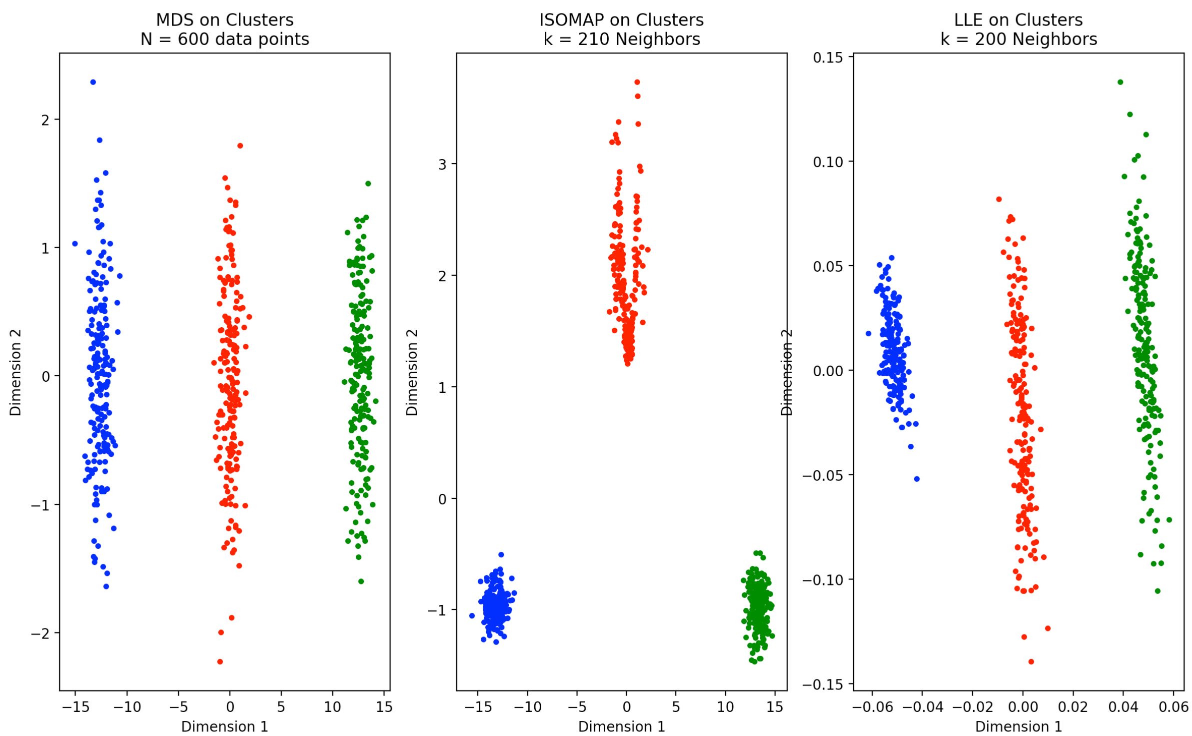
Question: *Are MDS, ISOMAP, and LLE able to identify the intrinsic manifold structure of the data? Why or why not? Be sure to address the goal!*

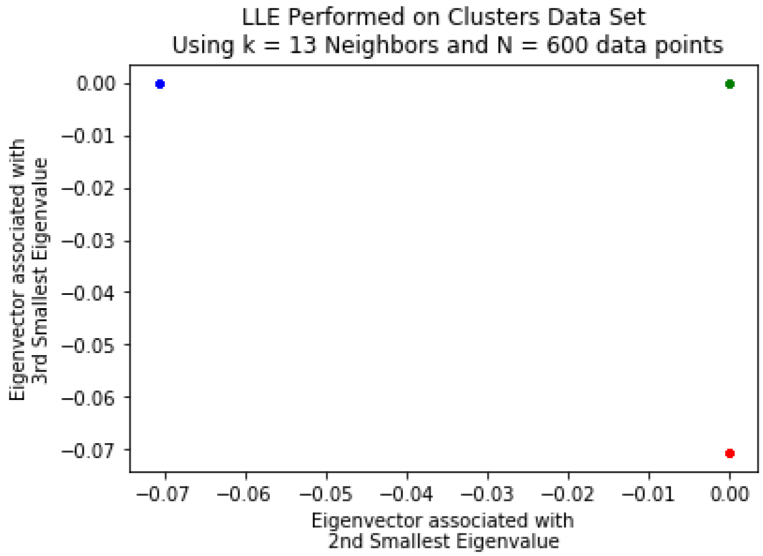


<u>CLUSTERS/3 GAUSSIAN CLOUDS</u>

Description/Goal:

Three groups of data where each group should be classified into separate classes.





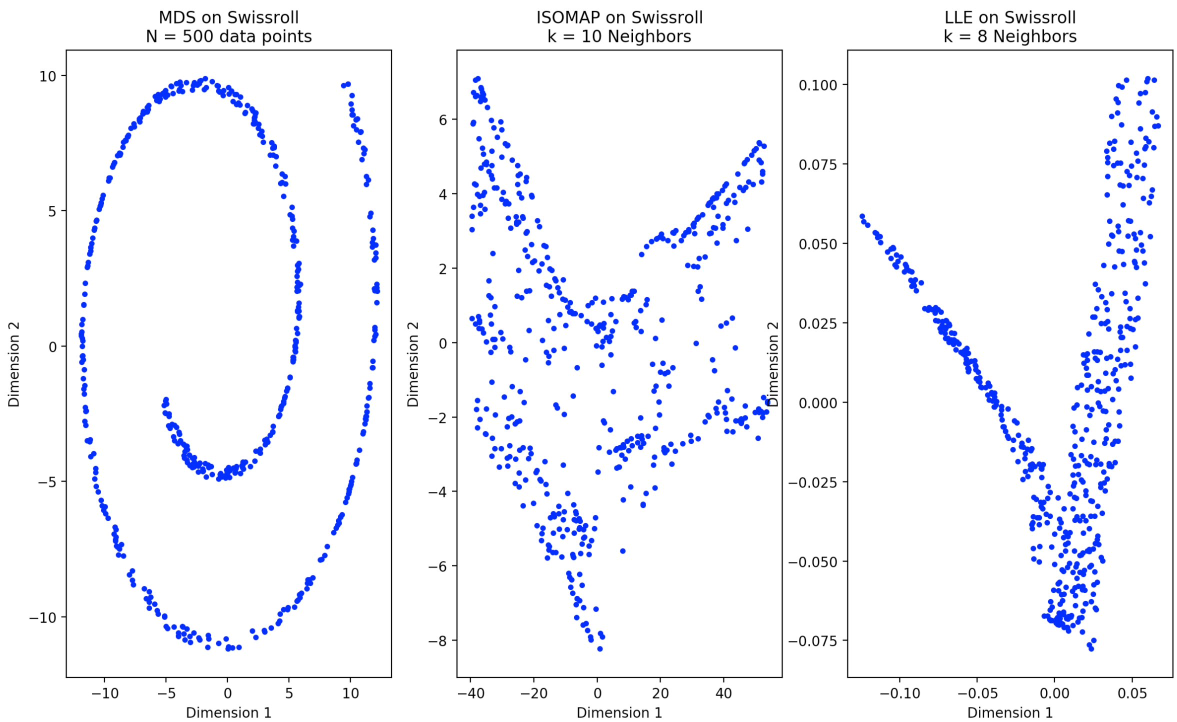
Techniques:

- **MDS:** Did the same job as PCA. Effectively separates the three groups into distinct classes. Doesn't necessarily maintain geodesic distances between neighboring points.
- **ISOMAP:** Reconstructed data points with lower dimensionality have the same proximity matrix which preserves the geodesic distance between neighboring data points compared to the original data. ISOMAP also effectively separates out the three types of data into their respective classes.
- **LLE:** Finally LLE also manages to separate the three classes of data using the locations of neighbors for reconstruction. Here, each cluster contains 200 data points. Interestingly, even if we choose k = 210 neighbors, the algorithm still manages to separate out the three groups. In the last plot above, using only k = 13 neighbors, the algorithm considered each point in its own group to be identical. We get clear classification between groups, but no distinction between data points!

<u>SWISSROLL</u>

Description/Goal:

One group of data where neighboring points on the surface should remain neighbors after dimensionality reduction and points far on the surface should remain far from each other after dimensionality reduction.



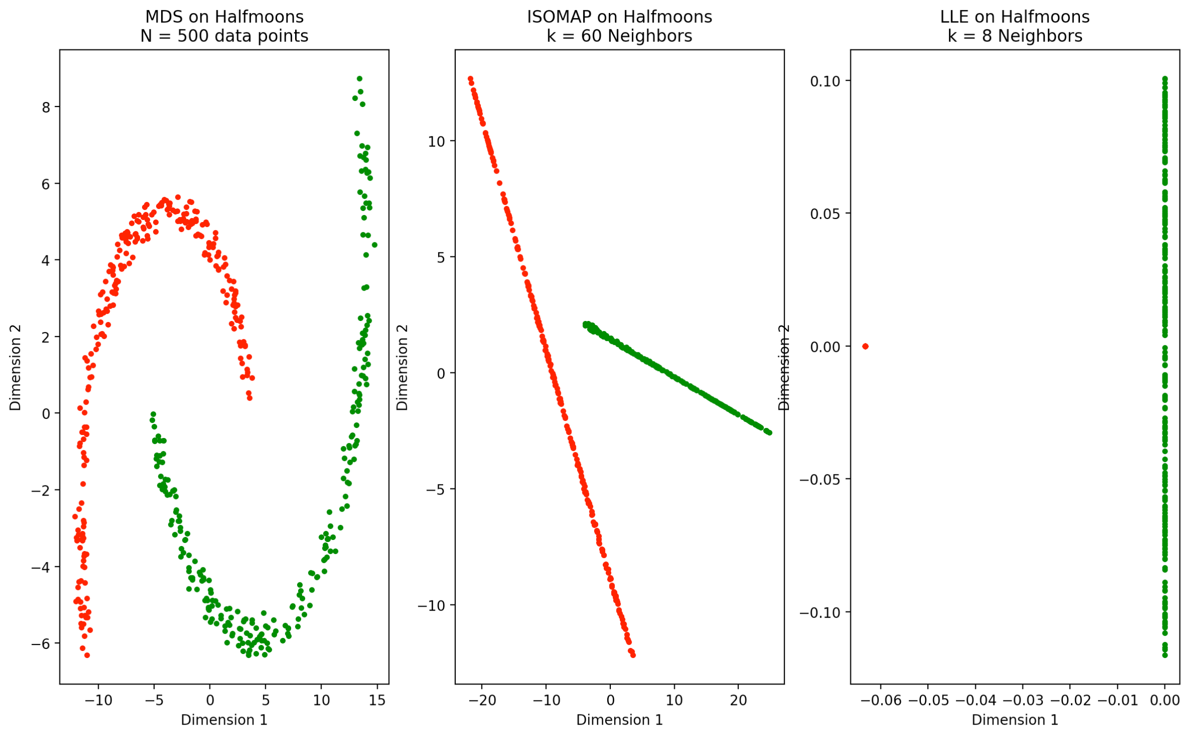
Techniques:

- **MDS:** Did the same job as PCA. Effectively smashes the swissroll into the sidewall. Neighbors before are still neighbors after. Doesn't necessarily maintain geodesic distances between neighboring points.
- **ISOMAP:** This algorithm once again preserves the proximity matrix using Euclidean geodesic distances. Between the three plots above, ISOMAP best shows the unfolding of the swissroll. At the very least puts each of data point to be around its neighbors.
- **LLE:** Here we can see a slightly unfolded and squeezed swissroll (along its side). This plot retains the approximate original location of the data points compared to before LLE, reconstructing each point to be as close to where it originally was.

<u>HALFMOONS</u>

Description/Goal:

Two groups of data where each group should be classified into a separate class.



Techniques:

- **MDS:** Did a similar job that PCA did. Separates the two groups into distinct classes. Doesn’t necessarily maintain geodesic distances between neighboring points. Almost flattens the two halfmoons onto the floor below.
- **ISOMAP:** We see a clear classification between the two types of data. The geodesic distance matrix is retained in the reconstructed data points, using this algorithm.
- **LLE:** Here we are viewing the two halfmoons as if their data points were strung together all in a line. One halfmoon is completely vertical and the other is perpendicular to your screen! All of the red halfmoon’s points fall exactly on that one spot. Therefore, separation between the two different classes is achieved, but I would say that reconstruction of the original data is poor, as the shapes of the halfmoons has been destroyed.