Local Linear Embedding

Katelyn Stringer

ASTR 689

December 1, 2015

Idea Behind LLE

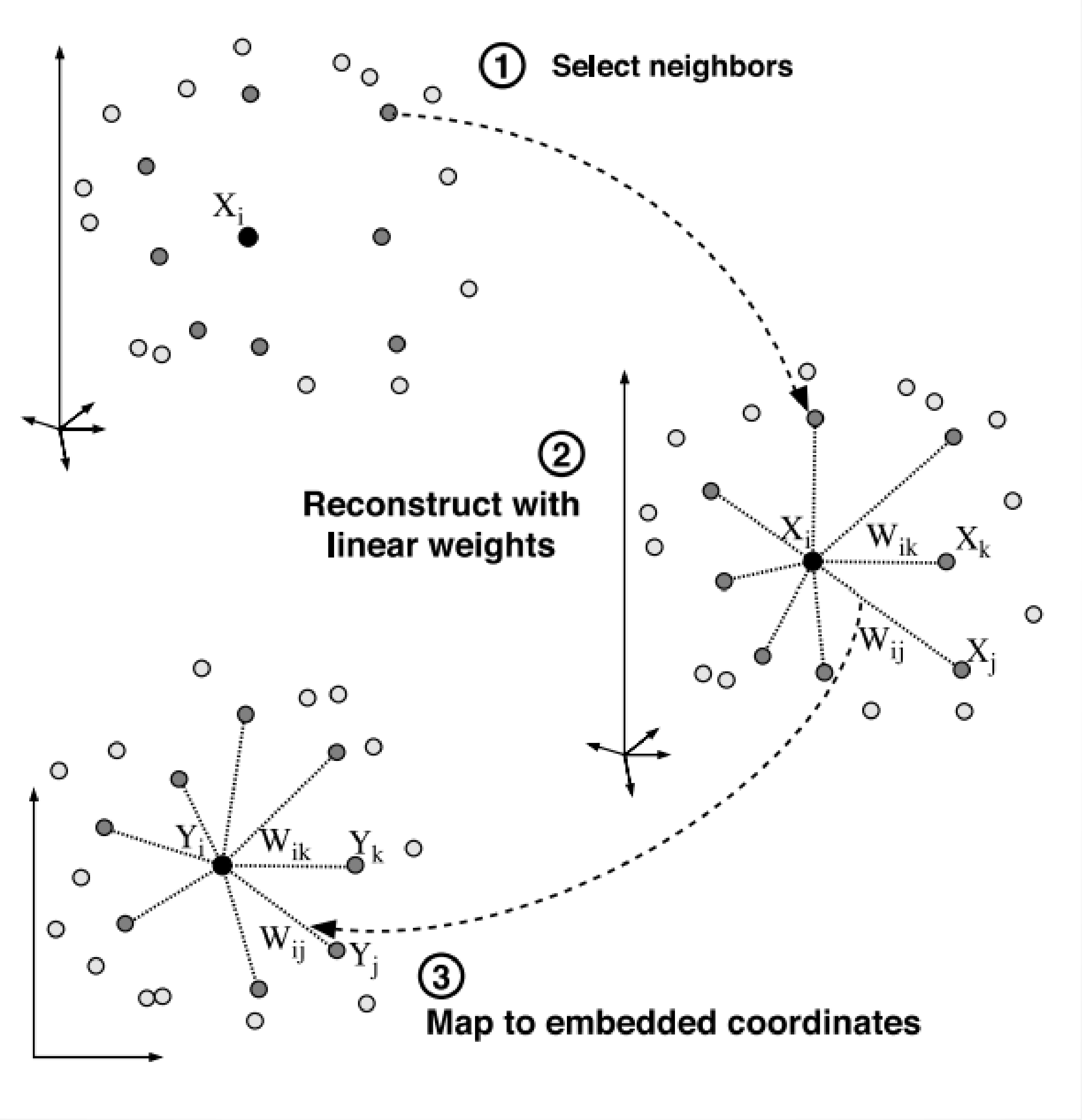
- Good at making nonlinear high-dimensional data easier for computers to analyze
- Example: A high-dimensional surface
- Think derivatives: local tangential hyperplane is a good approximation
- LLE records locations of points on this local tangential manifold based on locations of neighboring points
- It’s like describing your address in terms of how far your house is from other buildings

Idea Behind LLE

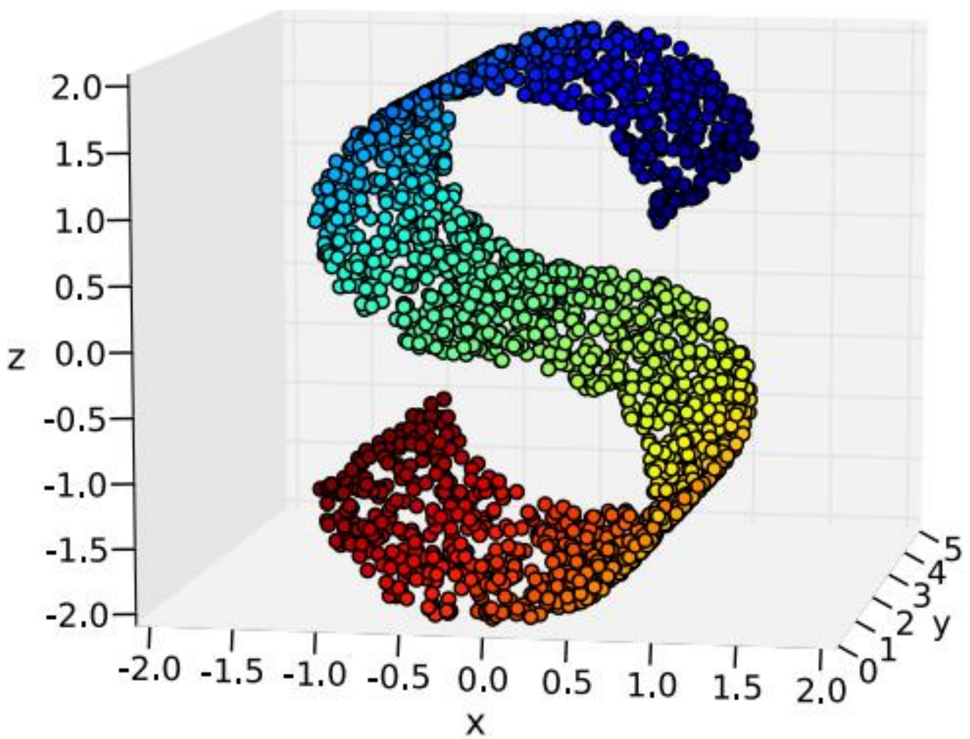
- Good at making nonlinear high-dimensional data easier for computers to analyze
- Example: A high-dimensional surface
- Think derivativ hyperplane is a good approxim

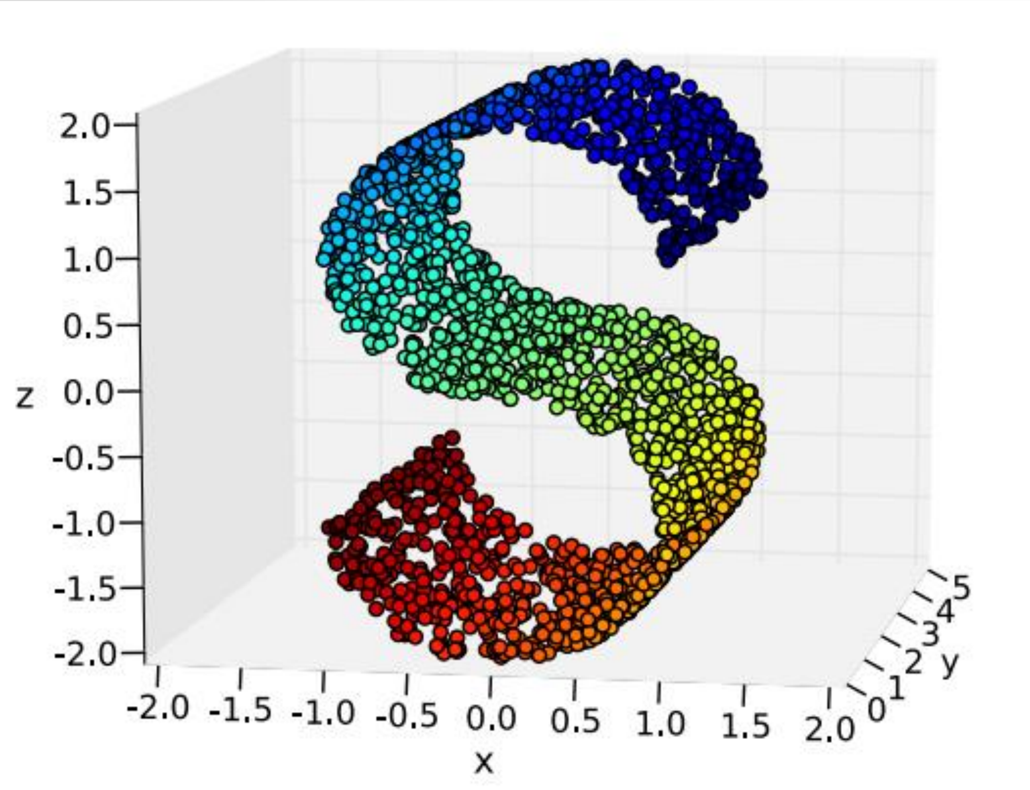
WHY?

- LLE records locations of points on this local tangential manifold based on locations of neighboring points
- It's like describing your address in terms of how far your house is from other buildings



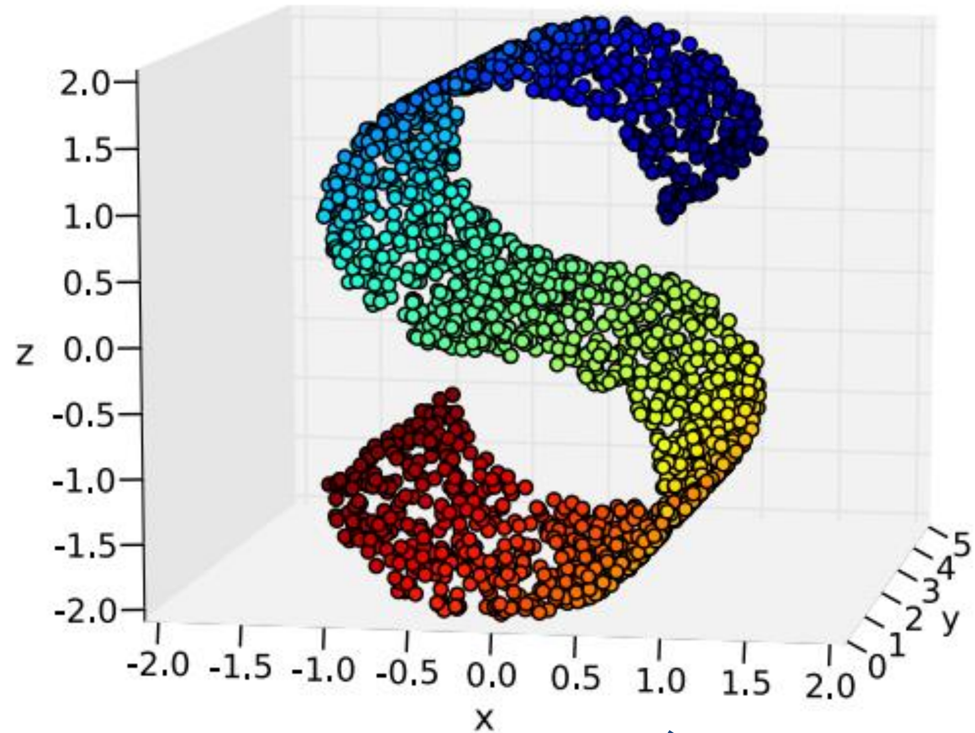
① Select neighbors
X_i
② Reconstruct with linear weights
X_i
W_{ik}
X_k
W_{ij}
X_j
Y_i
W_{ik}
Y_k
W_{ij}
Y_j
③ Map to embedded coordinates


2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
z
-2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0
x
0 1 2 3 4 5
y



2

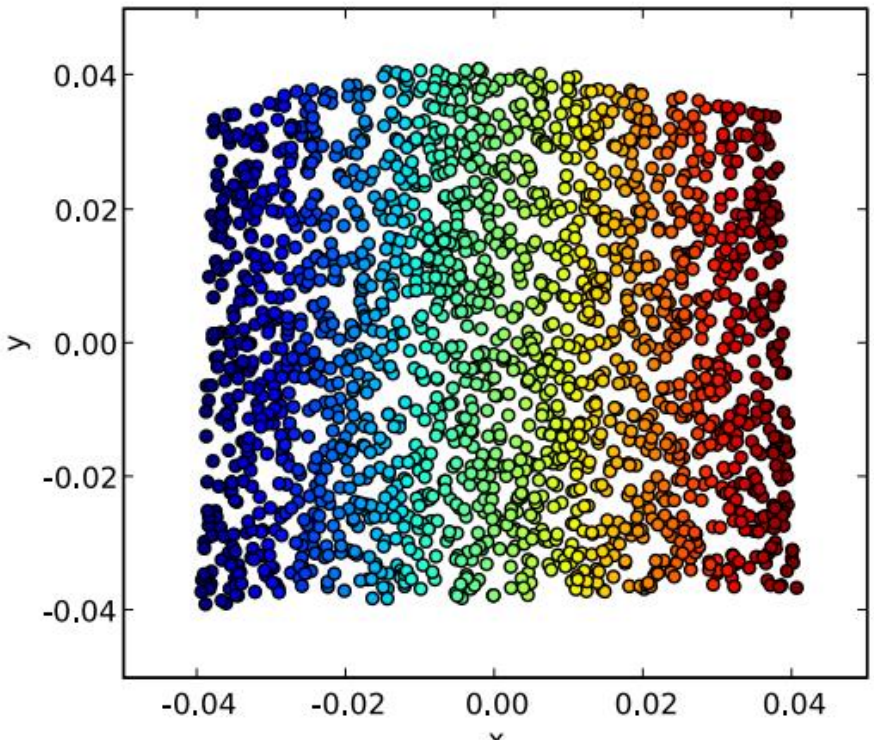
3


z
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0
x
0 1 2 3 4 5
y


LLE!

3

2

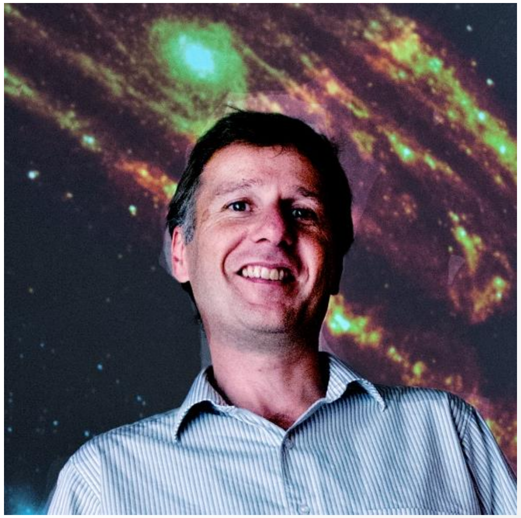

0.04
0.02
y 0.00
-0.02
-0.04
-0.04 -0.02 0.00 0.02 0.04

2

Reducing the Dimensionality of Data: Local Linear Embedding of Sloan Galaxy Spectra

Jake Vanderplas, Andrew Connolly
(University of Washington)

4



5

All of the following images and information came from this paper unless otherwise noted!

The LLE Algorithm

- You have a set of data vectors

$$\mathbf{X} = [\mathbf{x_0}, \mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_N}], \qquad \mathbf{x_i} \in \mathbb{R}^{D_{\text{in}}}$$

- Want to map them to coordinate system

$$\mathbf{Y} = [\mathbf{y_0}, \mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_N}], \qquad \mathbf{y_i} \in \mathbb{R}^{D_{\text{out}}}$$

$$D_{\text{in}} > D_{\text{out}}$$

- For each data vector $\mathbf{x_i}$, the indices of the K nearest neighbors are represented by

$$\mathbf{n}^{(i)} = [n_1^{(i)}, n_2^{(i)}, \ldots, n_K^{(i)}]^T$$

LLE Algorithm, cont.

- Assume that each point $\mathbf{x_i}$ lies near a locally linear low dimensional manifold
- Find K nearest neighbors (based on Euclidean distance)

$$d_{pq} = \sqrt{\sum_{s}^{D_{in}} (X_{ps} - X_{qs})^2}$$

Matrix element

- Find the local covariance matrix

($\mathbf{x}_{n_j^{(i)}}$ is the j^{th} nearest neighbor to $\mathbf{x_i}$)

$$C_{jk}^{(i)} = \left(\mathbf{x_i} - \mathbf{x}_{n_j^{(i)}}\right)^T \left(\mathbf{x_i} - \mathbf{x}_{n_k^{(i)}}\right)$$

- Regularize covariance matrix to produce a stable solution

$$\mathbf{C}^{(i)} = \mathbf{C}^{(i)} + \delta \, \mathrm{tr}(\mathbf{C}^{(i)})\mathbf{I}$$

(authors used $\delta = 10^{-3}$)

LLE Algorithm, cont.

- Determine the optimal weights for each nearest neighbor by minimizing the reconstruction error

$$\mathcal{E}_1^{(i)}(\mathbf{w}^{(i)}) = \left| \mathbf{x_i} - \sum_{j=1}^{K} w_j^{(i)} \mathbf{x}_{n_j^{(i)}} \right|^2$$

- Impose $\sum_{j=1}^{K} w_j^{(i)} = 1$

$$\boxed{C_{jk}^{(i)} = \left(\mathbf{x_i} - \mathbf{x}_{n_j^{(i)}}\right)^T \left(\mathbf{x_i} - \mathbf{x}_{n_k^{(i)}}\right)}$$

- Solve $\mathcal{E}_1^{(i)}(\mathbf{w}^{(i)}) = \sum_{j=1}^{K} \sum_{k=1}^{K} w_j^{(i)} w_k^{(i)} C_{jk}^{(i)} + 2\lambda_i \left(1 - \sum_j w_j^{(i)}\right)$ for eigenvalues
- Or solve $\mathbf{C}^{(i)}\mathbf{w}^{(i)} = [1, 1, 1, ..., 1]^T$, for $\mathbf{w}^{(i)}$ and scale to 1

LLE Algorithm, cont.

- Create overarching $\mathbf{W}$ matrix with all the weight vectors as columns, $W_{ji} = 0$ if point j is not a nearest neighbor of point i
- Create matrix $\mathbf{M} = (\mathbf{I} - \mathbf{W})(\mathbf{I} - \mathbf{W})^T$
- Solve for eigenvalues of $\mathbf{My}^{(i)} - \lambda_i \mathbf{y}^{(i)} = \mathbf{0}$.
- Eigenvectors $\mathbf{y}$ corresponding to $D_{out}+1$ lowest eigenvalues are the new basis vectors

  (except the first, $\lambda=0$ just gives a translation)

LLE Algorithm, cont.

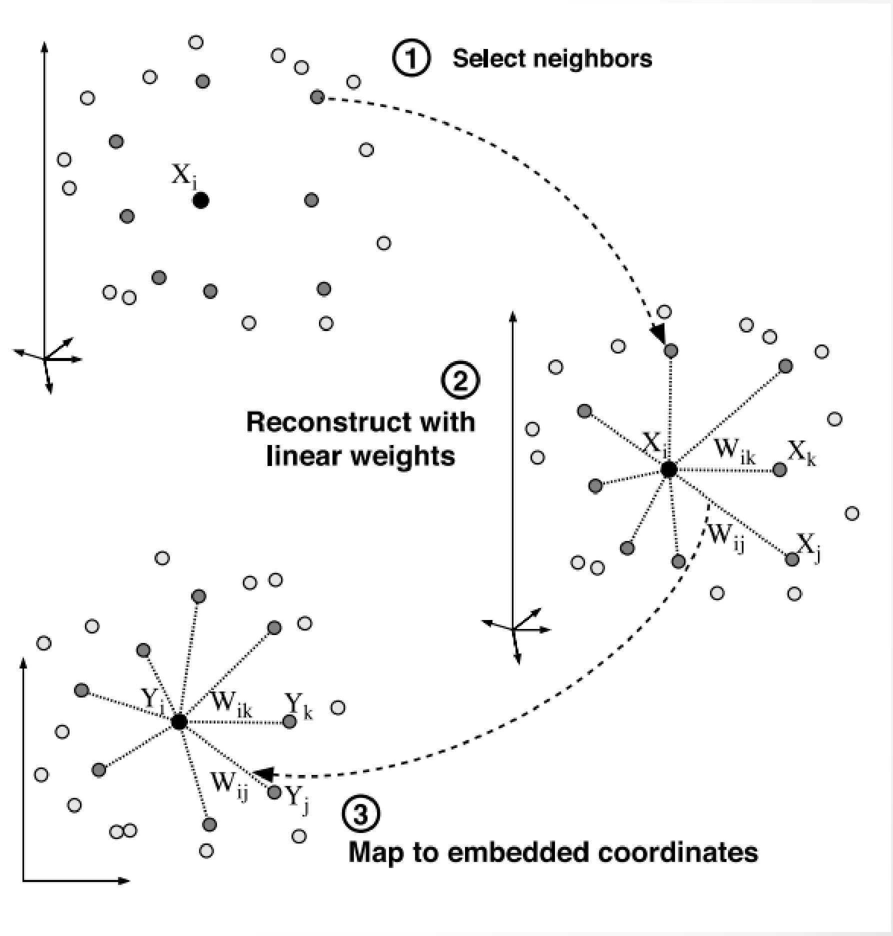
- Errors are given by cost functions:

$$\mathcal{E}_1^{(i)}(\mathbf{w}^{(i)}) = \left| \mathbf{x_i} - \sum_{j=1}^{K} w_j^{(i)} \mathbf{x}_{n_j^{(i)}} \right|^2.$$

$$\mathcal{E}_2(\mathbf{Y}) = \sum_{i=1}^{N} \left| \mathbf{y_i} - \sum_{j=1}^{K} w_j^{(i)} \mathbf{y}_{n_j^{(i)}} \right|$$

LLE Algorithm, Recap

- Find nearest neighbors
- Find weights to map each point in terms of its neighbors
- Find embedded coords
- Use weights to map points in new coords

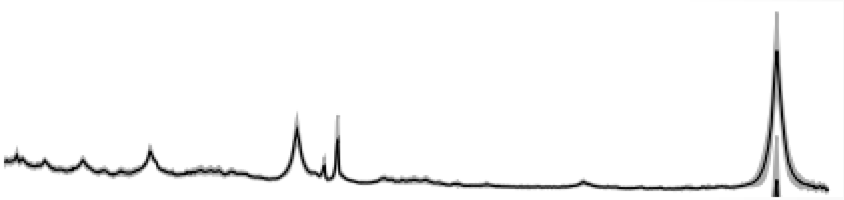


Our Sloan Spectra Sample

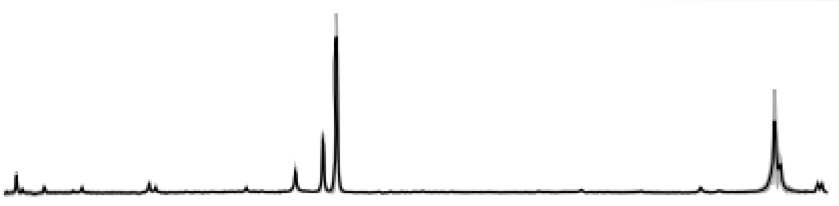
- 8711 total spectra, $z < 0.36$
- 1000 logarithmic wavelength bins (3800-9800 Å)
- Corrected for sky absorption & normalized
- Equivalent widths, line positions in headers
- Prior Classifications:
  - QSOs = Quasi-Stellar Objects , include Quasars
    - Broad line
    - Narrow line
  - Emission Galaxies
  - Quiescent Galaxies
  - Absorption Galaxies

QSOs: What we look for

- Hydrogen emission lines 3x > noise
- Broad line: larger redshift
  - Line widths > 1200 km/s

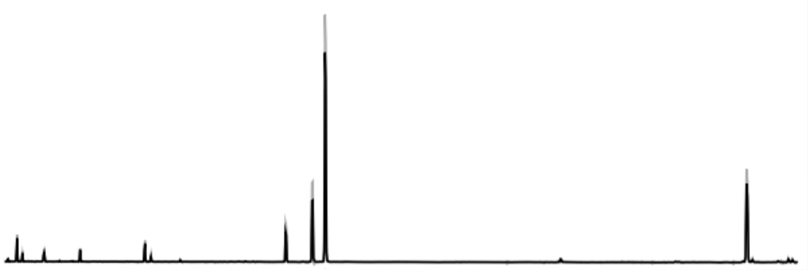


- Narrow line: smaller redshift



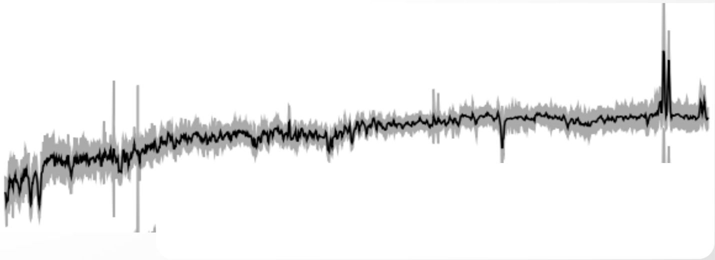
Galaxies

- Emission: Star-forming galaxies
    - Hydrogen emission > 3 x noise

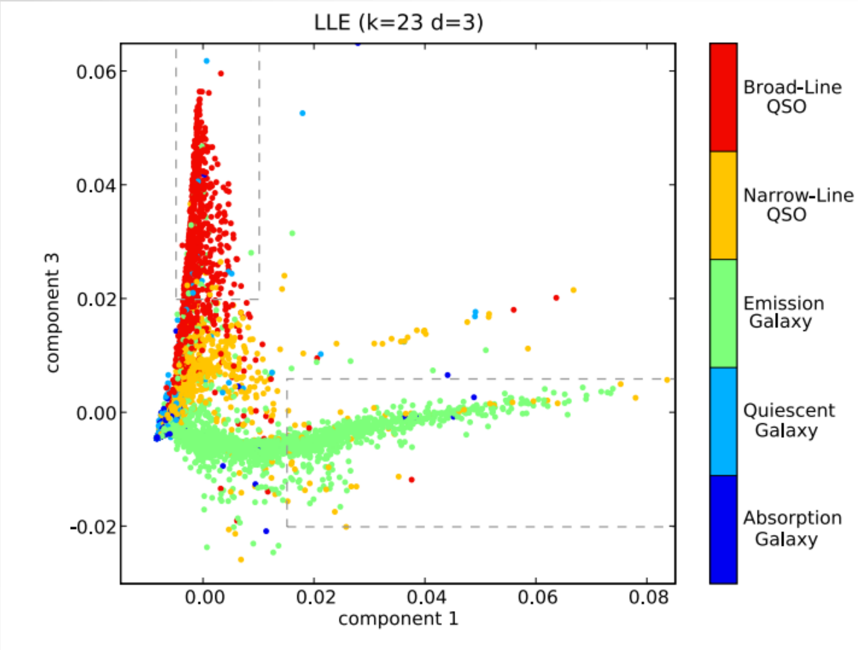


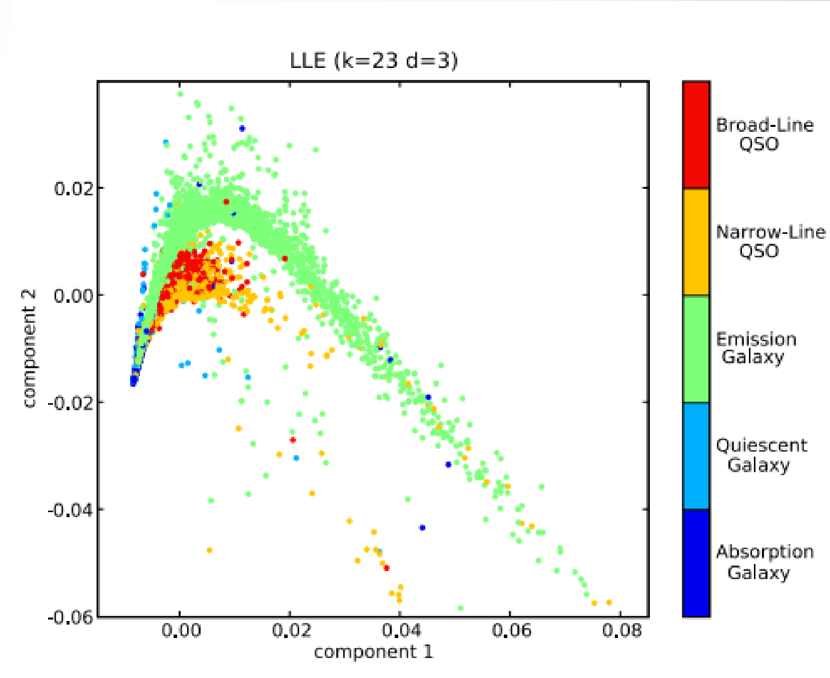
- Absorption: Balmer absorption > 3σ

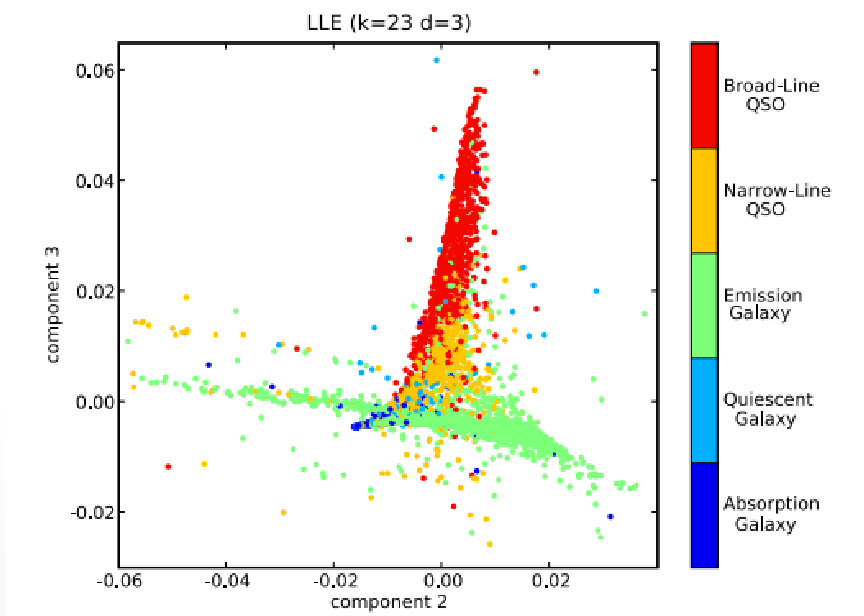
- Quiescent: “red and dead”,
  Balmer emission < 3σ



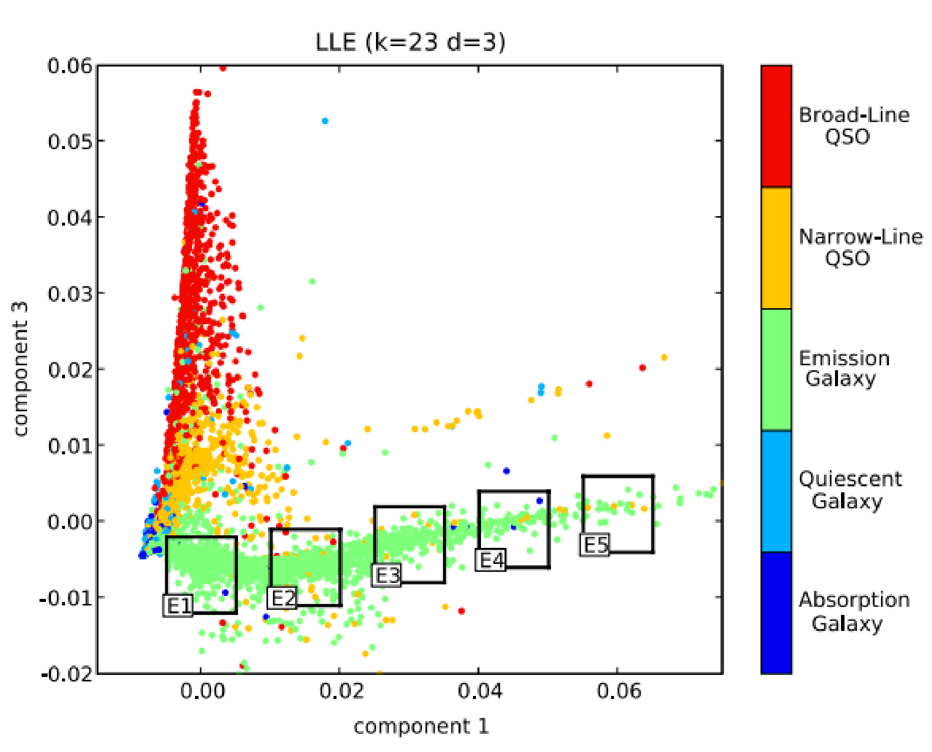
Apply LLE…

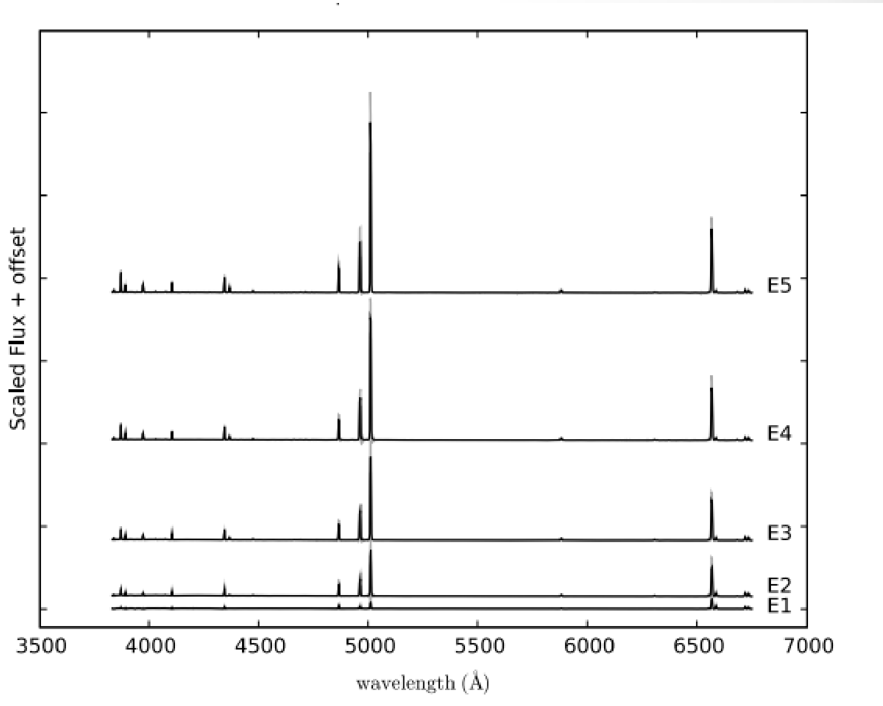

LLE (k=23 d=3)
component 3
component 1
0.06
0.04
0.02
0.00
-0.02
0.00
0.02
0.04
0.06
0.08
Broad-Line QSO
Narrow-Line QSO
Emission Galaxy
Quiescent Galaxy
Absorption Galaxy


LLE (k=23 d=3)
component 2
component 1
0.02
0.00
-0.02
-0.04
-0.06
0.00
0.02
0.04
0.06
0.08
Broad-Line QSO
Narrow-Line QSO
Emission Galaxy
Quiescent Galaxy
Absorption Galaxy

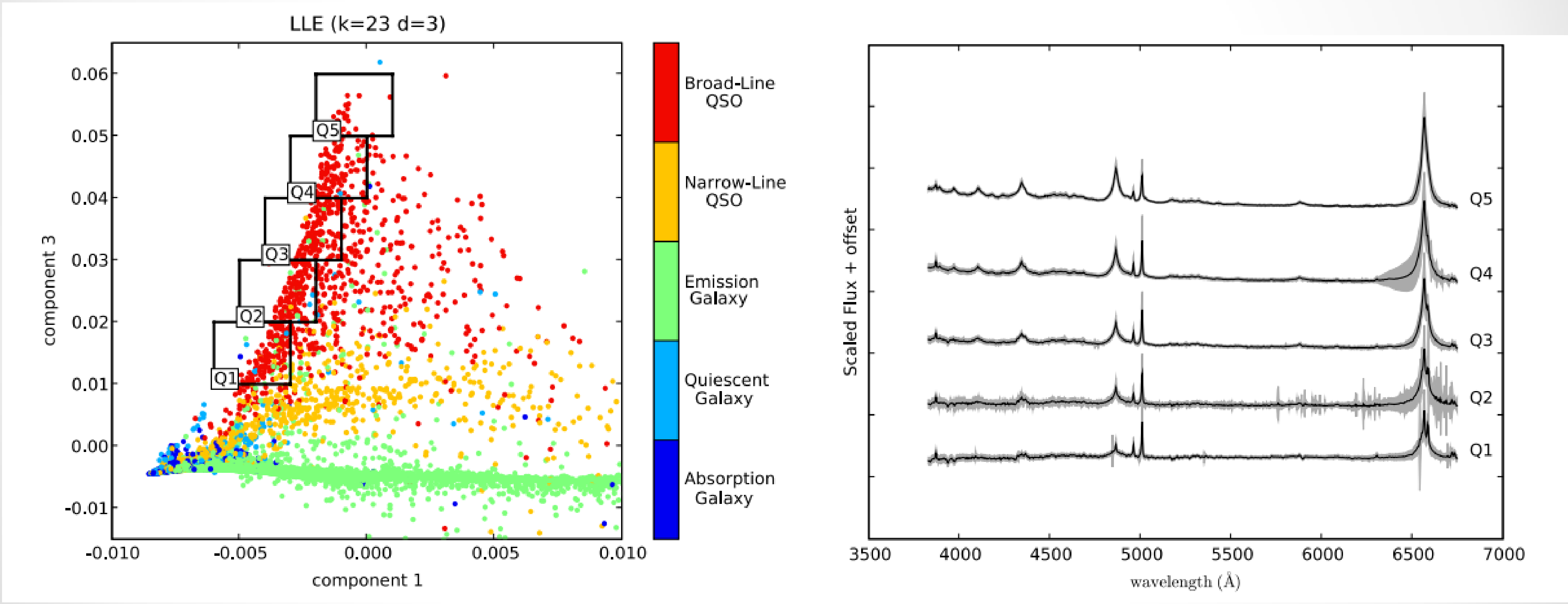

LLE (k=23 d=3)
component 3
component 2
0.06
0.04
0.02
0.00
-0.02
-0.06
-0.04
-0.02
0.00
0.02
Broad-Line QSO
Narrow-Line QSO
Emission Galaxy
Quiescent Galaxy
Absorption Galaxy

Progression of Emission Galaxy Spectra

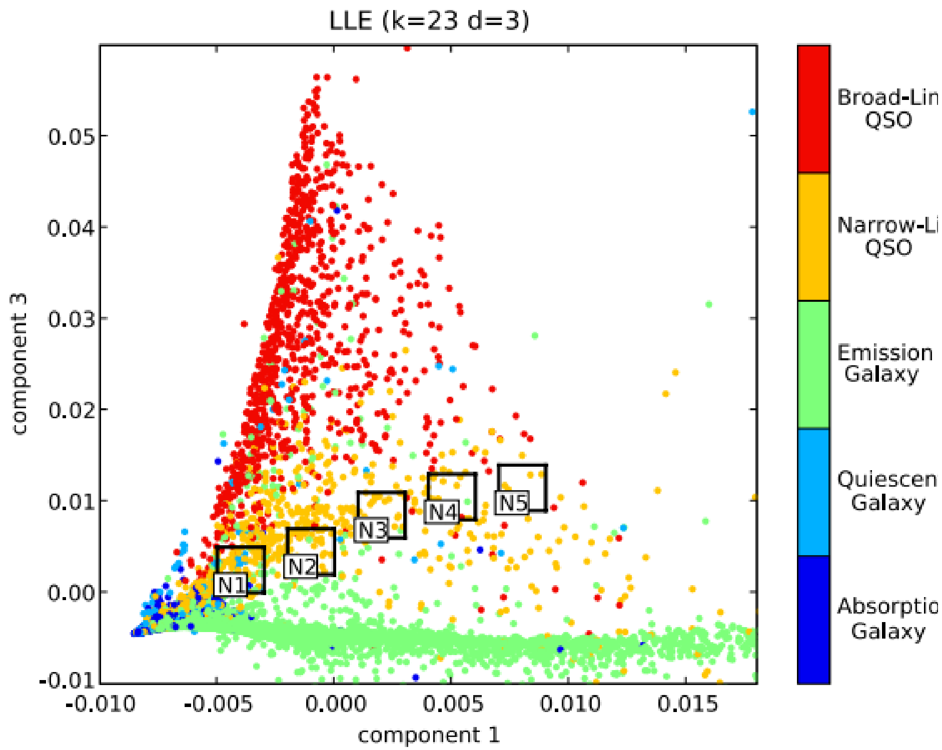


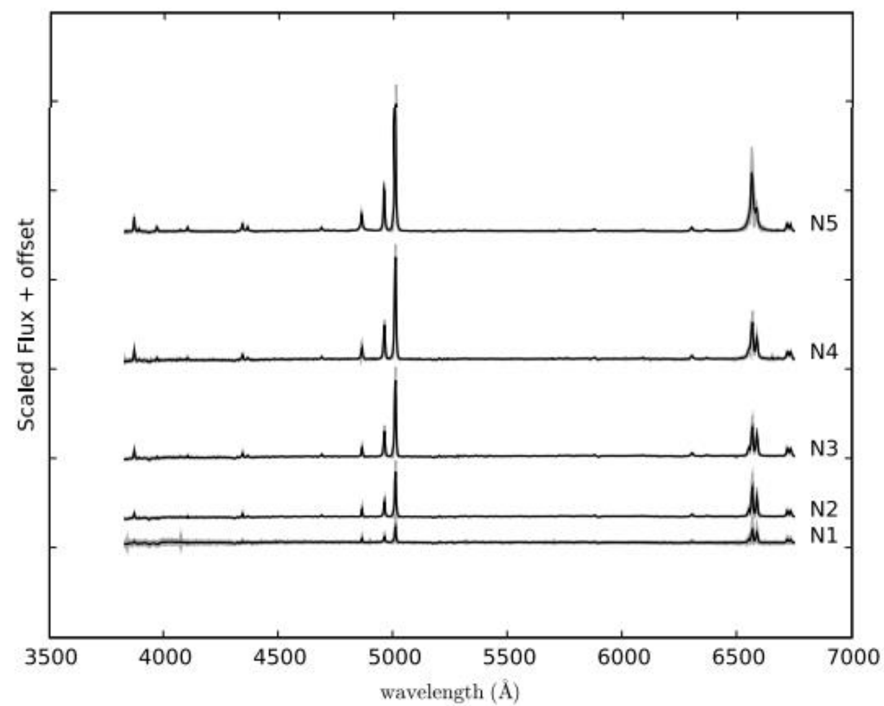


Progression of Broad-Line QSO Spectra

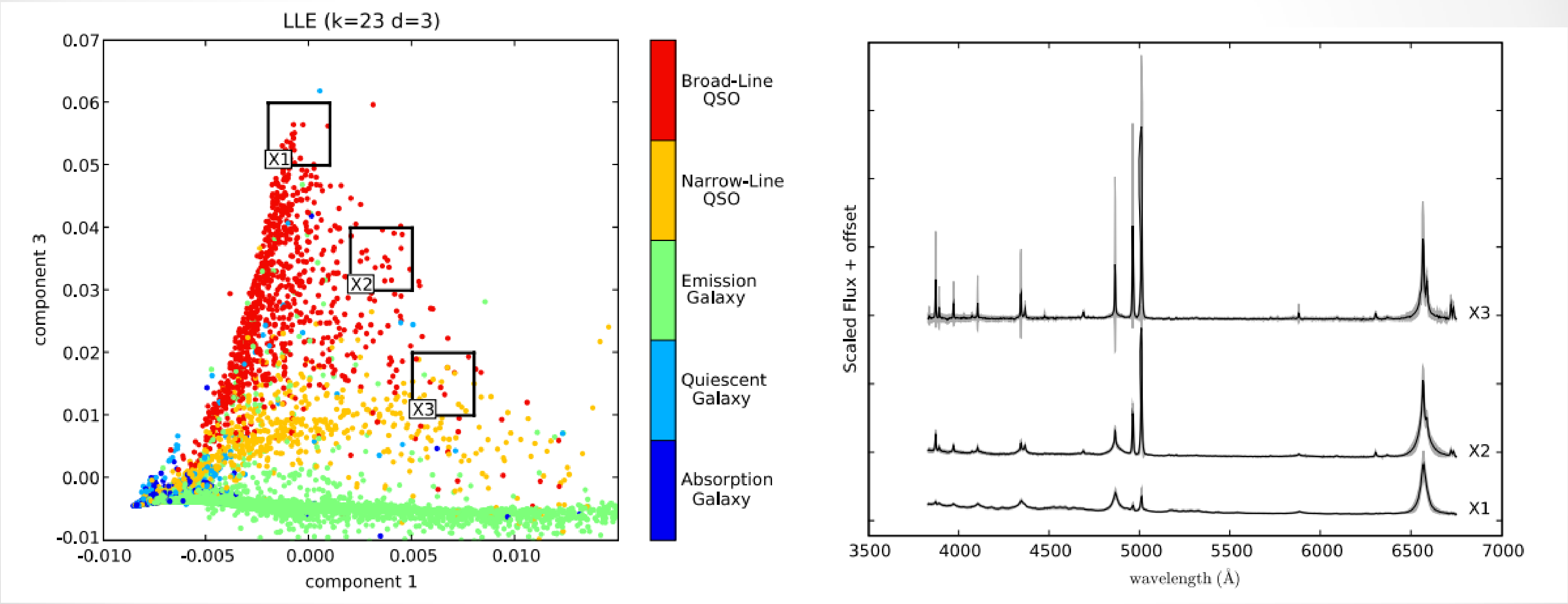


Progression of Narrow-Line QSO Spectra

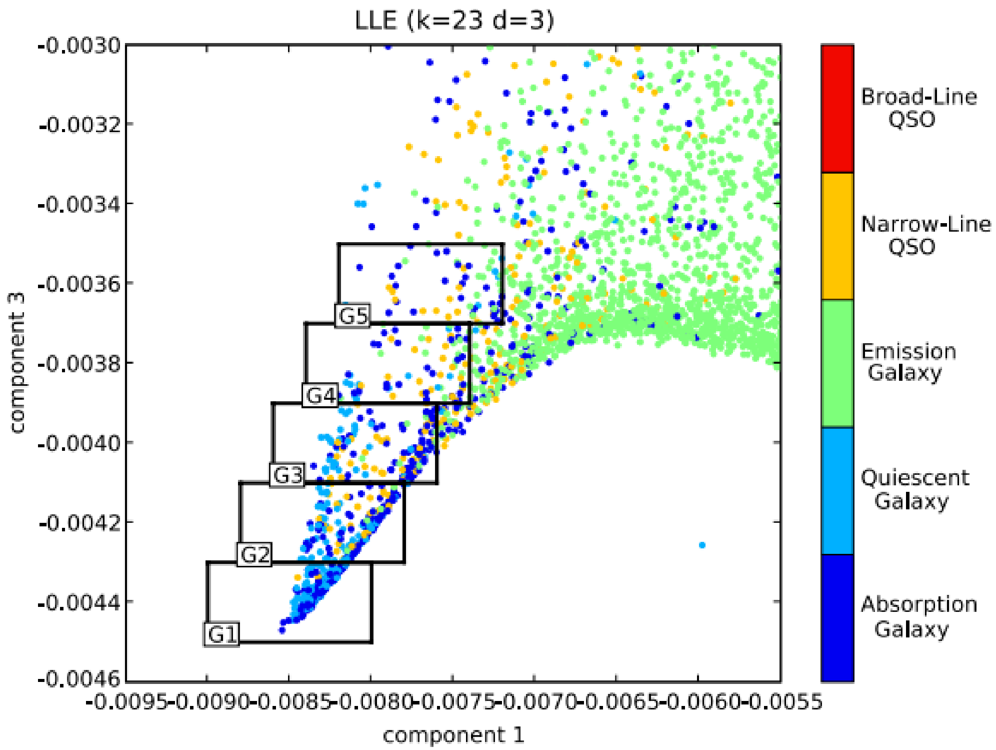


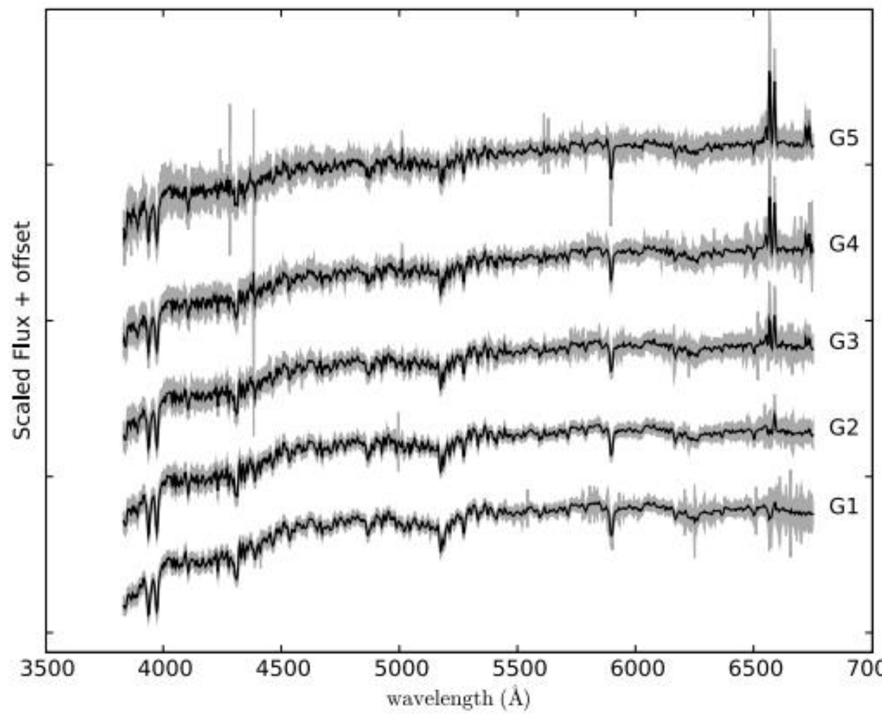


Progression from Broad to Narrow-Line QSOs

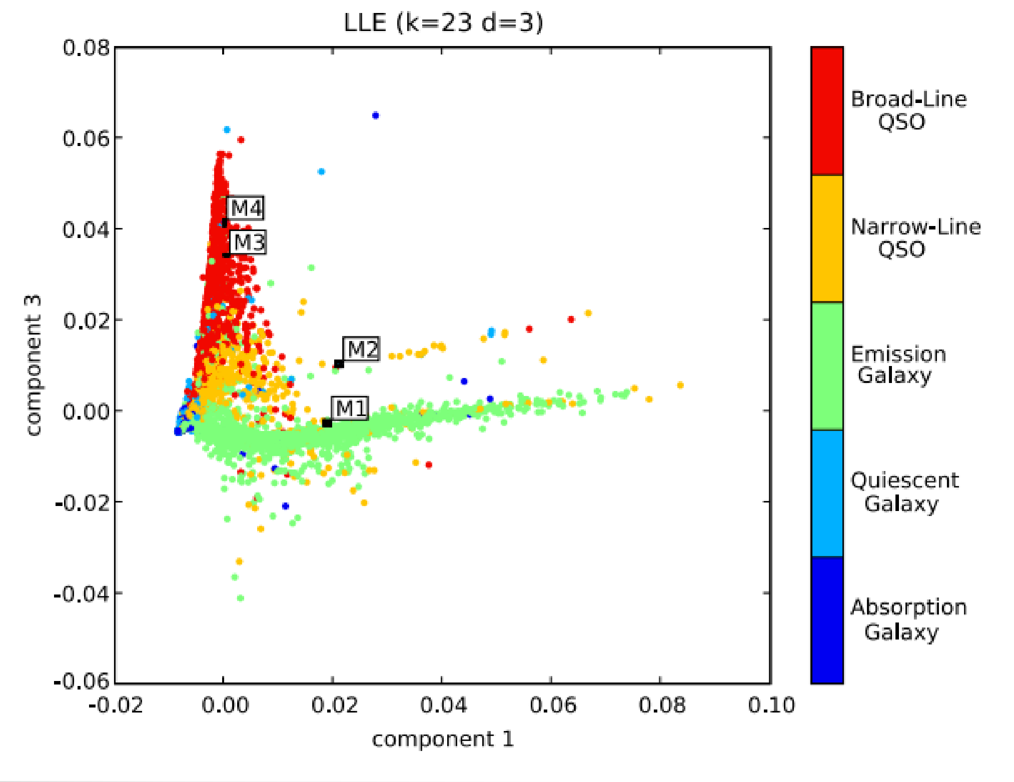


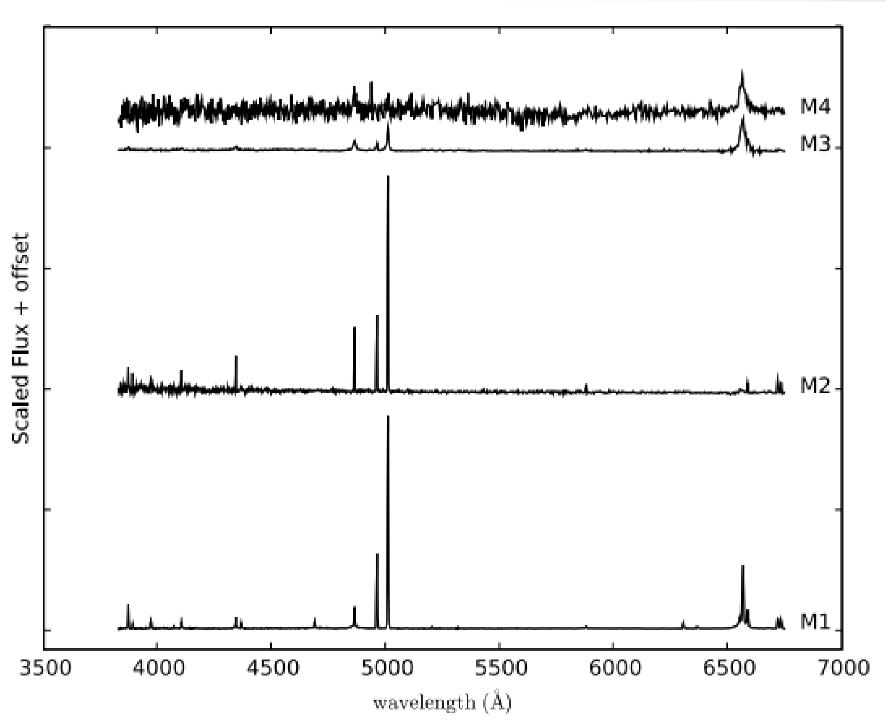
Progression of Quiescent Galaxy Spectra



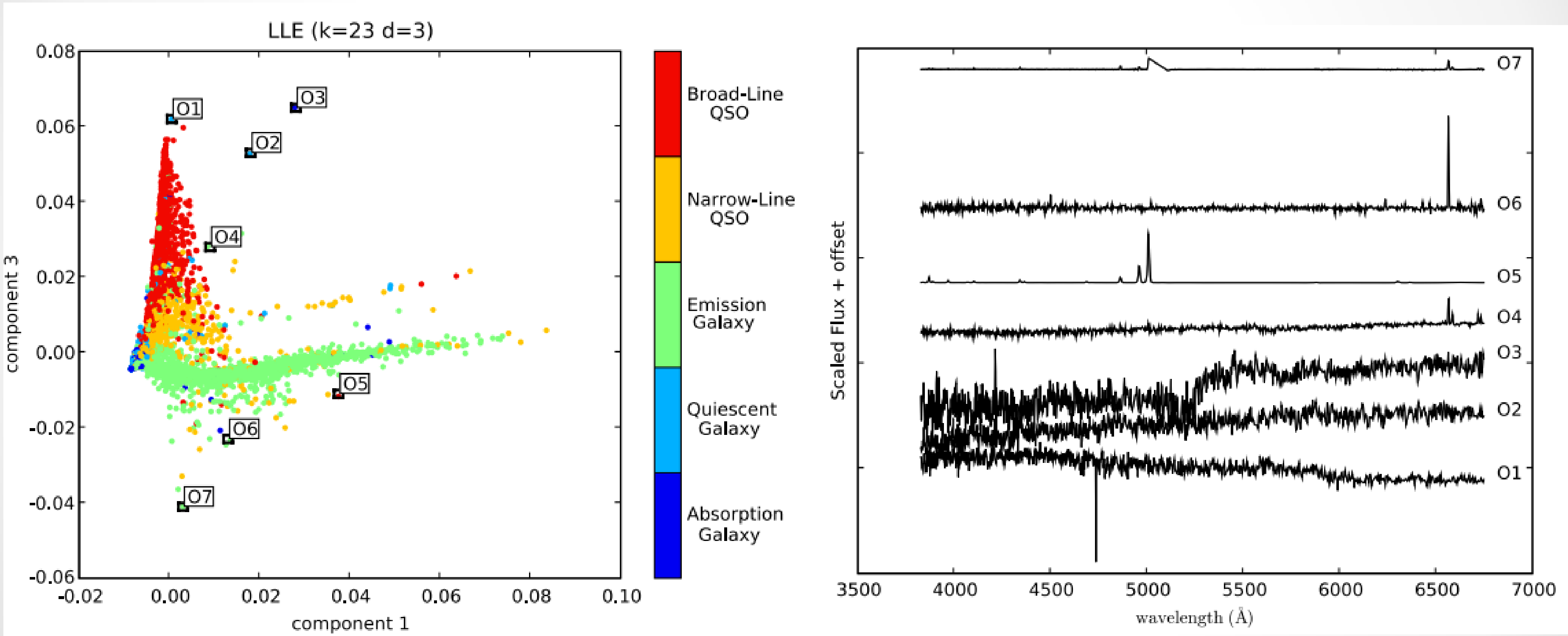


Objects Misclassified by Sloan



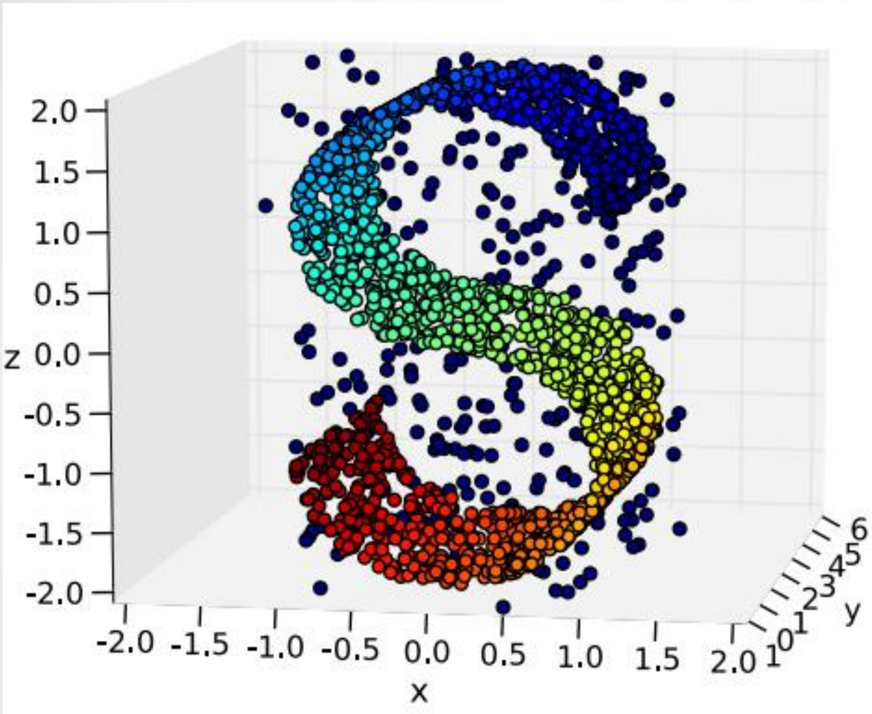


Outliers Found with LLE

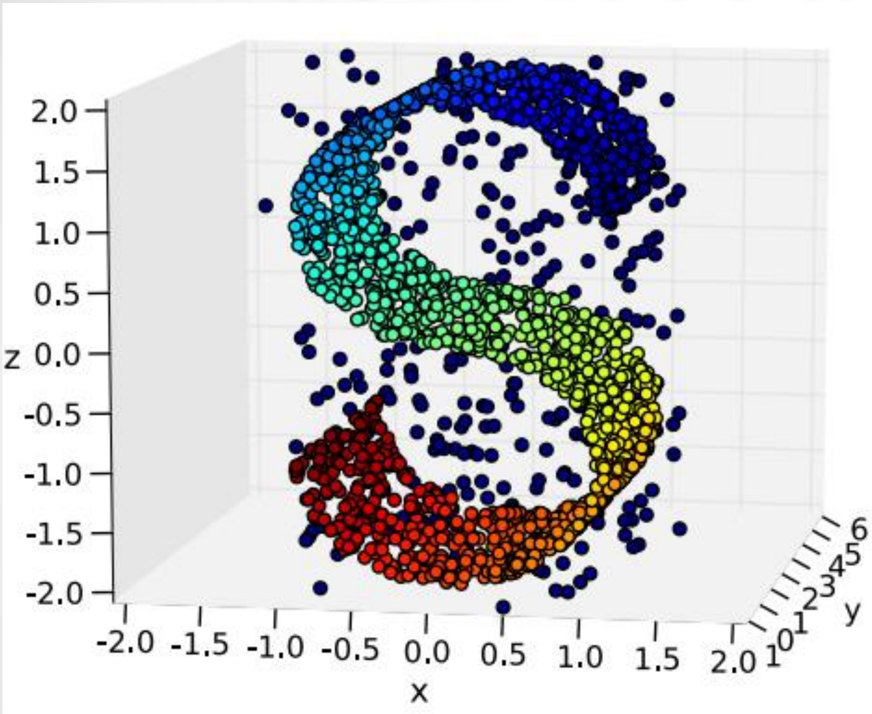


- How to keep outliers from skewing your results?

Must Account for Outliers!

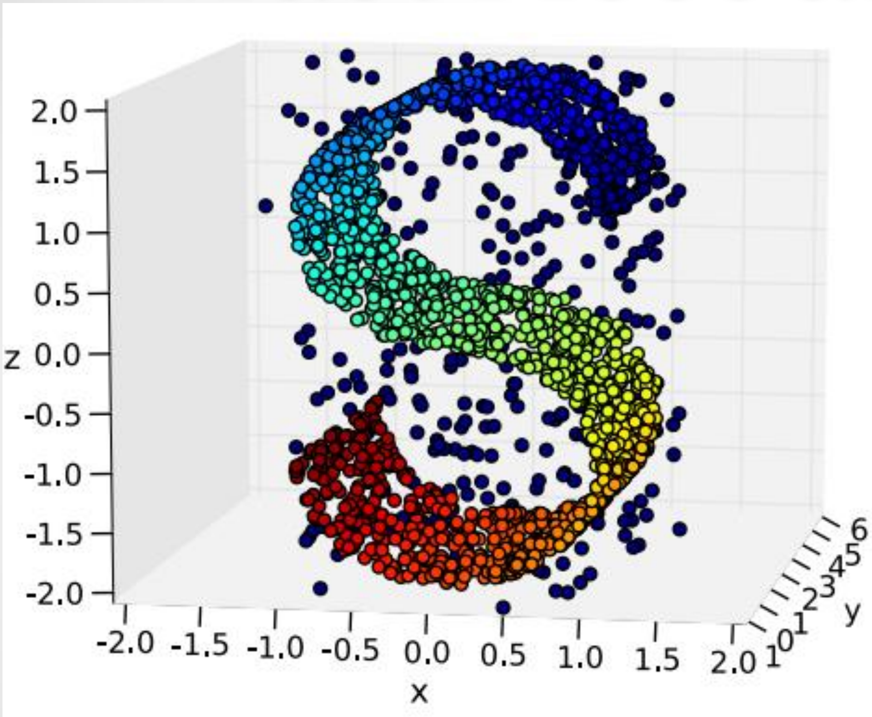


Must Account for Outliers!

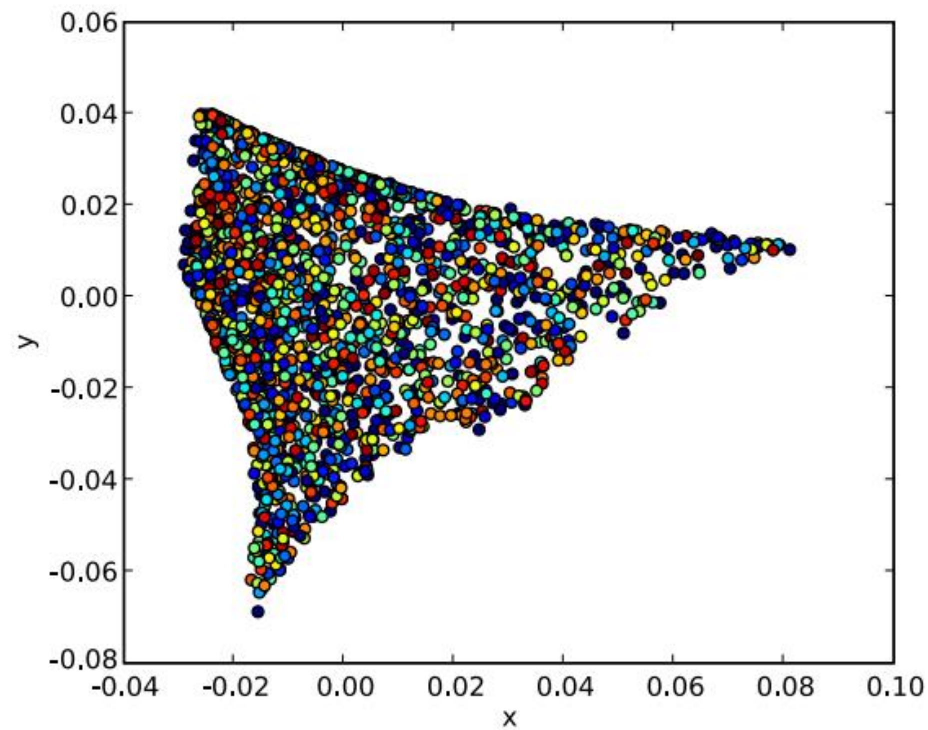


2

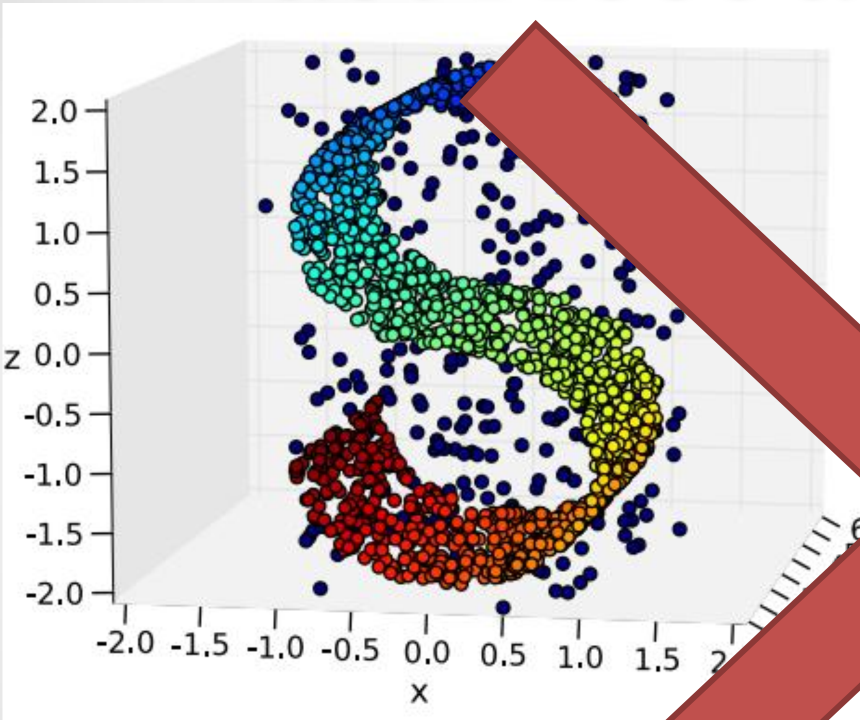
Must Account for Outliers!



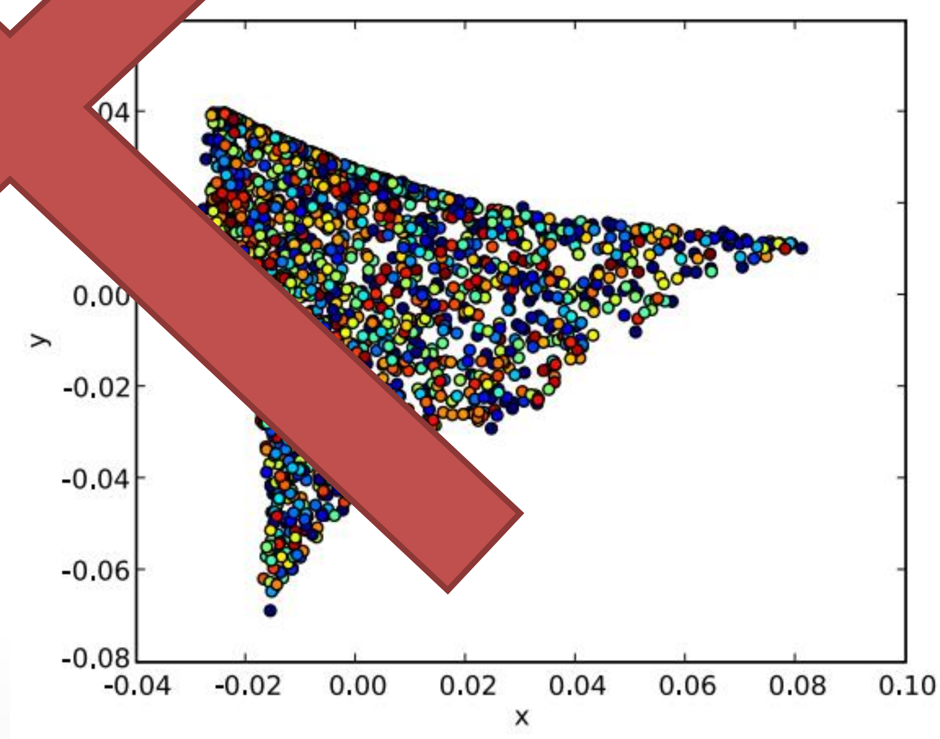
2



Must Account for Outliers!



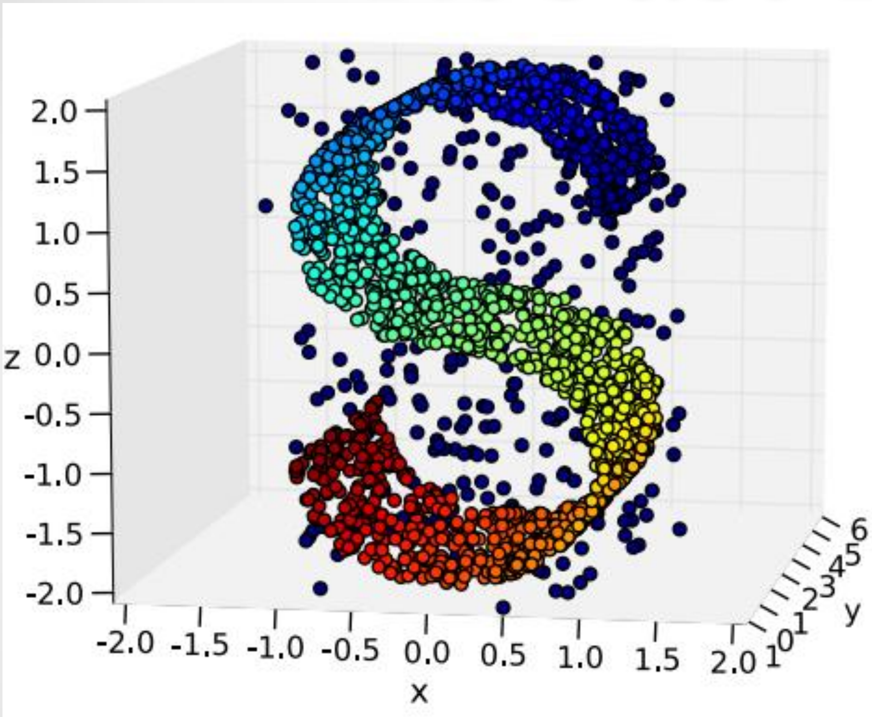
2

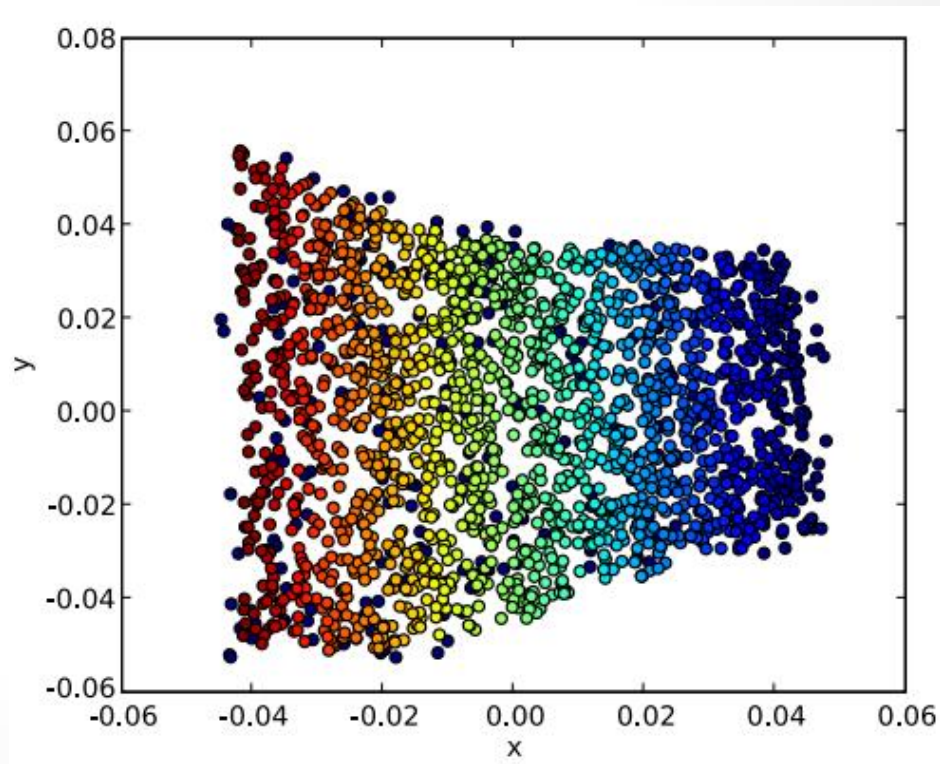


Robust LLE (RLLE)

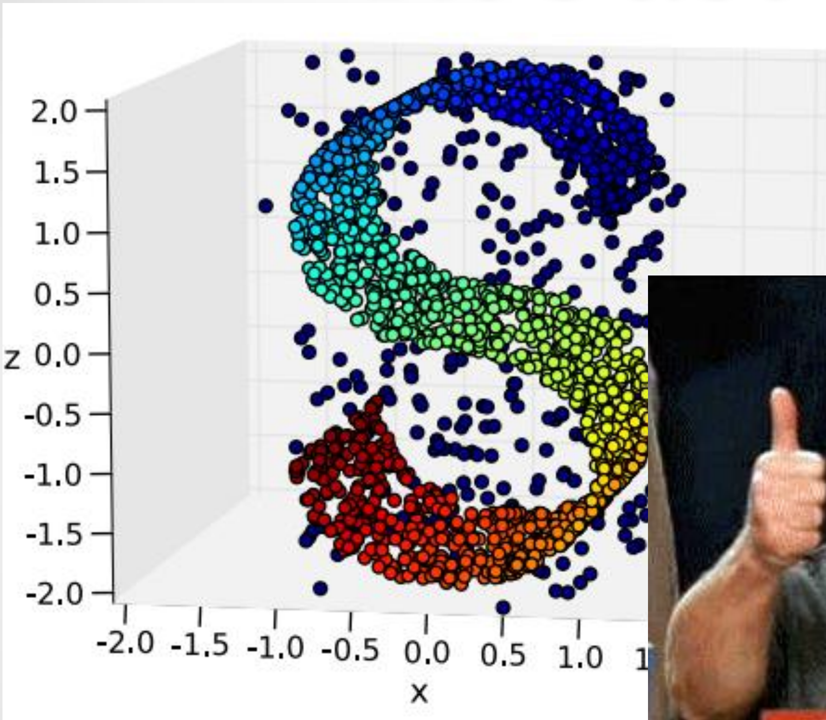
- Assign a "reliability score" to each data vector
- Outliers will not:
  - be a part of many local neighborhoods
  - lie near the best-fit hyperplane
- Perform an *iteratively reweighted least-squares reduction* on each data point to determine optimal weights for PCA reconstruction of data
  - Like finding best-fit hyperplane and determining each point's distance from it
  - Result: assigns local weights to each point for its contribution to local tangent space
- Sum all the local weights for each neighborhood
- Low reliability scores = outliers

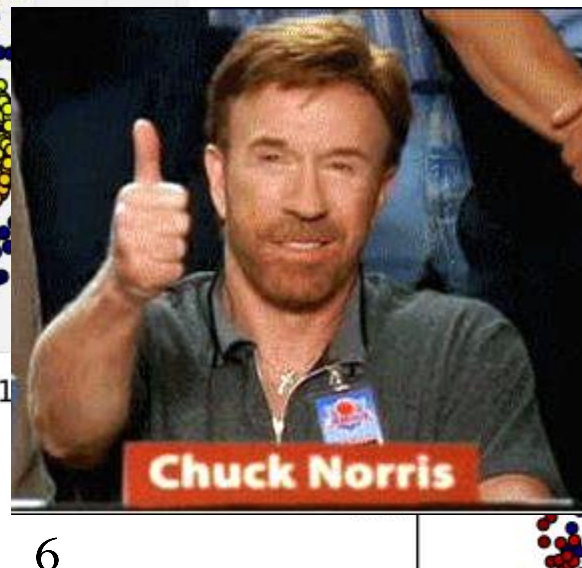
Robust LLE (RLLE)

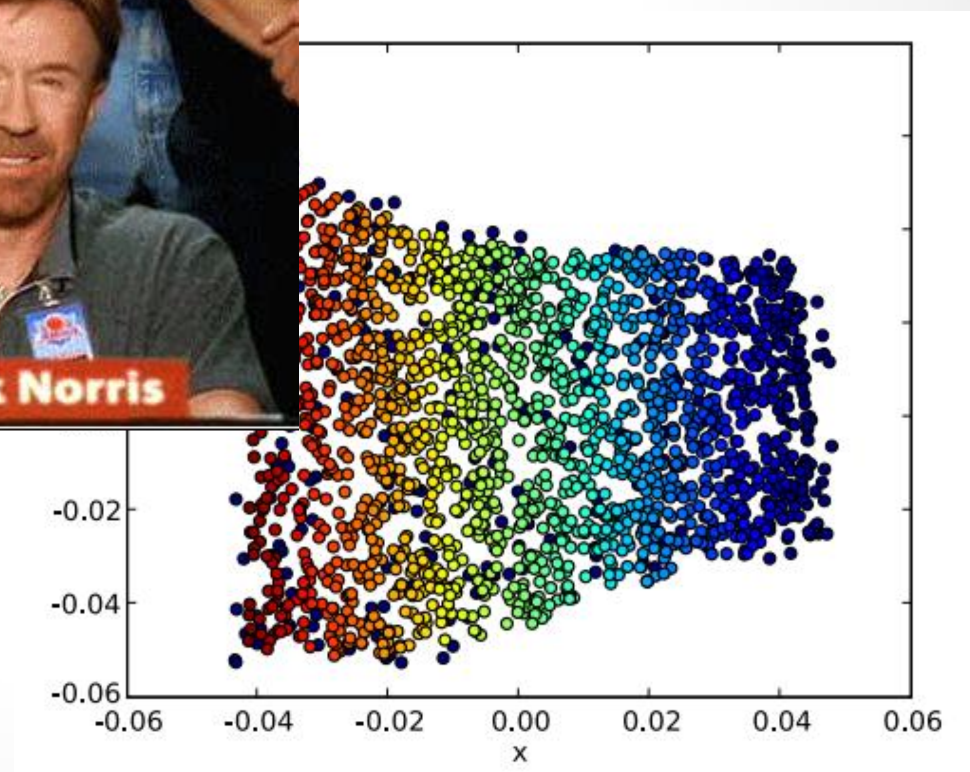




Robust LLE (RLLE)







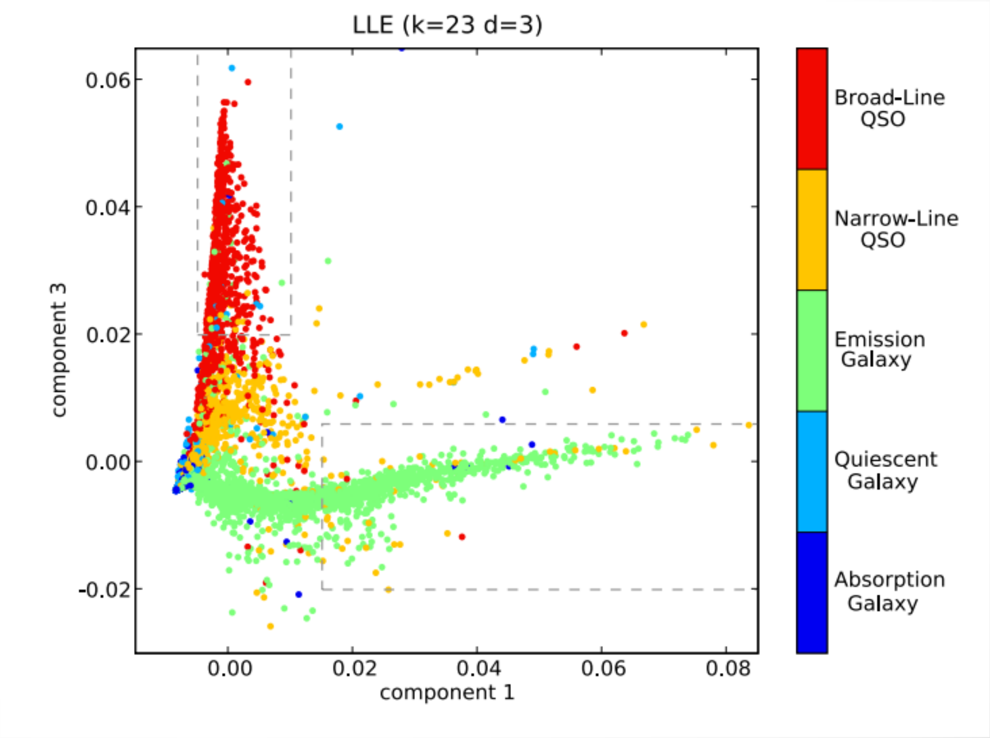
Choosing the Value of K

- Too small: undersampled, lose too much information
  - results easily skewed by outliers and noise
- Too large: manifold is oversampled, can no longer assume local linearity
  - cannot reduce the dimensions as much

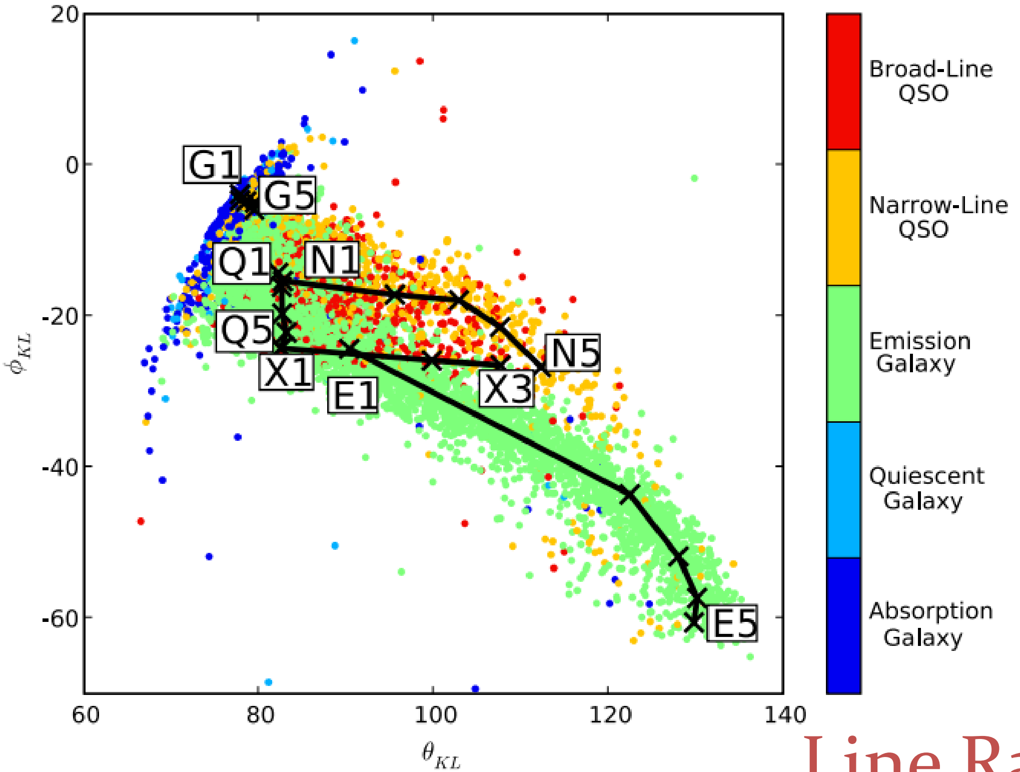
- How do we find the Goldilocks K value?

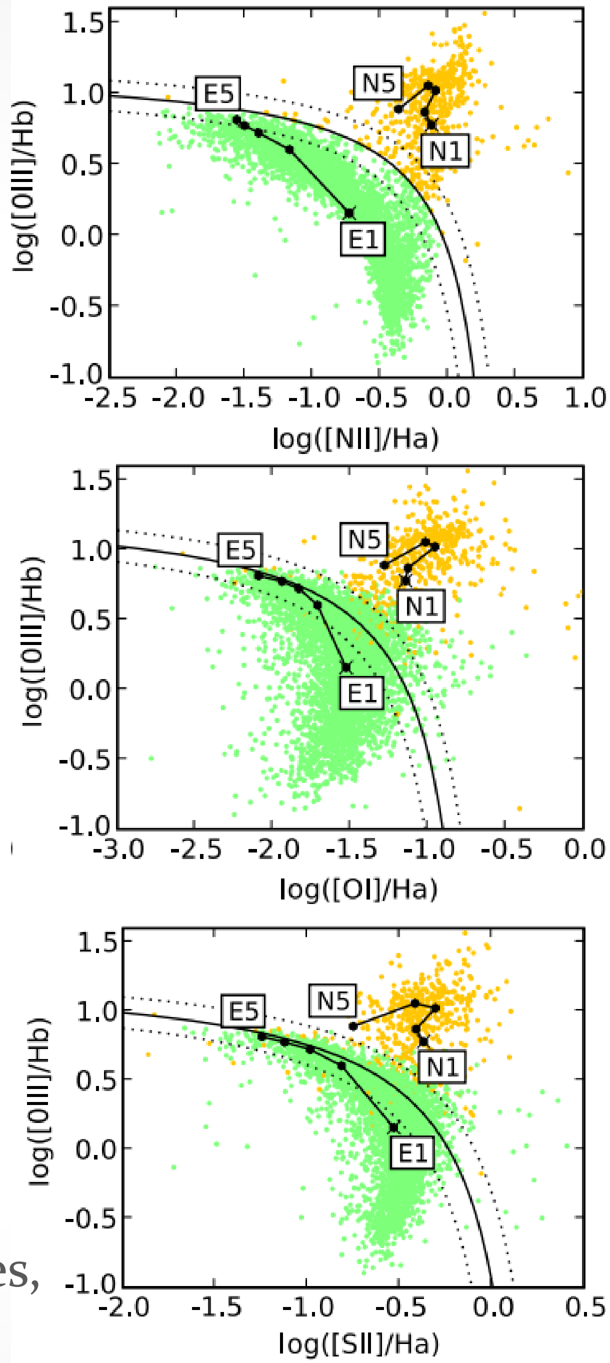
Finding K: Trial and Error

- Tried values from K=10 to K=30.
- Goal: Maximize the angle between the QSOs and the Emission galaxies
- Optimal K=23



Better than Other Methods?





PCA- good at showing variance,
weaknesses: nonlinear effects, Behavior of individual lines

Line Ratio Diagrams-
good w/ emission lines,
weaknesses: ignores continuum, works poorly w/ low S/N

Drawbacks

- High computational costs
- Bottlenecks:
  - Nearest neighbor search
  - Calculating optimal projection vectors

- Strategies:
- sample your data into smaller subsets
  - Vanderplas and Connolly did this! (~150,000 → ~9,000 spectra)

Sources

A. Roweis, S., & Saul, L. 2000, Science, 290, 2323

B. Vanderplas, Jake, and Andrew Connolly. 2009. "REDUCING THE DIMENSIONALITY OF DATA : LOCALLY LINEAR EMBEDDING OF SLOAN GALAXY SPECTRA," 1365–79. doi:10.1088/0004-6256/138/5/1365.

C. Saul, Lawrence K, and Sam T Roweis. n.d. "An Introduction to Locally Linear Embedding," https://www.cs.nyu.edu/~roweis/lle/papers/lleintro.pdf.

Image Credits

1. http://web.stanford.edu/class/ee378b/papers/roweis.pdf
2. http://iopscience.iop.org/article/10.1088/0004-6256/138/5/1365/pdf
3. https://i.imgur.com/96t6I9mh.jpg
4. https://twitter.com/jakevdp
5. http://www.washington.edu/storycentral/story/scanning-the-sky/connolly-andrew-11/
6. http://theactionelite.com/2012/07/the-return-of-chuck-norris/