AST 3162 - Spring 2014

DIMENSIONALITY REDUCTION

Due Date: 4/30/2014 @ 11:59 pm

In the era of large astronomical surveys the application of numerical methods for data analysis became a necessity. Among the most commonly used methods is the Principal Component Analysis (PCA). It is simple and performs very efficiently with normally distributed data, but fails to deliver when the underlying relationship between the data points is more complicated. Non-linear methods work a lot better in those cases aneasy to understandd there are a lot to chose from. We will have a closer look at Locally Linear Embedding (LLE).

- Implement the PCA method and test it on a 3D multivariate normal dataset that you generate with a (non-trivial) covariance matrix of your choice. Plot all possible 2D projections of the original and transformed datasets. Compute the covariance matrix of the transformed set and compare it to the original.

- Apply PCA and LLE to the dataset of 500 near-IR spectra sampled in 1000 wavelength points ranging from 8450 Å to 8750 Å. In both cases calculate the first three components and visualize the relationships between them. When using LLE try different values for the number of nearest neighbors. Use effective temperature, gravity, metallicity and signal-to-noise ratio data to see if the projected data are aligned with any of those parameters. Identify any potential outliers and check to see what the underlying spectra of those data points look like.