

1096-62-1122

Jeff Randell Knisley* (knisleyj@etsu.edu), Box 70663, Dept. of Math/Stat, East Tennessee State University, Johnson City, TN 37614-0663. *Consensus Spectral Techniques and Machine Learning.*

There is a long and fruitful history in the use of spectral methods in proteomics and genomics, especially when such techniques have been used to produce consensus spectra and consensus models for homological families of residue chains. Typically, these methods have focused on either periodicity – such as the fact that coding regions in DNA have 3-base periodicity – or on the use of biophysical measures such as the Electron-Ion Interaction Potential (EIIP) and its relationship to hot spots in amino acid chains. However, these methods can be extended into a more general framework, one that can be interpreted in terms of artificial neural networks and that subsequently can be applied both to complex networks and the data mining of large, complex data. Moreover, a major component of this approach is a resampling technique that allows not only the construction of a consensus model, but also a means of estimating confidence intervals from the resulting empirical distribution. We begin by establishing this framework, after which we move to its implementation, focusing on the machine learning toolkit scikit learn. The algorithms and examples in the talk will be freely available as an IPython Notebook shared via the data analysis framework Wakari. (Received September 13, 2013)