**Facilitating Discovery in Big Data Sets**

Kiri Wagstaff, Jet Propulsion Laboratory, California Institute of Technology

Given hundreds, thousands, or millions of observations, how can we decide which ones to look at first? Discoveries are often made by studying unusual or curious observations that do not fit a preconceived pattern or theory. Our goal is to develop methods for quickly prioritizing observations in large data sets so that the observations most likely to inspire new discoveries are examined first. When the data volume is so large that in-depth examination of each item is impractical, this prioritization is vital.

By modeling the current understanding of the data, we can predict which observations may be the most interesting by choosing those which deviate most strongly from the model. These observations are likely to contain new information that challenges the model. We have developed a method called DEMUD (Discovery through Eigenbasis Modeling of Uninteresting Data) that iteratively discovers, then incorporates, those observations into the user model. As the observations are iteratively chosen, the model evolves to reflect what new information has been seen. This can be done in a standalone, objective fashion, or user feedback can be incorporated into the model development to tailor the results for individual needs and interests.

We are using these methods to aid the analysis of data from a variety of instruments and scientific campaigns, including stellar time series from Kepler's search for stars with exoplanets, martian rock and soil emission spectra collected by the Mars Science Laboratory's ChemCam laser spectrometer, and atmospheric aerosol observations from the Multi-angle Imaging Spectroradiometer in Earth orbit. In collaboration with scientists from various disciplines, we are exploring the limits of known science by using DEMUD to quickly find the observations that are helping us advance our understanding of the universe.