White Dwarf-M Dwarf Pair Classification: Project Proposal

Ricardo Medina

April 14, 2011

1 Introduction

The classification of spectral data from sky surveys to isolate celestial bodies of interest is a matter of considerable importance to astronomers and astrophysicists. Indeed, it is a striking challenge to sift through the 600,967 stars with spectral data on the Sloan Digital Sky Survey Data Release 8. When a specific type of star is of interest, scientists have to come up with algorithms or other such techniques to find what they need. Processing spectral data (that is, the intensity vs. wavelength) is a computationally intensive and often unpromising burden. Another method used is to look at the intensity of large bands (*u, g, r, i, z*) the data for which is available from the SSDS. Because these intensities are largely based on the proximity of the stars, these bands are essentially useless unless somehow normalized. To account for this, the differences between two adjacent bands are taken: *u-g, g-r, r-i, i-z* (known as color-color). This data creates a four-dimensional color-space. Typically, however, no one thinks about the color-color data as such. Manipulations based on color-color bands are typically done as cuts where everything above a certain threshold is targeted for further inspection.

One specific kind of star that has interested scientists is White dwarf-M dwarf binaries. These binaries contain one low-mass, M dwarf, red star and one blue white dwarf. This wide range of temperatures (and therefore frequencies) makes the classification of these stars particularly difficult. Although studying these stars has astronomical significance because of their potential evolution to Type Ia supernovae, this project completely disregards that and is concerned only with finding them.

Two papers, one by Sean N. Raymond et al. and Nicole Silvestry et al., describe various procedures used to isolate as many as 800 of these stars from the fourth data release (DR4) of the SSDS. In 2003, Raymond applied harsh color cuts based on an earlier report by Silvestri that gave him a 60% accuracy (which was determined by examining the spectral data of the candidates).

In 2006, Sylvestri again sought out to find white dwarf-M dwarf binaries. After trying to refine the color cuts and realizing how inaccurate it was, she decided to visually inspect all of the spectral data for DR4 filtered only by dips in the absorption bands corresponding to the hydrogen emission of the bright white dwarfs. With this procedure she isolated approximately 700 white dwarf-M dwarf binaries.

2 My Project

My project will rely on Silvestri's data to make a four-dimensional logistic-regression to produce an ellipsoid in space that ideally contains, with improved accuracy, white dwarf-M dwarf binaries.

3 Functionality

My program should be able to do two things. First, it should be able to take a set of data of known classification (positive and negative) and find it's own regression coefficients. With these, the program should be able to process new data given to it.

The regression coefficients will be the linearized components of an ellipse which can then be represented graphically. For a two-dimensional version this will be 5 numbers. The conversion back to an equation of a recognizable ellipse will be done algorithmically with linear transformations (that I have to continue to do research on).

I plan on implementing all the steps of the logistic regression on my own.