Change Log

1. Title

I changed redshift throughout the paper to either observed or radial velocity depending on the context of the sentence. I made these changes in bold.

2. Keywords

3 SDSS data

I did not change this because I downloaded the spectra from Data run 12, 13, and 14 respectively. Basically I have three dr 12 copies, two dr 13 copies, and one dr 14 copy. I combined them into a single dataset. When I balance the dataset I generate a new random radial velocity for each sample, making each sample unique. Essentially there are no duplicates after data balancing.

5 Feature Selection

I changed the images by removing them. (This was not very clear) This is important because the Luminosity classes are based on the widths of the absorption lines. What I do is take each wavelength value from both regions and make them a dimension in a 34 dimension space where the flux measurement at a wavelength value is the length of the vector in that dimension. Then as the width of an absorption line changes with luminosity class, the position of this reduced spectrum in that 34 dimension space changes, allowing separation in luminosity classes. For the exact same reasons, this allows separation in spectral classes.

9. Table

Table 2 and Figure 11 do not represent the same thing. Table 2 represents how radial velocity causes machine learning complications because data of the same meaning appears in the wrong columns.

For example, the Iris flower dataset, which has three classes of flowers, and 4 features, as seen below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Petal Width (cm) | Petal Length (cm) | Sepal Width (cm) | Sepal Length (cm) | Class |
| 5.1 | 3.5 | 1.4 | 0.2 | Setosa |
| 4.9 | 3.0 | 1.4 | 0.2 | Setosa |
| 4.7 | 3.2 | 1.3 | 0.2 | Setosa |

In this example, each column has a unique feature of a flower. In stellar spectra, these unique features are the wavelengths. The purpose of table 2 is to demonstrate that this uniqueness is broken with radial velocity.

Figure 11 represents how the feature selection window is large enough to capture realistic radial velocities for stars.

Table 2 and Figure 11 are better used now that I have addressed changes in comments 4, 5, and 6.

10. Conclusions

I added a fourth conclusion about how correcting for radial velocity is unnecessary. I also added to the second conclusion a statement about how it works for small radial velocities. However, when dealing only with stars, if the bounds in algorithm 1 is to small, the number can be increased to account for larger radial velocities.

For the pre-processing conclusion, I added a statement stating that pre-processing should exclude any transformations to rest frame. From what I understand, the SDSS spectra are collected on a “blue” CCD and a “red” CCD and need to be recombined and re binned. Maybe I am citing the wrong papers, but that is what I mean by pre-processing.