

Machine Learning and Data Science Summary

Joseph Patrick Allen

April 2024

1 The Models Used

Machine Learning models are like computer detectives in the universe, but instead of solving crimes, they're using data to make predictions of objects in the night sky. 'Regression' models were used to find temperatures of objects in the night sky, and 'Classification' models were used to find out if that object one of the special categories we looked for.

Imagine you're trying to guess the temperature of a star based on its colour, and distance from Earth. A 'regressive' machine learning model is like a recipe that takes these factors and gives you a predicted temperature. It's like saying, "If a star is blue, and close to Earth, it's likely to have a temperature of 5,000 degrees Celsius". Now you're looking for couples in space that have a bigger and a smaller companion. That's where a 'classification' machine learning model comes in. It's like a filter that decides if a couple belongs to one of these special types based on the same factors (brightness, level of x-rays shot out) compared with the other uninteresting objects out there. So, if the bigger companion of a couple appears dimmer, shooting out lots of x-rays, with the smaller of two providing all the brightness, the model might say, "This belongs to the 'low mass x-ray binary' group." Difference being whether the couple is high or low mass lies in the size difference of the smaller companion in the couples. If the smaller companion was our sun, it would certainly be a 'low mass x-ray binary' compared with a object millions of times larger in a couple.

2 Predicting Temperatures

Below, every single temperature prediction is represented by a point and coloured according to the computer detective it came from. Whilst the green predictor lies closest the true temperature (the black line), it suffered from under predicting at lower temperatures. It is possible that the green predictor was so good in training, and that when tested with other night sky objects it couldn't perform as well. We call this 'overfitting'.

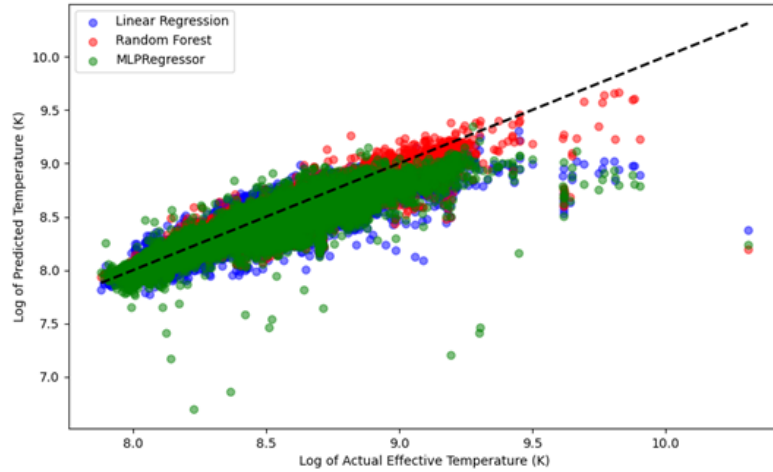
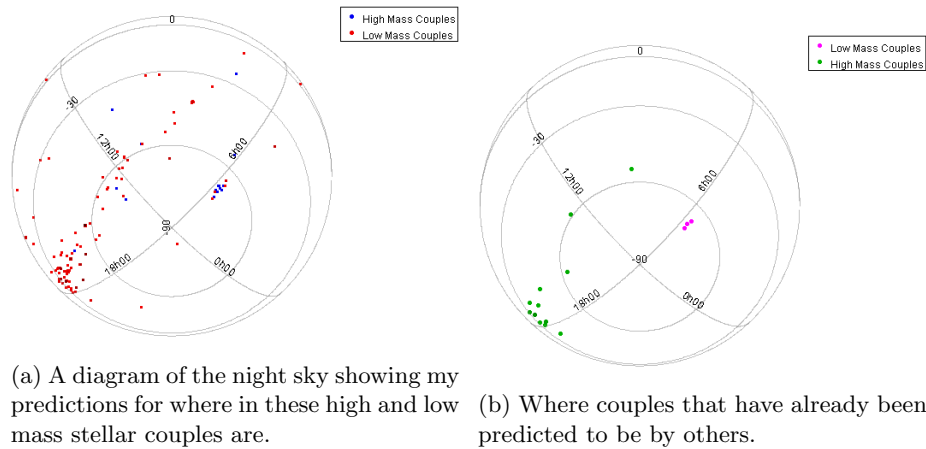


Figure 1

3 Hunting for Space Couples

After feeding three sets of 'Machine Learning models' hundreds of thousands of objects data, it has decided that below, each red and blue point represents a couple in the night sky that might not have been discovered yet! On the right are some of the predictions that have already been discovered.



(a) A diagram of the night sky showing my predictions for where in these high and low mass stellar couples are.

(b) Where couples that have already been predicted to be by others.

Figure 2: Notice how the couples seem to be in the same areas of the sky, this shows that these computer detectives (Machine Learning models) are working very well!