1. What is the difference between a rule-based system and a machine learning system?

A rule-based system is implemented as a set of rules, which is usually executed in some sequential or conditional flow, which means the exact steps of execution is known.

A machine learning system includes some kind of model which can be updated iteratively and used to predict a value; this can be singular or multi-dimensional. Mathematically, this equates to learning a function f(x) = y where x and y can be singular or multi-dimensional.

For real-world problems, there are too many unknowns for a rule-based system to be useful for solving these problems, whereas a machine learning system can be used heuristically.

1. What is the difference between unsupervised and supervised learning?

The major difference is the input data, where data for supervised learning includes labels for the ground truth, whereas data for unsupervised learning doesn’t include these labels. The reason for unsupervised learning is that data is expensive to label, since it requires human interaction, and so unsupervised learning can be used on extremely large, unlabelled datasets. But it’s limited in output, since it can’t predict values or classifications, but can cluster or autoencode data. The mid-ground is hybrid learning, where some data is labelled but most isn’t, meaning predictions can be made while learning on mostly unlabelled data.

1. What do we mean when we say that a machine learning system is overfitting?
2. Part 1 question 2.1 - include all steps
3. Part 1 question 2.2 – include Python snippets and RMSE performance
4. Part 1 question 2.3 – Write a table of metric score, write a small summary (500 words) of how preprocessing data, feature choosing, trained and evaluated model