

Question 3.1: Build a classification engine for being able to correctly classify given a set of readings the category they belong to. Note that, as per data above, different vehicles induce readings of differing lengths on the metal.

Since the classification task is trivial, I suggest the use of a decision tree model.

A decision tree is a simple model which will more than suffice for this task. We can also use a higher order regression function to solve this problem but I prefer a decision tree since the results can be nicely visualized.

Usually I would use bootstrap sampling in such a dataset because of small samples but there isn't much need as the data seems to be linearly separable.

Question 3.2: Demonstrate how that classification engine can be used for prediction given unclassified data.

The classification engine uses the decision tree model to predict the vehicle. The tree is visualized in the ipython notebook.

Decision trees use the concept of entropy to create divisions in the data set.

We can avoid overfitting by giving the max depth.

Question 3.4: How would your predictive model (built in Question 2) perform when there are 1000s of vehicles moving at 60km/hr every second on a national highway?

I would not suggest using a decision tree in that case. Decision trees are used for simplicity and interpretability.

I would prefer using Random Forests which are an ensemble model of decision trees.