Introduction to Data Science Portfolio

Week 6 – Classification Walkthroughs

TianQiu

This week we will use a parameter-free supervised learning algorithm to classify our data.

Firstly we collected a dataset called [pokedex ] dataset. The dataset is a collection of different Pokémon characteristics, such as attack power, defence, life value, etc. We can try to classify the Pokémon by their height, weight, attack power, defence power, and Pokémon Life Value.

For better calculation, I converted Pokémon's life value into a binary data, the Pokémon with greater than or equal to 50 life value is a category, and the Pokémon with less than 50 life value is a category.

I try to classify Pokémon with different hp by decision tree classifier by drawing decision boundaries in feature space, and try to classify Pokémon by different features, different decision depths, and compare the accuracy between them.

A screen shot of a graph

Description automatically generatedA graph showing height vs weight

Description automatically generated

Accuracy:85.0 Accuracy:81.75

A graph of a graph with orange and blue dots

Description automatically generatedA screen shot of a graph

Description automatically generated

Accuracy:76.0 Accuracy:81.0

A graph showing a number of dots

Description automatically generated with medium confidenceA graph showing a number of dots

Description automatically generated with medium confidence Accuracy:76.5 Accuracy:76.75

We can see from the accuracy that classifying a Pokémon's hp by attack and defence is the most accurate, and that increasing Max\_Dept does not increase accuracy, but rather decreases it.

## Inspecting the tree

We can see what a decision tree with the best feature pairs looks like and try to see how it has balanced both large splits and low gini impurity scores by way of images.

A diagram of a algorithm

Description automatically generated

We can see that the first level of classification is segmented by the value of defence, the second level is segmented by the value of attack and the third level is segmented mainly by the value of defence.

We again try to use all the features for input and build the inspecting tree.

A diagram of a diagram

Description automatically generated with medium confidence