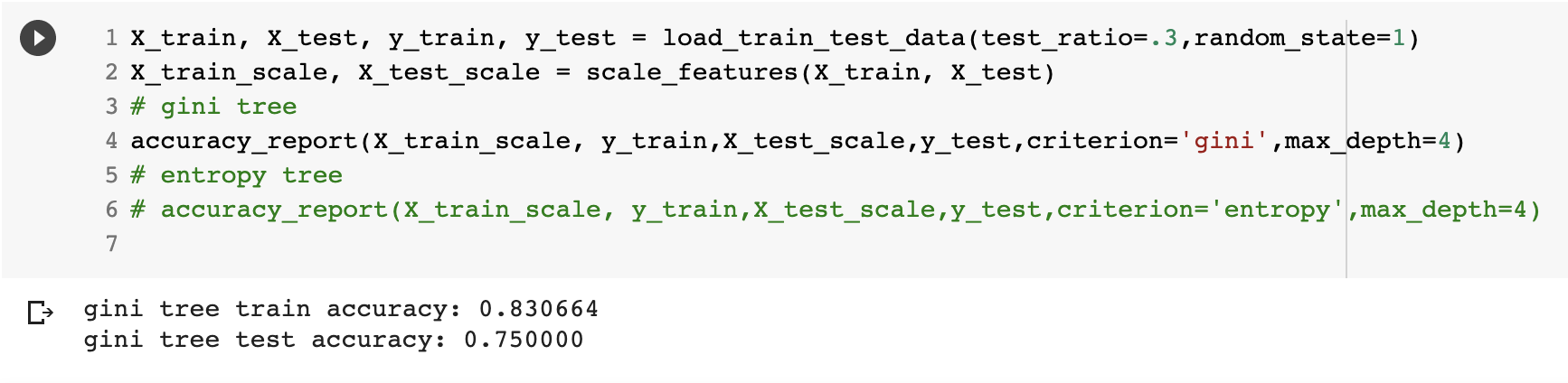
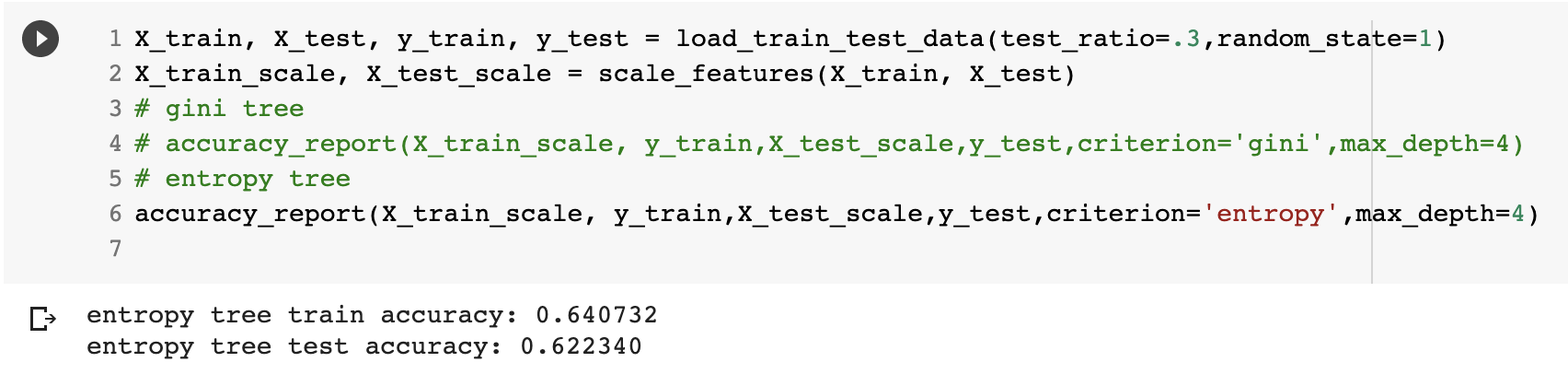
資料科學導論 exercise2 Decision Tree

109525009 張祐綸

Colab:<https://colab.research.google.com/drive/1jKe4wmJN2BaHI8GNt6EPFfmQ-IYxAtWd#scrollTo=mTPGsFnriUp7>

* The accuracy on both training and test data





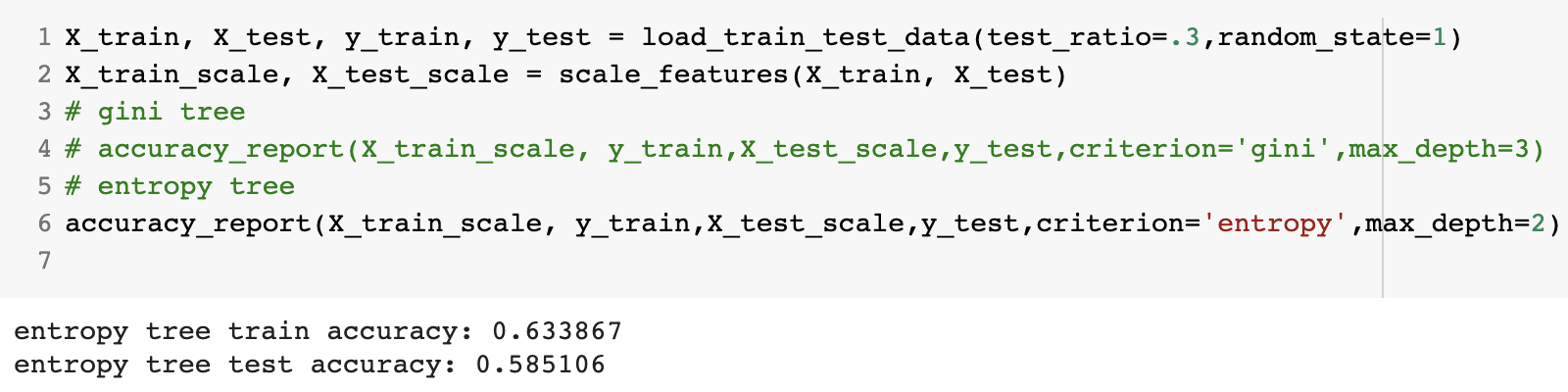
* The effect of different parameters (Ex: criterion、max\_depth)

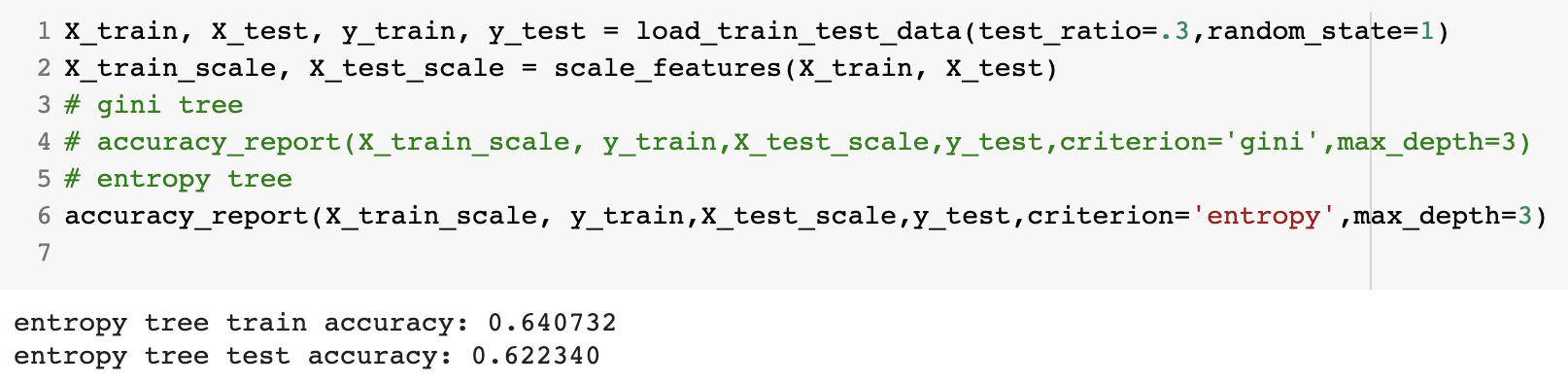
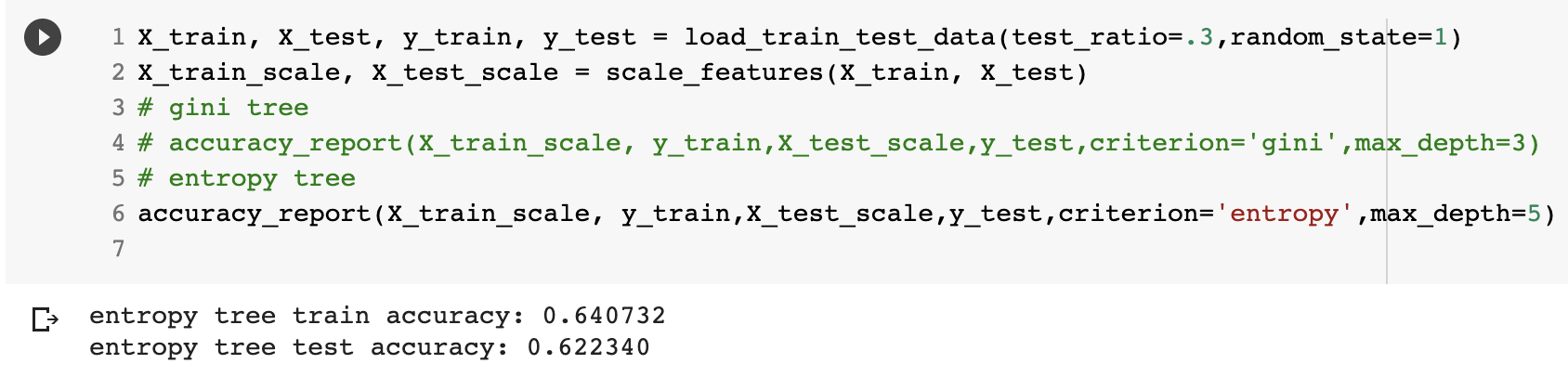
1. Criterion：through the experience I had try.

Gini tree is much better than entropy tree.

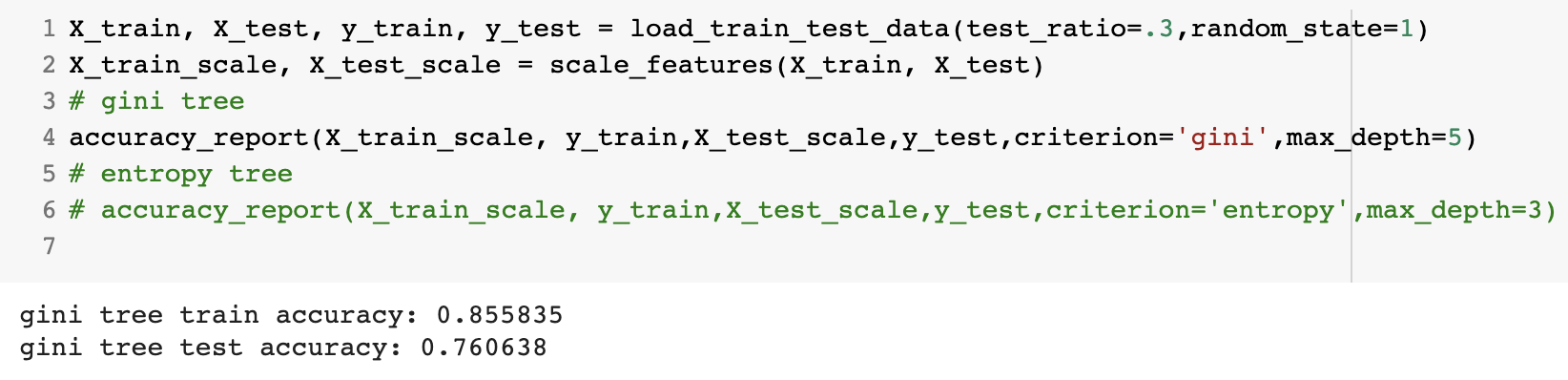
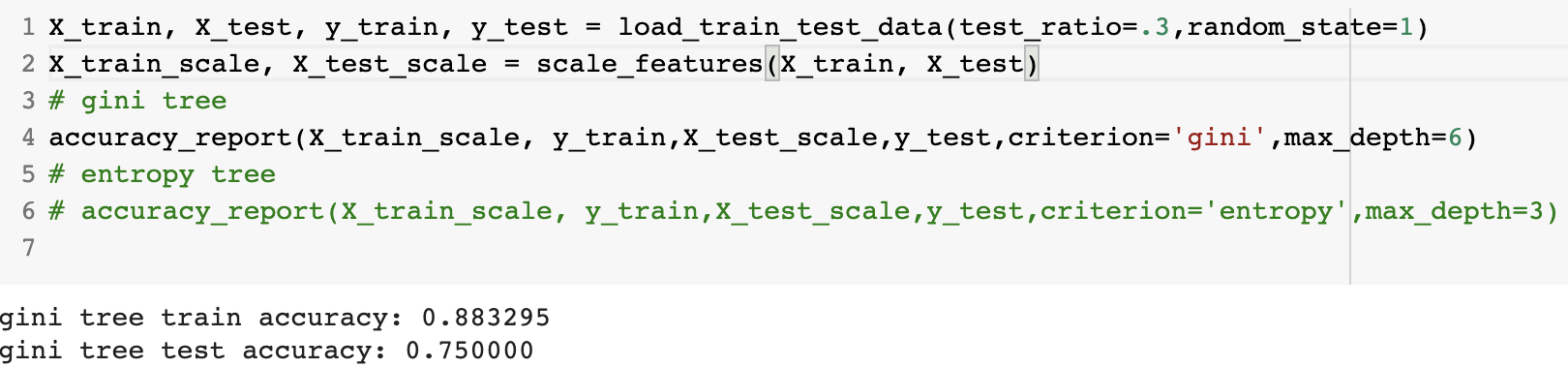
1. max\_depth：

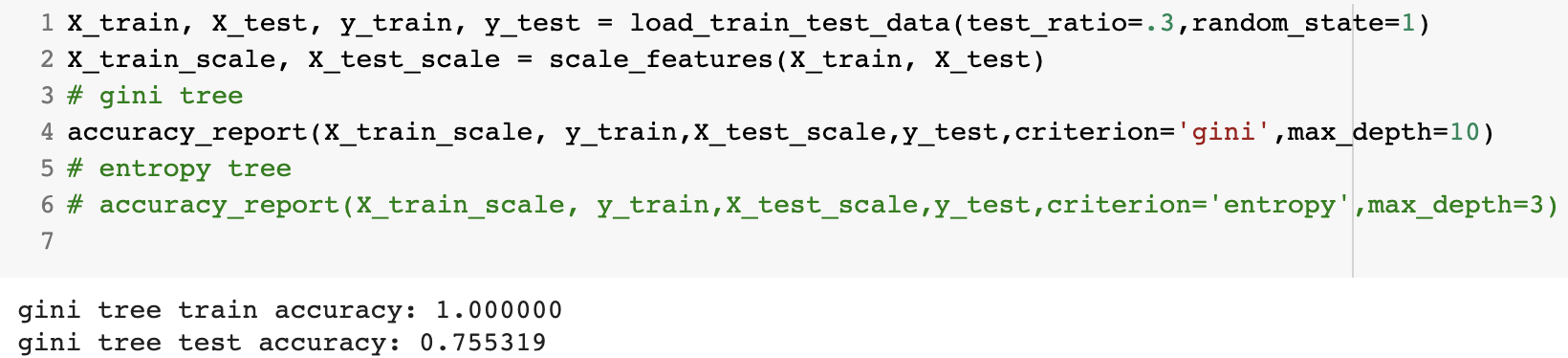
If the max\_depth of entropy tree is over 3, there is no improvement in both train and test accuracy. I think the best max\_depth of entropy tree is 3. Because over 3 both test and train accuracy won’t improve.



If the max\_depth of gini tree is over 10, there is no improvement in both train and test accuracy. It train accuracy equals 100. I think the best max\_depth of gini tree is 5. Because over 5 the test accuracy won’t improve a lot, but the train accuracy will close to 100% which may lead overfitting.



* A brief discussion of the results

Overall, gini tree is much better than entropy tree. No matter in train accuracy or test accuracy.