**Machine Learning in Finance - Homework 4 – Due 10.01.2024**

Textbook reading:

Chapters 5 to page 129

* **start from 'max depth DecisionTreeExample.ipynb' in colab**
* **read in the load\_wine database from** [**https://scikit-learn.org/stable/datasets/toy\_dataset.html**](https://scikit-learn.org/stable/datasets/toy_dataset.html)
* **create decision tree, random forest, and gradient boost trees of depth 3**
* **Display a pandas data frame containing the feature importances with columns the 3 tree types and rows for the features**



* **What are the 3 most important features for each of 3 different tree methods?**

Decision Tree: flavonoids, color\_intensity, proline

Random Forest: proline, color\_intensity, flavonoids

Grandient Boosting: proline, color\_intensity, od280/od315\_of\_diluted\_wines

* **Compute the accuracy for the 3 methods as you go from a depth of 1 up to a depth of 9 and display in a graph? (hint: save the outputs for each depth to a pickle file and write code to read the back in once you have created all the pickle files)**

**A graph of a test set

Description automatically generated**

**A graph of a training set

Description automatically generated**

* **Given this information (since you don't have a test set), argue that there is a best max depth**

We do have a test set (?). Max depth on Gradient Boosting Classifier, surprisingly, has a maximum accuracy at 1. For the other two models, a max depth of 3 seems to be the most optimal choice as accuracy score not longer improves.