ChE 197/297: Intro to AI/ML for Chemical Engineers
Case Studies in ChemE

Instructions: Answer each problem then create a solution using Python code via Jupyter Notebook.

Problem: Classifying wines from their chemical analysis using weak learners and ensemble learning

Figure 1. Different wine cultivars (cultivated varieties of wine) from Italy.

You are given a data set of the composition of 178 different wines, coming from 3 different cultivars. The composition contains Alcohol, Malic acid, Ash, Alkalinity of ash, Magnesium, Total phenols, Flavanoids, Nonflavanoid phenols, Proanthocyanins, Color intensity, Hue, Protein purity, and Proline.

Our task is to classify the cultivar where the wine came from. Do the following.

1. View the spread of the wine composition values for each class using box plots.

2. Split the data into training (70%) and testing (30%). Stratify this according to the cultivar classes. For this problem, the models that we will use do not require normalization.

3. Train a k-Nearest Neighbor classifier with k=5, a Gaussian Naïve Bayes model, and a Decision tree with max_depth=10. Report the training and testing accuracies and confusion matrices for all models.

4. For the Decision Tree classifier trained in item 3, view the tree itself using “plot_tree”.

5. Train a Random Forest model and an XGBoost model on the same data set. Use their default settings. Report the accuracies and confusion matrices.

END OF EXERCISE