Submission Date: 13th December 2023

1 Bagging

1. Bootstrapping is random sampling with replacement. Discuss the advantages and disadvantages of bootstrapping.
2. What is bagging and why do we use it?
3. What is the difference between Bagging and Boosting? Given a scenario where your base model is performing well on the training dataset but not on the validation dataset. Which of the two ensemble techniques you learned so far would be suitable to apply here? Comment on how the complexity of the model would matter when choosing an ensemble method.

2 Random Forests

1. Implement the missing code parts in `Python Scripts/08_assignment.py`:
   (a) The `bootstrap` function
   (b) The `__init__`, `fit`, and `predict` functions of the class `RandomForestClassifier`.
2. What is the accuracy you get with the default hyperparameters?
3. Set the value of `bootstrapping_size` to a lower number and then to a higher number than the default. How does the accuracy change? Why is this happening?

3 Boosting

1. Consider the data

x_1	x_2	x_3	y
-1	+1	+1	-1
+1	+1	+1	+1
-1	+1	-1	+1
+1	+1	-1	+1

Perform three rounds of Adaboost with one-level decision trees (stumps). For calculating the splits use the loss:

$$L(y, \hat{y}, w) = \sum_{i=1}^{N} w_i e^{-y_i \hat{y}_i} \tag{1}$$

2. Calculate the predictions for the following test set

x_1	x_2	x_3
+1	-1	-1
-1	-1	-1
+1	-1	+1