



COMP4388: Machine Learning

Fall 2021/2022

Assignment sheet #3: Deadline 20 January 2022

Please answer the following questions with each in no more than 2 pages. Please make sure to include proper references and citations.

Q1) Bagging and Boosting are two methods for ensemble learning algorithms. Please answer the following:

- a. What is Random Forest?
- b. What is Boosting?
- c. Explain in your own words the difference between Random Forests and XGBoost.

Q2) In this part, you will experiment the difference in the behavior of Random Forests and C4.5 Decision Tree. You have to report the Accuracy, Precision, Recall, F-Score, Bias, and the Variance of each.

Use the following binary-class dataset

(<https://www.dropbox.com/s/8r3nfglhg96n63/COMP4388-Dataset1.csv?dl=0>). Split the data 80 training/20 test and explain the steps of your work (i.e., if you scale data, how did you split, the performance measures, bias and variance, and state your findings in your own words). You have to provide your Python code as well.

This is an individual assignment. Please make sure that the submission of this assignment is as a reply to the message sent on Ritaj and make sure that the file is names "A3-XXXXXXX.docx" where XXXXXXX is your student ID.

If you have any questions, please feel free to ask during the lecture or by contacting me via Ritaj or email: rjarrar@birzeit.edu

Good luck!

Dr. Radi Jarrar