FIN 6930

AI & ML Applications for Finance & FinTech

Assignment 2

Due June 11, 2023 @ 11:59 PM

1. A **larger data** set for the **Lending Club** case is **attached.** It contains additional features, including loan amount, term of the loan (i.e., when it is supposed to be paid off), interest rate on the loan, etc. that significantly increases the number of features that can be used to predict the likelihood that a borrower will default or not. Most of the features are fairly straightforward to understand. For those are less common, I provided a description on a separate sheet (cf. "Feature explanation".)

    a. For this data set run both a logistic regression and a decision tree as in the last homework to predict "good/bad" loans
    b. How do these models compare with the previous ones? (Does the addition of more features change anything?)

    Note: For this question, you are going to "clean your data" as some entries are missing. You will also need to convert some data into categorical forms.

2. Refer back to the **Python code for the implied volatility** discussed in class. Consider the **effect of the following changes** (keeping all the other elements of the NN structure in place):

    a. Number of hidden layers changed from 3 to 1
    b. Number of hidden layers changed from 3 to 5
    c. Number of nodes per layer changed from 20 to 10
    d. Number of nodes per layer changed from 20 to 40
    e. "Sigmoid" activation function changed to "ReLu".