FIN 6779

AI & ML Applications for Finance & FinTech

Assignment 2

Due June 8, 2025 @ 11:59 PM

1. Use the **Lending Club Data** posted under "**Data and Code**" under the page labelled "**Module 2: ML Applications in Finance (Week 2)**"**.** Refer now to **slide 4** of the "**02.06-Supervised Learning_DecisionTrees -Credit Application"** slide deck. Consider 11 potential values of the **threshold X**, starting with $34,000 and increasing in increments of $8,600 up to $120,000. In other words, $34,000, $42,600, …, $11,140, $120,000. For each of these values find the corresponding values for $P_1, P_2, \text{and } P_3$, and expected entropy associated with the income feature. Display the results in tabular form with column headers X, P_1, P_2, P_3, and E_{income}. Out of the possible values of X considered here, which would you pick to make income a good candidate for the root node?

2. Refer to the Python **Code for the Lending Club Decision Tree**. Based on the displayed tree, which combination of features leads to the highest likelihood of no default? (in other words, you need to find the path that generates such likelihood.)

3. Modify the **Python Code for the Lending Club Decision Tree** so that the threshold values for the probability of a "good loan" are changed from **.75, .80, and .85** to **.90 and .95**. If you were to pick between these five thresholds, which one would you pick? Explain your answer.

4. Refer to the **Implied Volatility Surface** Python code under the page labelled "**Module 2: ML Applications in Finance (Week 3)**." Modify the neural network so that it has four hidden layers with the same number of neurons. Use a-one or more ReLU activation function(s) anywhere you wish. Does this modification affect the result (i.e., does the MSE increase or decrease)?