DSO 569: Homework 1

Due by the end of Saturday, March 30, 2024

1. Explain the difference between artificial intelligence, machine learning, and deep learning

2. Explain supervised learning and unsupervised learning. Give 3 application examples for each type of machine learning method. For each application, give examples of its corresponding loss function.

3. Explain underfitting and overfitting, and optimization and generalization.

4. We consider the same stock price data used in class. This data set contains the daily price of 470 stocks from Feb 8, 2013 – Feb 7, 2018. We are interested in predicting whether the return of ***Adobe Inc*** (stock symbol: ADBE) will be negative, low, or high using the remaining 469 stocks. Let us use the first 4 years (2/8/2013-2/7/2017) for training and the last year (after 2/8/2017) for testing. Here, a stock has a negative return if the return is negative, a low return if its return falls between 0 and 2%, and a high return if its return exceeds 2%. Follow the steps below to achieve this goal:
   a. Download the Python script from Blackboard used in class for analyzing the same data set. Use the code to convert the stock price data to return data. Modify the code to convert the continuous return to a categorical variable. In particular, the response variable takes the value 1 if negative, 2 if low, and 3 if high.
   b. Divide the data into training and test sets.
   c. Fit a multinomial regression model. The document below should be able to help. Note that the Python code I used in class works for the binary response, so you should not directly apply it here.
      https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html#sklearn.linear_model.LogisticRegression

   d. Use your model in step 3 to predict the value of y on both the training and test data. Note that the model you built returns a probability vector of dimension 3 specifying the probabilities of y=1,2,3 at each time point. For example, if on 1/1/2018, your model returns a probability vector (0.9,0.1,0), then you should set yhat=1 because the probability vector achieves its maximum value at location 1.

e. Calculate the prediction error on both training data and test data. Here, the prediction error is defined as the percentage of times that yhat differs from y.

Attach your script and relevant outputs from running your code.