Take Home Exam

Includes coding tasks and comments

Task 1 – The synthetic dataset

You have been provided with a synthetic dataset generated from a polynomial target function. You are also provided with a larger validation dataset generated from the same target function.

Your tasks are as follows:

Step 1: Train a linear model and collect the training and validation errors.

Step 2: Train a linear model using quadratic, 3rd order, and 4th order polynomial transforms and collect the training and validation errors.

Step 3: Plot the training and validation errors by order of polynomial transforms, i.e., linear, quadratic, 3rd, and 4th order.

Step 4: Comment on the pattern you observe. Also, comment on which of the models you would pick and why you would pick it.

Task 2 – Breast Cancer dataset

You have been provided with a preprocessed dataset with 8 features extracted from the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. The column labeled 'y' is the class variable containing two possible entries, 'M' for malignant and 'B' for benign cancer diagnosis. Your task is to find the best machine learning model to predict the cancer type based on the features. When splitting the dataset into training and testing datasets, please use the last three digits of your SJSU ID as the random_state. For example, if your SJSU ID is 11100111, use `train_test_split(X,y,random_state=111)`.

Include a detailed description of the methodology you used, the best performing model based on your experiments, why you chose it, and the performance metrics you used to evaluate your models.

Submission instructions

In your GitHub repo, create a branch called *take-home-exam*. Frequently commit your code and make sure it is shared with the instructor and ISA. Include a link to the GitHub repository in your submission pdf. Also include links to Google Colab notebooks if you used them.

<u>VERY IMPORTANT:</u> Include ALL the references you used for this exam, including names of classmates you discuss with. Failure to cite your sources counts as an act of academic dishonesty and will be taken seriously without zero tolerance. You will automatically receive a "fail" grade in the exam and further serious penalties may be imposed.

NOTE: You can look for help on the Internet but refrain from referencing too much. Please cite all your sources in your submission.

CMPE 257 Machine Learning

When you submit your exam, you automatically agree to the following statement. If you do not agree, it is your responsibility to provide the reason.

"*I affirm that I have neither given nor received unauthorized help in completing this exam. I am not aware of others receiving such help. I have cited all the sources in the solution file.*"

CMPE 257 Machine Learning