

Homework 2
Programming, Due 21:00, Wednesday, October 27, 2021

Late submission within 24 hours: score*0.9;
Late submission before post of solution: score*0.8 (the solution will usually be posted within a week); no late submission after the post of solution)

Total 120%

1. **(60%)** Consider the bias and variance example covered in the class. Suppose we now have a hypothesis set consisting of all horizontal lines $h(x) = b$. The input variable x is uniformly distributed in the interval $[-1, +1]$. The training data \mathcal{D} consists of only two points $\{x_1, x_2\}$. The target function $f(x) = \sin(\pi x)$. The data set is $\mathcal{D} = \{(x_1, \sin(\pi x_1)), (x_2, \sin(\pi x_2))\}$. The learning algorithm returns the line at the midpoint $b = \frac{\sin(\pi x_1) + \sin(\pi x_2)}{2}$ as $g^{(\mathcal{D})}$ (\mathcal{H} consists of functions of the form $h(x) = b$). Modify the code distributed in Hw2-1.ipynb template from Google Colab and record the bias and variance.

Pass these answers and call `output_csv` function to save your answers in `yourStudentId_p1.csv`.

2. **(60%)** Cross validation can be used to help us find the best hyperparameters in an algorithm. In this exercise, we will use an in-house *C&RT tree with random forest bagging algorithm* as a black box to predict whether the class of test data is 1 or -1. In order to find the best hyperparameter combination of this classification algorithm, you will do two parts following the instructions in Hw2-2.ipynb template from Google Colab.

(1) Please record your best tree number, best tree depth and test accuracy from the holdout validation. All the codes have been implemented. You will receive 40% of the total score. (2) Please implement a 3-fold cross validation, take the average of accuracy from 3 folds and observe which hyperparameter combination gives the best results. Record your best tree number, best tree depth and test accuracy from the 3-fold cross validation. You will receive 60% of the total score.

Pass these answers and call `output_csv` function to save your answers in `yourStudentId_p2.csv`.

- **Submission Format:** Please compress these two .csv files into `yourStudentId_hw2.zip`, then upload it to NTU COOL.