

CAP6610 Assignment 1 – End-to-End ML Projects

Total: 50 pts

Due Date: Sunday, 2/4/2024 11:59 PM

In this project, using the same dataset discussed in Chapter 2, you will experiment a few more regression models provided in the scikit-learn package. When ready to submit, zip your Assignment 1 directory containing assignment1.ipynb and questions.txt, and submit the zipped file to Canvas.

Steps:

1. From Canvas download assignment1.ipynb, a template that you start with for this assignment.
2. For all the models learned next, you will use these dataset variables:
housing_prepared (training X), *housing_labels* (training y), *X_test_prepared*, and *y_test*.
3. Create a new cell at the end of the notebook. In the new cell created, train a K-Nearest Neighbors (*KNeighborsRegressor*) model with n_neighbors=5, using 5-fold cross validation (*cross_val_score*), and print the square root MSE scores. **(10pts)**
4. Create another new cell at the end of the notebook. Train an Artificial Neural Networks (*MLPRegressor*) with model solver='lbfgs', hidden_layer_sizes=(15,), random_state=42, max_iter=5000, using 5-fold cross validation (*cross_val_score*), and print the square root MSE scores. **(10pts)**
5. Create another cell at the end. Use Grid Search (*GridSearchCV*) to search for the best hyperparameter among 'n_neighbors': [10,12,15], 'p': [1,2,3], 'algorithm': ['auto', 'ball_tree', 'kd_tree', 'brute'], using, again, 5-fold cross validation. **(10pts)**
6. Create last cell at the end. Pick the best K-Nearest Neighbors model from the previous Grid Search, and test it finally on *X_test_prepared* and print the square root MSE. **(10pts)**
7. From Canvas download questions.txt, and answer the two questions in the file. **(10pts)**

Note Steps 4 and 5 can be time consuming. If you'd like to see the time taken by running a cell, you may put `%time` at the top the cell.