Module 1 - Bonus Challenges

§M1: Introduction, Gradient Descent, Nearest Neighbors

- Below are open-ended bonus challenges; solving them is not required but can help you better understand ML/AI in the context of engineering, and how to use them in practical cases.
- Bonus points earned in all homework assignments will be averaged (6 bonus points for each assignment) and then directly added to your final score to calculate your final letter grade.

Challenge 1.1. As shown in Figure 1a, the property value v is given for a set of points (x,y), which reside on a circle with a radius r_1=1. The task is to predict the property values for unknown points with KNN regression (Figure 1b). Please use the built-in function `KNeighborsRegressor` in `sklearn` to achieve this task, following the instructions below: (*6pts*)

1. Use two nearest neighbours in the KNN.
2. Use the arc length on the circle as the distance metric.
3. Implement a customized weighted KNN. Assume the arc length distances from an unknown point to its two nearest neighbors are d_1 and d_2, respectively. The weights for the two neighbors are given as:

$$w_1 = \frac{d_2^2}{d_1^2 + d_2^2}, \quad w_2 = \frac{d_1^2}{d_1^2 + d_2^2}, \tag{1}$$

4. All the data are stored in the excel file '`m01_bonus.xlsx`'. The sheet '`Known`' contains the coordinates (x,y), and property values v for the known points. The coordinates for unknown points are given in the sheet '`Unknown`'.
5. Please submit your Jupyter notebook file with necessary comments and results visualization as in Figure 2. If your implementation is correct, you shoul dbe able to see similar results.Please also upload your predicted data in a separate csv file.

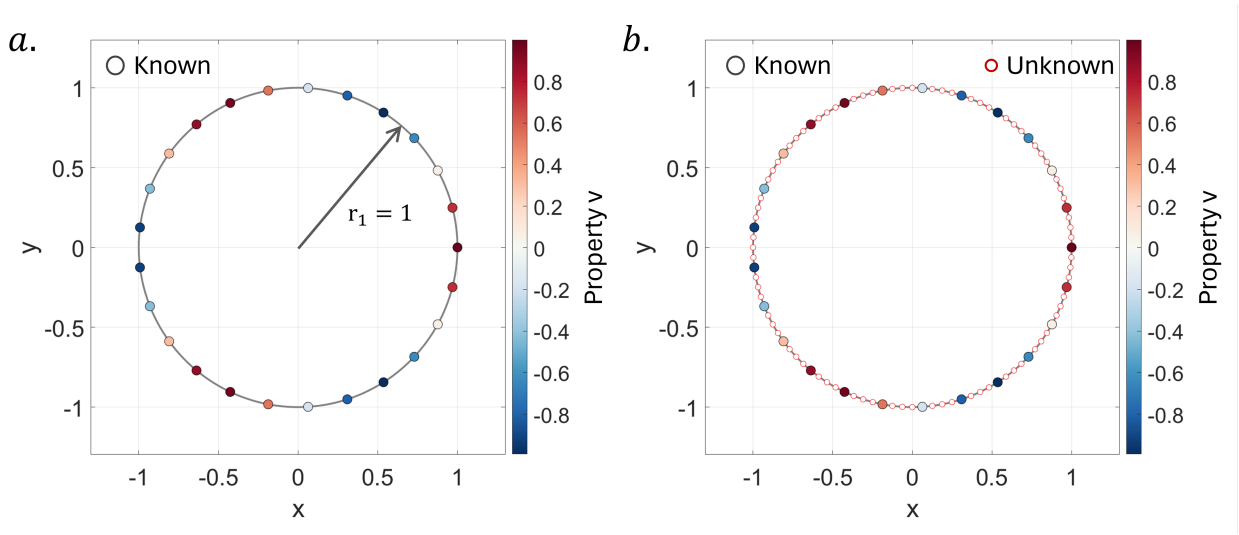

Figure 1: Problem setting for the KNN regression on a single circle

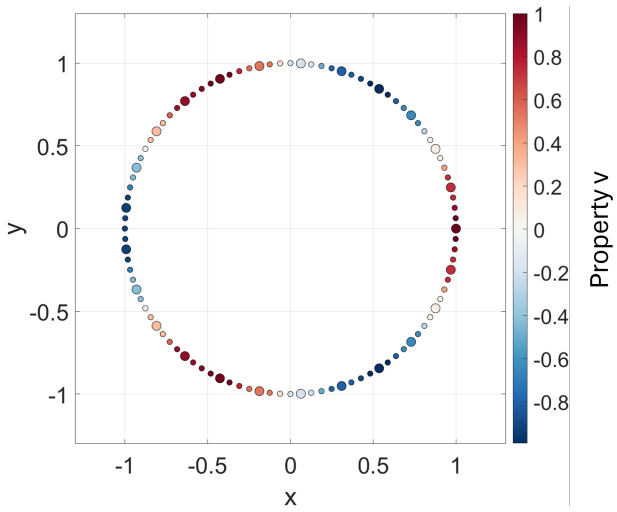

Figure 2: KNN predicted results

Tips: Please check the document for the built-in functions here: https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsRegressor.html

Note: When using customized weights, sklearn may not provide accurate predictions for batch data. Therefore, to ensure accuracy, you could iterate over each sample individually for prediction, rather than processing them as a batch.