SVM & Decision Trees - Evaluation Questions

1. Write a bare python function named OvR satisfying the following: **(40 mins, 10 marks)**

Inputs: a 2-d numpy array of size n x 2 named **data** containing floats

a list of length n named **label** containing values from the set {0, 1, 2}

a 1-d numpy array of length n named **value** containing floats

Operation: The One vs Rest training and prediction.

i.e In each iteration,

values of **data** associated with each of the **label** from two of the **label**s will be averaged and stored.

Then, for every **data** associated with the remaining **label**, based on to which of the **data** associated with the two **label**s in previous step it is close to, its value will be predicted as the corresponding average already computed and stored.

Note: Every label and its associated data must be part of prediction, and only once.

Output: Print the given **value** and the predicted value for each of the data

To test the function, you are encouraged to use the following data:

$$\textbf{data} = \begin{pmatrix} 1 & -1 \\ 2 & 1 \\ -1 & -.5 \\ -2 & .5 \\ 0 & 2 \\ 0 & -2 \\ .5 & -1 \\ -.5 & 2 \\ 1 & 1 \\ -1 & 3 \end{pmatrix} \qquad \textbf{label} = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \\ 1 \\ 2 \\ 0 \\ 1 \\ 2 \\ 1 \end{pmatrix} \qquad \textbf{value} = \begin{pmatrix} 0.1 \\ 1.1 \\ 2.5 \\ 0.5 \\ 1.6 \\ 2.4 \\ 0.3 \\ .41 \\ 2.9 \\ 1 \end{pmatrix}$$

The code will be in pure python with the only exception of using numpy.

You may write intermediate functions.

You may refer to the NumPy documentation if needed, but at this level, it shouldn't be necessary for solving the given problem since the requirement of NumPy for the given problem is basic. However, the choice is yours.

2. In the decision tree exercise, you have got a best estimator based on grid search, and you tested this estimator on the test set and printed the performance measures. Now, you print for each sample in the test set, the leaf index associated with its prediction. You may go through the DecisionTreeClassifier documentation. **(25 mins, 7 marks)**

3. Speak about your attempt at SVM and decision tree models in 2 minutes. **(3 marks)**