Hi everyone, my name is An. Today, I’m going to share my work for assignment two with you.

Slide 2:

Let me start with a dataset. The original dataset was data used by Candanedo et al. in 2017.

In this assignment, I choose four hundred random data to predict. And here are the details for each column.

Except for the first column, the following five columns are features X, and the last column is output Y. Here are their details and my code to extract them.

Slide 3-4-5-6-7:

After extracting the data, I wrote code to draw a scatter plot of each variable X with Y and histograms. It is hard to identify a relationship between X and Y, so I wrote a code to calculate the correlation between them.

Most variables X have a positive relationship with Y, except X2 and X5 have a negative relationship with Y.

Based on histograms, I can identify a positive skew for X1, negative skew for X2, positive skew for X3, positive skew for X4, negative skew for X5, and positive skew for Y

Slide 8 9 10

Based on the correlation result from task one, I chose X1, X2, X3, and X4 to predict because they have a stronger relationship with Y than X5

To reduce skewness,

I used log transformation for X1, X3, X4 and Y because they have a positive skew.

I apply a third power for only X2 because it has a negative relationship.

Then, I applied min-max normalisation to all of them on a scale from 0 to 1.

As X2 has a negative relationship with Y, I applied a negation function for it.

So here is the histogram of them after transforming.

The skewness has been reduced, and all data are between 0 and 1

Slide 11

After transforming the data, I ran the code to find the best model. Here is the table for error measures. According to the table, OWA has the lowest RMSE and average absolute. It also has the lowest Pearson and Spearman correlation compared with others, but the difference is not too much. It indicates that OWA has the smallest error here, so it should be chosen to predict task 4

Slide 12

Here is the table of weights for each model.

The OWA function sorts input from low value to high value for the importance of variables. Then, the weight of the OWA is used for the sorted input.

Slide 13

After finding the best model, in task 4, I assigned the input and OWA weight.

I transformed input as the same pre-process in task 4, including the third power for the second input and log transformation for others.

Then, I use min-max normalisation to make them on the same scale from 0 to 1.

After normalisation, I apply the negation function for the second input and use the OWA function to predict.

I reversed the transformation back from min-max normalisation and log transformation to make the predicted output on the same scale as the expected output

So I got sixty-five point seventy-one, greater than the expected output of nine point fifty two per cent.

It is a reasonable output because the difference is smaller than ten per cent.

Slide 14:

The best condition for a low energy use of appliances is

* Temperature in living room area should be low (because the correlation is positive)
* Humidity in living room area should be high (because the correlation is negative)
* Temperature in the office room should be low (because the correlation is positive)
* Humidity in office room should be low (because the correlation is positive)

Slide 15

For implications of OWA:

Implications:

We can flexibly favour higher or lower inputs according to preference on defined function.

Then, The order of inputs does not affect output because the OWA function will sort the order based on our preference.

Limitations:

The OWA can be affected by outliers. For example, If input values have outliers, the outliers can be sorted at the beginning or end of the input range. It may affect the weight.