Part A: Analysis of Energy Efficiency Dataset for Buildings
Description:

In order to design energy efficient buildings, the computation of the Heating Load (HL) and the Cooling Load (CL) is required to determine the specifications of the heating and cooling equipment needed to maintain comfortable indoor air conditions. Energy simulation tools are widely used to analyse or forecast building energy consumption. The Dataset provides energy analysis of Heating Load (denoted as Y1) and the Cooling Load (denoted as Y2) using 768 building shapes that are simulated using a building simulator. Select one of Y1 or Y2 as your variable of interest and focus the analysis on this variable. The dataset comprises 5 features (variables), which are denoted as X1, X2, X3,X4,X5. The description of the variables is given below:

X1: Relative compactness in percentage (expressed in decimals) - A measure of building compactness. A high value means highly compact.

X2: Surface area in square metres

X3: Wall area in square metres

X4: Roof area in square metres

X5: Overall height in metres

Y1: Heating load in $kWh.m^{-2}$ per annum

Y2: Cooling load in $kWh.m^{-2}$ per annum

Tasks:

1. Understand the data [**10 marks**]

   (i) Download the txt file (ENB18data.txt) from CloudDeakin and save it to your R working directory.

   (ii) Assign the data to a matrix, e.g. using

   ```
   the.data <- as.matrix(read.table("ENB18data.txt"))
   ```

   (iii) Decide whether you would like to investigate Heating Load (Y1) or Cooling Load (Y2). This is your *variable of interest*. Generate a subset of 300 data, e.g. using:

   To investigate Heating Load Y1:

   ```
   my.data <- the.data[sample(1:768,300),c(1:5,6)]
   ```

   To investigate Cooling Load Y2:

```
my.data <- the.data[sample(1:768,300),c(1:5,7)]
```

(iv) Using scatterplots and histograms, report on the general relationship between each of the variables X1,X2, X3, X4 and X5 and your variable of interest Y1 (heating load) or Y2 (cooling load). Include a scatter plot for each of the variables X1, X2, X3, X4, X5 and your variable of interest Y1 or Y2. Include a histogram for X1,X2,...,X5, and Y1 or Y2. Include 1 or 2 sentences about the relationships and distributions.

2. Transform the data **[15 marks]**

(i) Choose any **four** from the first five variables X1,X2,X3,X4,X5.

Make appropriate transformations to the variables (including Y1 or Y2) so that the values can be aggregated in order to predict the variable of interest (your selected Heating Load Y1, or cooling load Y2). The transformations should reflect the general relationship between each of the four variables and the variable of interest. Assign your transformed data along with your transformed variable of interest to an array (it should be 300 rows and 5 columns). Save it to a txt file titled "name-transformed.txt" using

```
write.table(your.data,"name-transformed.txt",)
```

(ii) Briefly explain each transformation for your selected variables and the variable of interest Y1 or Y2. (1- 2 sentences each).

3. Build models and investigate the importance of each variable. **[15 marks]**

(i) Download the AggWaFit.R file (from CloudDeakin) to your working directory and load into the R workspace using,

```
source("AggWaFit718.R")
```

(ii) Use the fitting functions to learn the parameters for

- Weighted arithmetic mean (WAM),
- Weighted power means (PM) with $p = 0.5$, and $p = 2$,
- Ordered weighted averaging function (OWA), and
- Choquet integral.

(iii) Include two tables in your report - one with the error measures (RMSE, Av.abs error, Pearson correlation, Spearman correlation) and one summarising the weights/parameters that were learned for your data.

(iv) Compare and interpret the data in your tables. Be sure to comment on:
(a) How good the model is,
(b) The importance of each of the variables (the four variables that you have selected),
(c) Any interaction between any of those variables (are they complementary or redundant?)
(d) better models favour higher or lower inputs (1-2 paragraphs for part (iv)).

4. Use your model for prediction. **[10 marks]**

   (i) Using your best fitting model, predict the Heating Load Y1 or the Cooling Load Y2 for the following input:

   X1=0.82, X2=612.5, X3=318.5, X4=147, X5=7.

   Give your result and comment on whether you think it is reasonable. (1-2 sentences)

   (ii) Comment generally on the ideal conditions (in terms of your 4 variables) under which a low heating or cooling load will occur. (1-2 sentences)

For this part, your submission should include:

1. A report (created in any word processor), covering all of the items in above. With plots and tables it should only be 2 - 3 pages.
2. A data file named "name-transformed.txt" (where 'name' is replaced with your name - you can use your surname or first name - just to help us distinguish them!).
3. R code file, (that you have written to produce your results) named "name-code.R", where name is your name;