Statistical Modelling & Machine Learning HW2
(Due: 11/01/2020, Sunday)

Instruction:

- There is no correct or unique answer in this homework.
- I will give your HW score based on your results and model building procedure.
- Submit your HW solution with `R` code.
- Your `R` code should show the procedure that you obtain the final model (**Do NOT include all models that you have tried**).

1. Consider the data in '`house.csv`' file. The our goal is to build the best model predicting house price based on some factors in a city. The description of the variables in the data as follows:
`date`: House transaction date (e.g., 2012.25 = 2012 March).
`age`: House age (unit: year).
`dist`: Distance to the nearest subway station (unit: meter).
`store`: The number of convenience stores in the living circle (unit: integer).
`lat`: The geographic coordinate, latitude.
`lon`: The geographic coordinate, longitude.
`price`: House price per $3.3m^2$. Consider `price` as an output variable.

(1) Using the data modelling techniques (parametric modelling), build your model and fit it to the data for attaining the lowest AIC value. To compute AIC in `R`, use `AIC` built-in function.

(2) Based on the model obtained from part (1), interpret the relationship between house price and each input variable as detail as possible.

2. Consider the training data in '`pm25_tr.csv`' and the test data in '`pm25_te.csv`'. Suppose that our interest is to predict pm 2.5 concentration based on some meteorological factors. In the dataset, the output variable is `pm25`. The training set has data measured from March, 1st to May, 20th and the test set has data measured from May 21st to May 25th (next 5 days). The descriptions of the variables in the dataset are as follows:
`month`: Month of data.
`day`: Day of data.
`hour`: Hour of data.
`pm25`: PM2.5 concentration.
`DEWP`: Dew Point.
`TEMP`: Temperature.
`PRES`: Pressure.

`cbwd`: Combined wind direction.
`Iws`: Cumulated wind speed.

(1) Using the data modelling techniques (parametric modelling), build your best prediction model from the training data [**NOTE:** You might need the transformation of variables or variable selection].

(2) Compute the test MSE of your model obtained in part (1) using the test set.