Pre-class Learning Activities for Session 6

- Complete the following learning activities (listed below) before class.
- References are optional and not meant to be read from first page to last. Refer if necessary.
- You may be asked to volunteer/present your work for class participation points.
- If you presented your work in class, remember to email your work to instructor after class.
- You may be asked to provide comments/add-ons to work presented by students.

Learning Activities:

1. In NTULearn Main Site, watch Lecture Videos S6.1 and S6.2.
2. Activity 1:
    i. Run RScript flatsales-mars.R
    ii. What is the MARS model coefficients and RMSE if 10-fold CV is used to prune instead of GRsq? Which of the following two ncross levels is more stable?
        - Set Seed = 2 vs 2020
        - pmethod="cv"
        - nfold = 10
        - ncross = 1 vs 5
    iii. Create a copy of the 5-room flat sales dataset as an Excel workbook. Using Excel, show that the linear regression model with the selected 4 hinge functions from R output (see below screenshot) has the same model coefficients as R output.
    iv. Advanced option: Compute GCV, GCV.null and GRSq in excel, and verify if the numbers match R output below.

```
> summary(m.mars1)
Call: earth(formula=Sales.5rm~t, data=data.sales, degree=1)

             coefficients
(Intercept)    3994.2043
h(t-11)        -172.7576
h(t-26)         105.0363
h(32-t)         -77.5735
h(t-32)          96.0591

Selected 5 of 6 terms, and 1 of 1 predictors
Termination condition: RSq changed by less than 0.001 at 6 terms
Importance: t
Number of terms at each degree of interaction: 1 4 (additive model)
GCV 109300    RSS 3886456    GRSq 0.5730591    RSq 0.696496
```

3. Activity 2: In Chang et. al. (2012)[1] Analysis of freeway accident frequency using multivariate adaptive regression splines p. 827 states "…(5) if the degree of horizontal curve is greater than 8.20, the accident frequency will increase by 0.992 for additional increase in degree of horizontal curve (indicated by BF7)."

    Does this mean that if horizontal curve degree = 9.2, then accident frequency will increase by 0.992?

---

[1] PDF article provided in the NTULearn content folder.

R References:

- Section 9.4 of textbook (free PDF): https://web.stanford.edu/~hastie/ElemStatLearn/
- MARS example: http://uc-r.github.io/mars
- Rpackage earth: https://cran.r-project.org/web/packages/earth/index.html

Python References:

- scikit-learn Py-earth: https://contrib.scikit-learn.org/py-earth/
- MARS Example: https://machinelearningmastery.com/multivariate-adaptive-regression-splines-mars-in-python/