

Coding Assignment 2

Due Monday, October 11

Part I: Select optimal span for loess (2pt)

Write your own functions to use LOO-CV and GCV to select the optimal span for `loess`.

- Test your code on data set `Coding3.Data.csv`, which can be downloaded from Campuswire.
- Report your CV and GCV for 15 span values: 0.20, 0.25, ..., 0.90.
- Report the optimal value(s) for span based on CV and GCV.
- Plot the fitted curve(s) using the optimal value(s) for span.

Part II: Clustering time series (3pt)

Download the `Sales_Transactions_Dataset_Weekly` dataset from UCI Machine Learning Repository Link. This dataset contains weekly purchased quantities of 811 products over 52 weeks.

1. The feature for each product is a time series with 52 measurements. Remove the mean from each time series and store the data as an 811-by-52 matrix **X**.
2. Fit each time series with a NCS with $df = 10$, which is equivalent to a NCS with 8 interior knots. Save the corresponding coefficients as an 811-by-9 matrix **B**.
3. Run k-means algorithm on **B** to cluster the 811 products into 6 clusters. Display the time series for products in the same cluster in one figure along with the cluster center; arrange the 6 figures in 2-by-3 format.
4. Run k-means algorithm on **X** to cluster the 811 products into 6 clusters. Display the time series for products in the same cluster in one figure along with the cluster center; arrange the 6 figures in 2-by-3 format.

What you need to submit?

An R Markdown file in HTML format, which should contain all code used to produce your results.

Name your file starting with `Assignment_3_xxxx_netID` where “xxxx” is the last 4-dig of your University ID.