

HW7

Part-1:

Random forest approach for classification.

Use the same dataset that you used for HW5 for the classification

(It probably won't perform well; say why.)

Part-2:

Introduce 5 new features based on the citation numbers. and use them in the RF instead of the citation numbers directly.

Each new feature is:

$((\text{citation number in year } n+1) - (\text{citation number in year } n)) / (\text{citation number in year } n)$
for $2016 < n < 2022$.

Due: 11/25/2023