HW4 Feature Selection

Using the SAT dataset ('CASE1201.ASC.txt')

Create a .ipynb notebook to submit your assignment. The file name should start with “HW4\_yourwsuid”.

Build a simple (intercept + takers) and a full linear regression model to predict the **SAT scores**. (that is, simply copy the first two models from the SAT\_example.ipynb presented in class.)

Choose (at least) one method from each of the three feature selection techniques (filter, wrapper, and embedded). Select (and report) the “best” features using these three methods.

(extra credit) Use PCA to reduce the dimensionality of the model to 2

Compare the different models with each other. Report the results in a cell that includes (at least) R^2, F1 and P(F1) for each model (keep it simple).

Submit the .ipynb for your assignment.