uc3m | Universidad Carlos III de Madrid

OPENCOURSEWARE
ADVANCED PROGRAMMING
STATISTICS FOR DATA SCIENCE

Ricardo Aler

ADVANCED PROGRAMMING
MASTER IN STATISTICS FOR DATA SCIENCE.
FOURTH ASSIGNMENT: PYTHON PLOTTING AND SCIKIT-LEARN PIPELINES

0.5 POINTS

Introduction

The aim of this assignment is to use the dataset of the previous assignment for practicing with pipelines and plotting:

What to do

In this case, you have to use all meteorological attributes (not only those at location 13/Sotavento). For model evaluation, we will use exactly the same holdout approach than in the previous assignment.

1. Use a Pipeline of feature_selection+decision_tree and Grid-Search to determine the optimal number of attributes. Hyper-parameters of the decision tree should also be tuned. Evaluate the final model on the test set and determine whether this result is better than the one you obtained in the previous assignment
2. Use Matplotlib or SeaBorn to obtain a plot similar to the one below. That plot is used to compare the prediction of the model (power prediction) with the actual power (ground truth). You can use either the best model you obtained in the previous assignment or the model you are obtaining in this assignment.