

H/W #2:

Briefly describe the mechanisms of operation of four different data imputation methods.

Using the iris dataset, randomly create missing values to occupy x% of the data. Use R to apply 3 different methods of imputation and compare their performance when x=2, 5, 10, 15, 20, 25 (All code snippets must be included in the report appendices and on github). Use the following methods to evaluate the performance of the methods:

Root mean square error (RMSE) between imputed and true values

Supervised classification error (Use k-NN classifier)