Assignment II.

1. Try different imputation methods on the Titanic Dataset, and evaluate classifier accuracies for each of these. A package you can use is fancyimpute.

2. Briefly describe how gradient boosting differs from bagging. Implement gradient boosting as invoked in scikit-learn, and evaluate classifier accuracy for the Titanic dataset

3. Theoretically, increasing the number of decision trees (n\_estimators?), increases classifier performance and/or generalizability. How would you design and evaluate a computational experiment to test this, on the Titanic dataset. Range() n\_estimator

4. Pick any Kaggle regression dataset. Train, tune and evaluate performance of a Random

Forest Regression model. How will you use the feature importance calculations from this

to perform feature selection. Please demonstrate this using the Kaggle regression dataset

you picked.

Module7 part4 bottom