

## Cross Validation/Regularization Homework

CMPE 257

Instructor: Jahan Ghofraniha

1. Perform full EDA on fish dataset (FishDataset.csv provided under ML data bank on Canvas). Assume the **weight** column to be the output and the rest of columns as inputs. Column zero (categorical data) should be converted to numbers using encoding (use `preprocessing.LabelEncoder()` from sklearn library).
2. For step 3, perform standardization on the data before performing cross-validation.
3. Cross-validate the following models using the sample code provided on Canvas (`crossValidation_regression.py` and `mlr_lasso_ridge.py`) as a reference. You should cross-validate all the models together in one run under the same condition:
  - Linear regression with 6, 5 and 4 features (you should use RFE to find out top ranking features).
  - Regularized version of all the linear models (RFE with 6, 5 and 4) features with Lasso and Ridge regularization. Remember, you need to find the right value of  $\alpha = \text{Lambda}$ . These variations (different lambda values) should be included as part of the cross validation.
4. Include your code, the results and explanation of the results either as a Jupyter notebook file (.ipynb) or a (.py) plus the output/results as comments in your code or as screenshots if it involves plots/graphics and upload to Canvas under cross-validation assignment.