CSE 590 **Introduction to Machine Learning** Spring 2023

**Homework No. 1**
**Due Jan. 25 (11:00 am), 2023**

**Objectives**

1. *Build and analyze simple regression algorithms based on KNN and linear models*
2. *Identify cases of underfitting and overfitting*
3. *Select parameters that optimize performance (generalization)*

**Problem #1**

For this problem, you will use the *Wine Quality* database (posted in Blackboard). Use the provided training data subset to train your model and the testing subset to predict and analyze your results.

a) Build and train a KNN Regression model. Vary the parameter K and analyze the results by identifying cases of **overfitting** and **underfitting**. Select the optimal value of K and justify your choice.

b) Build and train an OLS Regression model. Analyze the results and indicate if the learned model is a good choice for this data. Justify your conclusions.

c) Build and train a Ridge Regression model. Vary the constraint parameter $\alpha$ and analyze the results by identifying cases of **overfitting** and **underfitting**. Select the optimal value of $\alpha$ and justify your choice.

d) Build and train a LASSO Regression model. Vary the constraint parameter $\alpha$ and analyze the results by identifying cases of **overfitting** and **underfitting**. Select the optimal value of $\alpha$ and justify your choice.

**What to submit?**

- A report that
  - **Describes** your experiments,
  - **Summarizes**, **explains** (using concepts covered in lectures) and **compares** the results (using plots, tables, figures)
  - Identifies the best method for each dataset.
- Do not submit your source code
- Do not submit raw output generated by your code!
- Your report needs to be a single file (MS Word or PDF)
- Your report cannot exceed 10 pages using a font of 12
- Assign numbers to all your figures/tables/plots and use these numbers to reference them in your discussion