Title: Penalized regression model and its application I

High-dimensional data typically describes datasets with a greater number of predictors or a larger number of features in comparison to the number of observations or samples. When handling this type of data, statistical models tend to suffer from overfitting (to the training data) which may lead to poor prediction results in a new dataset. Penalized regression model, or sometimes called shrinkage or regularization, typically refers to a set of methods to address the overfitting issue by adding a penalty term to the model's error function.

In this project, students are going to study three basic penalized regression models including ridge regression, lasso, and elastic net from both theoretical and practical perspectives. Specifically, for model implementation, students are expected to

• determine a specific topic that can be studied by analysing some real-world high dimensional data

• apply the abovementioned methods to the dataset using programming language such as R or Python

• give sound interpretation of results, and

• compare the results of different model