This task is about doing linear regression on transformed features, where the features are given and fixed. In our coding, we first load the training data from “train.csv”. Then we compute the feature transformation according to the given Linear, Quadratic, Exponential, Cosine and Constant features. Finally, we build our linear regression model, do the training and learn the coefficients that linearly combine the features in order to fit the training targets y, while maintaining generalization ability to test dataset.

To achieve these two goals, i.e. fitting the training data well and also generalizing well, we decide to choose the RidgeCV model from the sklearn module. It is basically ridge regression with built-in Cross-Validation. We can set multiple regularization hyperparameter values (here, after several trials, we set alphas = [0.01,1,100]), and the algorithm would try and choose the best value among them which produces the best estimations on validation sets. For “fit\_intercept”, we set it to be false because we already have a constant feature. At last, we set the hyperparameter cv=7, which means we use 7-fold Cross-Validation. During trials of this cv value, we find the final performance barely changes when we vary cv from 5 to 50. So, we just tune it to keep the number of data points in each fold to be in a reasonable range.