**1.**

**Linear regression with various regularizers:**The UCI Machine Learning dataset repository hosts a dataset giving features of music, and the latitude and longitude from which that music originates [here](https://archive.ics.uci.edu/ml/datasets/Geographical+Original+of+Music). Investigate methods to predict latitude and longitude from these features, as below. There are actually two versions of this dataset. Either one is OK by me, but I think you'll find the one with more independent variables more interesting. You should ignore outliers (by this I mean you should ignore the whole question; do not try to deal with them). You should regard latitude and longitude as entirely independent.

* First, build a straightforward linear regression of latitude (resp. longitude) against features. What is the R-squared? Plot a graph evaluating each regression.
* Does a Box-Cox transformation improve the regressions? Notice that the dependent variable has some negative values, which Box-Cox doesn't like. You can deal with this by remembering that these are angles, so you get to choose the origin. why do you say so? For the rest of the exercise, use the transformation if it does improve things, otherwise, use the raw data.
* Use glmnet to produce:
* A regression regularized by L2 (equivalently, a ridge regression). You should estimate the regularization coefficient that produces the minimum error. Is the regularized regression better than the unregularized regression?
* A regression regularized by L1 (equivalently, a lasso regression). You should estimate the regularization coefficient that produces the minimum error. How many variables are used by this regression? Is the regularized regression better than the unregularized regression?
* A regression regularized by elastic net (equivalently, a regression regularized by a convex combination of L1 and L2). Try three values of alpha, the weight setting how big L1 and L2 are. You should estimate the regularization coefficient that produces the minimum error. How many variables are used by this regression? Is the regularized regression better than the unregularized regression?

Glment() function is used on raw data in order to to predict latitude and longitude from music features. According to the documentation of the cv.glmnet() function, the fuction does a 10-fold cross validiation by splitting the data into 10 sets, fitting the moder on 9 sets and evaluating on the one, repaeating this process 10 times, changing the sets each round. Therefore, the MSE values are obtained by fitted model being tested on one fold of data not used for training the model.

Two sets of linear regression models are fitted for latitude and longtitude vs. 166 features. A set of linear regression models fitted for each case include one cross validated unregurized linear model and five cross validated regulrized linear model: ridge (alpha = 0), lasso (alpha = 1), elastic net (alpha = 0.25) , elastic net (alpha = 0.50) , elastic net (alpha = 0.75). For being able to compare the results of the models with each other, all the models are fitted using the cv.glmnet() function and means squared errors (MSE) reported by the function are obtained and analysed. In case of the cross validated unregurized linear model, the lambda and alpha coeeficents are set to zero for cv.glmnet() function which according to the document of function, will cause function to perform as an unregularized linear model with no regularizer penalty. Table 1 and 2 show the MSE results obtained by fitting a set of linear models for latitude and longtitude labels, respectively.

Table 1. MSE results obtained by fitting linear models for latitude vs. 116-feature matrix

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Latitude | | | | | |
| Unregularized | Regularized - ridge | Regularized - lasso | Regularized - elastic net | Regularized - elastic net | Regularized - elastic net |
| Best Lambda | 0 | 5.78 | 0.458 | 1.83 | 0.916 | 0.611 |
| Alpha | 0 | 0 | 1 | 0.25 | 0.5 | 0.75 |
| MSE | 285 | 280 | 279 | 280 | 280 | 277 |
| Non-zero Variables | 116 | 116 | 21 | 34 | 22 | 21 |

Table 2. MSE results obtained by fitting linear models for longtitude vs. 116-feature matrix

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Longtitude | | | | | |
| Unregularized | Regularized - ridge | Regularized - lasso | Regularized - elastic net | Regularized - elastic net | Regularized - elastic net |
| Best Lambda | 0 | 4.5 | 0.27 | 1.43 | 0.539 | 0.327 |
| Alpha | 0 | 0 | 1 | 0.25 | 0.5 | 0.75 |
| MSE | 1929 | 1873 | 1873 | 1887 | 1882 | 1851 |
| Non-zero Variables | 116 | 116 | 58 | 97 | 88 | 86 |

In case of linear regression regularized by L2 (equivalently, a ridge regression), the best regularization coefficient Lambda for latitude and longtitude values are 5.78 and 4.5 producing a minimum MSE values of 280 and 1873 (see Table 1 and 2). Comparing the regularized ridge regression with unregularized linear model, it can be concluded that ridge regularization is improving the accuracy of the model.

For linear regression regularized by L1 (equivalently, a lasso regression), the best accuracies are obtained with regularization coefficients of Lambda 0.458 and 0.27 along with alpha of 1 for latitude and longtitude, respectively. Obtained minimum MSE values corresponding to these regularization coefficients are 279 and 1873 for latitude and longtitude, respectively (see Table 1 and 2). Comparing the regularized lasso regression with unregularized linear model, it can be concluded that lasso regularized regression is better than the unregularized ones. In case of lasso regularized regression, 21 and 58 non-zero variables are used for latitude and longtitude cases, respectively.

Finaly, in regard to linear regression regularized by elastic net (equivalently, a regression regularized by a convex combination of L1 and L2) three alpha values of 0.25, 0.5 and 0.75 are used. The regularization coefficient that produces the minimum error are provided in Table a and 2 for three various values of alpha along with number of variables used by each regression model. Comparing the MSE values, it can be conluded that the best regression model is obtained by using an elastic net regularization technique and this methods definitely is better that the unregularized regression model.