Exercise 4: Part 4

Logistic Regression

Python Programming Bootcamp by Dr Rohitash Chandra
UNSW, 2021

Description

Apply logistic regression to the Pima-Indian diabetes dataset:
https://www.kaggle.com/kumargh/pimaindiansdiabetescsv

You can either use logistic regression from scratch code or use scikit-learn library.

Do the following:

1. 60/40 train test spilt and report training and test performance in terms of log-loss, classification performance, and RMSE
2. Report AUC and ROC and Precision-Recall curve, F1 Score.
3. Try L1/L2 and elastic net regularization and compare results.
4. Carry out 10-fold cross-validation on the original data-set.
5. Select any binary classification problem of your choice from UCI machine learning repository and carry out steps 1-4.