

Week 6 Practice Problems

Logistic Regression

The practice problems for this week will allow you to practice implementing a logistic regression model:

1. Import the typical libraries needed for ML and data analysis such as numpy, pyplot, and pandas.
 - Read the CreditCard.csv dataset provided. The dataset can also be obtained from:
<https://www.kaggle.com/mlg-ulb/creditcardfraud>
 - Display last 7 records of the dataset.
2. Check to see if the classes in the data are imbalanced. Do this by plotting count of the values in the target class and displaying count of each values in the target variable as number to the screen.
3. Normalize the data using standard deviation with SK Learn
4. Separate the data into features and target variable, and convert from dataframe to numpy arrays
5. Since the classes are imbalanced:
 - Use numpy's random method to generate K random indices of non-fraudulent classes. Here K is the total number for transactions with class labeled as fraudulent.
 - Use the generated indices to filter through the data such that you end up with equal percentage of fraudulent and non-fraudulent classes.
 - Show the percentage of each class as well as the total number of transactions (records) in the new dataset.
6. Split your data into training and testing set
7. Use Logistic Regression algorithm to train a model, and make predictions on the testing set created in 6.
8. Use evaluation metrics such as confusion matrix, classification report and accuracy score to show performance of your classifier.

Week 6 Practice Problems Solution: Jupyter Notebook

(Jupyter%20Notebook/Week%206%20Practice%20Solutions.ipynb?_&d2lSessionVal=A3zrhW82FAbprEN6UpM5B79gC)