Classification and Regression Perceptron

Exercise A: Classification

You are asked to design and implement a perceptron classifier that determines whether a bank transaction is genuine or fraudulent.

The file 'm_creditcard.csv' contains a modified part of the Kaggle dataset available from: https://www.kaggle.com/dalpozz/creditcardfraud/data

Each row in the file represents a card transaction. The set of features that describe the transaction are the columns V2 and V11, followed by the target class of the transaction: 0 for genuine and 1 for fraud.

Please note that in this dataset; the features have been manipulated using the PCA technique to anonymize the dataset guaranteeing the user and bank privacy. PCA keeps the relative importance of the features; hence you will use these features in the same way as normal features.

This dataset is not strictly linearly-separable, hence you will need to introduce an error tolerance value for the perceptron to converge.

Requirements and Grading Criteria

1. Perceptron Design and implementation. 2%
2. Train the perceptron using the included dataset. Repeat the training process for training rates 0.2 and 0.8. Comment on the number of iterations it takes for the perceptron to converge in each case. 2%
3. Plot the decision boundary with respect to V2 and V11 at the end of the training process. 1%
4. Documentation, code quality, demo and discussion with team members. 1%

For this assignment, you may use Numpy and Pandas. You're asked to write the perceptron yourself, rather than use library versions such as the Scikit perceptron.

While you may use online references, only your own code may be submitted.

Exercise B: Regression

Modify the perceptron you implemented in part 1 to predict the time a flight will take based on its distance. For this exercise you will be using flight data collected in 2015 for several American airlines.

Please note that some of the records in the dataset have missing values. You should use the dataframe.dropna() method to drop these records.

Divide the dataset you're assigned into parts of size 80% and 20% for training and testing respectively. Calculate the mean square error over the test set.

Each team will be assigned a particular airline to work on. Please check the accompanying file for a list of teams and corresponding airline. You can download your assigned dataset from

https://drive.google.com/open?id=0B7k61JKDD_vdOE1mSV85eTYtWUE

Requirements and Grading Criteria

1. Design and implementation modifications 1%
2. Training and testing. Comment on the number of iterations it takes for the perceptron to converge. 2%
3. Documentation, code quality, demo and discussion with team members. 1%

Submission

Please submit your python code files and pdf report via email to mla.w17.guc@gmail.com by 11:59PM Tuesday 10th October 2017.

Late submissions will not be accepted.