DIGITAL STACK
Knowledge. Performance. Growth.

powered by Bloomberg

How to Use **Python** and **Machine Learning** for **Financial Fraud Detection:** from Zero to Hero

Corina SARARU

community partner Women Techmakers

CONTENTS

1.Introduction and Objectives of Workshop

2.Classic Machine Learning Algorithms Used in Financial Fraud Detection: Intuitive description -no math!-

3.Introduction in Python: Motivation and Basic Syntax

4.Some Python Libraries for Machine Learning

5.Coding Exercise: Financial Fraud Detection with Python Libraries

6.Conclusions

1. INTRODUCTION AND OBJECTIVES OF WORKSHOP

➢ **INTRODUCTION**

In this workshop, **Python** and **Machine Learning** will meet to demonstrate solving a case of **Financial Fraud Detection**.

➢ **OBJECTIVES:**

- **The Good:**
  - ✓ What are some classic Machine Learning algorithms?

- **The Bad** [in a good way]**:**
  - ✓ What makes Python so popular? What are the strengths it brings to the market?

- **The** [not so] **Ugly:**
  - ✓ How does a problem like Financial Fraud Detection get solved using Machine Learning in Python?

2. CLASSIC MACHINE LEARNING ALGORITHMS USED IN FINANCIAL FRAUD DETECTION: INTUITIVE DESCRIPTION -NO MATH!-

Financial fraud detection, in particular credit card fraud can be explained this way:

- Either a credit card transaction is legitimate (this is one class of transactions)
- Or a credit card transaction is fraudulent (this is another class of transactions)

Objective: determine automatically for every transaction, **if it is or not fraudulent.**

More generically:

- Given two sets of blue and red dots**, the objective** is to determine their color **– by separating them**.

2. CLASSIC MACHINE LEARNING ALGORITHMS USED IN FINANCIAL FRAUD DETECTION: INTUITIVE DESCRIPTION -NO MATH!-

In this workshop, three supervised learning classifiers will be used (among the many available on the market today):

A. Perceptron classifier: can be trained to draw a line to separate the dots

More info here: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Perceptron.html

B. Multi Layer Perceptron (MLP) classifier: can be trained to draw a curve to separate the dots

More info here: https://scikit-learn.org/stable/modules/neural_networks_supervised.html

C. Support Vector Machine (SVM) classifier: can take things to the next level. Of detail. Literally.

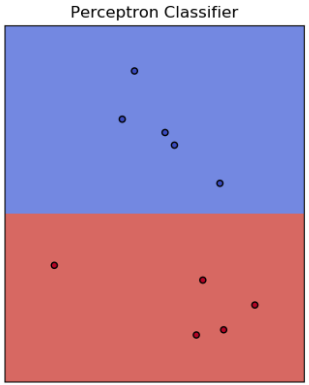
More info here:

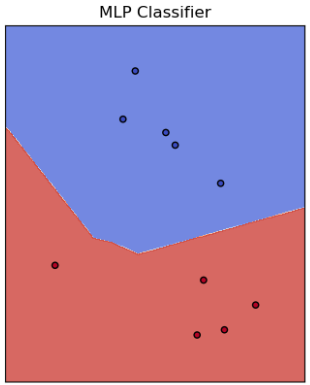
http://scikit-learn.org/stable/modules/svm.html

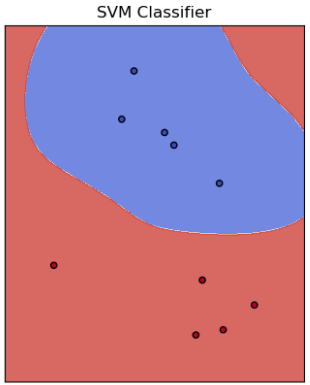
https://codingmachinelearning.wordpress.com/2016/08/19/svm-plotting-the-hyperplane/

2. CLASSIC MACHINE LEARNING ALGORITHMS USED IN FINANCIAL FRAUD DETECTION: INTUITIVE DESCRIPTION -NO MATH!-

First case: easy dataset (well separated blobs)

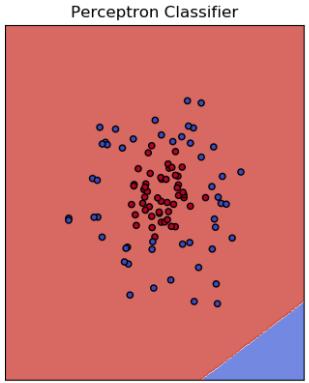


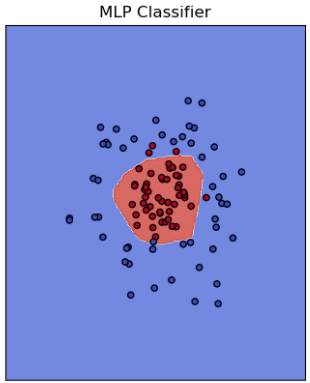


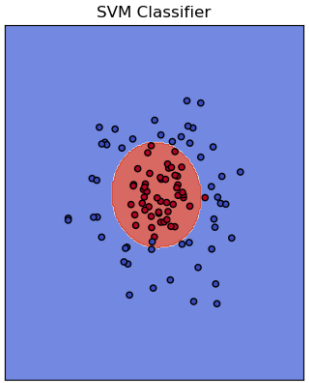


2. CLASSIC MACHINE LEARNING ALGORITHMS USED IN FINANCIAL FRAUD DETECTION: INTUITIVE DESCRIPTION -NO MATH!-

Second case: concentric dataset (dots organized in concentric circles)

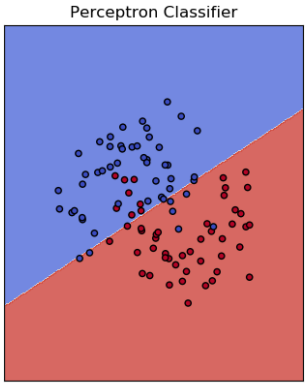


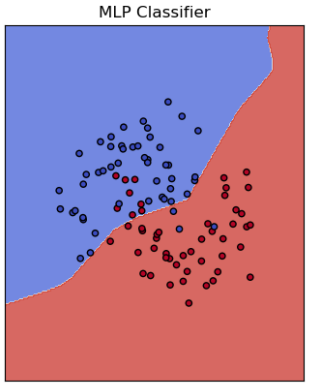


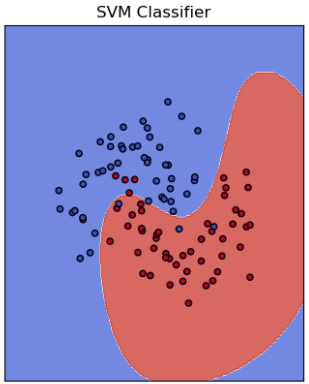


2. CLASSIC MACHINE LEARNING ALGORITHMS USED IN FINANCIAL FRAUD DETECTION: INTUITIVE DESCRIPTION -NO MATH!-

Third case: interaction dataset (dots organized in moon shapes)







3. INTRODUCTION IN PYTHON: MOTIVATION AND BASIC SYNTAX

Python is a high-level programming language, ranked as most demanded on the market in 2020.

Why? Because of many advantages:

- Supports scripting, procedural programming and object-oriented programming
- Data is dynamically defined, no need for type definition
- Many libraries to be used out of the box, and hence, easy manipulation of data structures
- Community open to development and community documented solutions
- Portable on various platforms
- Many IDE available: Pycharm Community – which is integrated with Github

3. INTRODUCTION IN PYTHON: MOTIVATION AND BASIC SYNTAX

Hello world example: **Print** (“Hello world!”)

Important: tabs (as in tab key) are important in Python. All code must be aligned on blocks.

Basic syntax examples:

Assignation:

```
x = y
```

Conditional:

```
if True:
    Print (“hello world!”)
else:
    Print (“bye”)
```

Loop:

For

```
girlNames = [“Ana”, ”Betty”, “Chris”]
for girlName in girlNames:
    Print (girlName)
```

Subfunction definition:

```
def NameOfFunction([parameters]):
    #block of code with tab
return variable_1, variable_2,..., variable_n
```

4. SOME PYTHON LIBRARIES FOR MACHINE LEARNING

Numpy – Numbers in Python: makes it really easy to handle data structures

Matplotlib - plotting data for visualization

Scikit-learn - scientific kit learning – easy to use for machine learning –spoiler alert-

Tensorflow – one of the most popular libraries for machine learning; made by Google

Pytorch – created by Facebook, also for machine learning

More info here: https://towardsdatascience.com/best-python-libraries-for-machine-learning-and-deep-learning-b0bd40c7e8c

5. CODING EXERCISE: FINANCIAL FRAUD DETECTION WITH PYTHON LIBRARIES

Choosing of problem and dataset:

- Financial fraud detection is a problem where automated systems come in handy for the data analysts.
- In order to develop and benchmark such systems, there are many datasets available for proving efficiency (have a look here)
- For this workshop, a dataset which contains credit card frauds is used: https://www.kaggle.com/mlg-ulb/creditcardfraud

5. CODING EXERCISE: FINANCIAL FRAUD DETECTION WITH PYTHON LIBRARIES

Plan of attack:

1. Get dataset
2. Read dataset
3. Load dataset into classifiers (Perceptron, MLP, SVM)
4. Train each classifier
5. Test each classifier to check which one is more successful

Example of implementation: https://github.com/Irome13/TTW_2020V1.git

5. CODING EXERCISE: FINANCIAL FRAUD DETECTION WITH PYTHON LIBRARIES

RESULTS FOR 10000 DATA:

Perceptron

Wrong classified data: 41.0 (on training dataset)
Perceptron Cross validation scores with 5 Folds: 0.4003016498754125

MLP

Wrong classified data: 37.0 (on training dataset)
MLP Cross validation scores with 5 Folds: 0.993902948475737

SVM

Wrong classified data: 0.0 (on training dataset)
SVM Cross validation scores with 5 Folds: 0.9962000992500248

6. CONCLUSIONS

- In this workshop, the objectives were to learn about:
  - ✓ **Python**: easy to use, brings strong tools in the background, usable in an intuitive way
  - ✓ **Machine Learning algorithms**: Perceptron, MLP, SVM
  - ✓ **Financial Fraud Detection**: results

- Perspectives: all the ranges of desktop application industry and a tool for data scientist jobs.
- However: making a difference in the domain, requires deep understanding of the models.
- Challenge: try this at home and for any kind of questions and curiosities, always feel free to ask the community (hence, also me: corina.sararu@draexlmaier.com)

Q&A