

You are currently looking at **version 1.2** of this notebook. To download notebooks and datafiles, as well as get help on Jupyter notebooks in the Coursera platform, visit the [Jupyter Notebook FAQ \(https://www.coursera.org/learn/python-machine-learning/resources/bANLa\)](https://www.coursera.org/learn/python-machine-learning/resources/bANLa) course resource.

Assignment 3 - Evaluation

In this assignment you will train several models and evaluate how effectively they predict instances of fraud using data based on [this dataset from Kaggle \(https://www.kaggle.com/dalpozz/creditcardfraud\)](https://www.kaggle.com/dalpozz/creditcardfraud).

Each row in `fraud_data.csv` corresponds to a credit card transaction. Features include confidential variables V1 through V28 as well as Amount which is the amount of the transaction.

The target is stored in the `class` column, where a value of 1 corresponds to an instance of fraud and 0 corresponds to an instance of not fraud.

```
In [1]: import numpy as np
import pandas as pd
```

Question 1

Import the data from `fraud_data.csv`. What percentage of the observations in the dataset are instances of fraud?

This function should return a float between 0 and 1.

```
In [2]: def answer_one():
    from sklearn.datasets import load_digits
    data_frame = pd.read_csv('fraud_data.csv')
    X, y = data_frame.drop('Class', axis=1), data_frame.Class;

    result = len(y[y==1]) / (len(y[y==1]) + len(y[y==0]))

    return result

answer_one()
```

```
Out[2]: 0.016410823768035772
```

```
In [3]: from sklearn.model_selection import train_test_split

df = pd.read_csv('readonly/fraud_data.csv')

X = df.iloc[:, 1:]
```