No arquivo **acesso.csv** estarão todos os dados referente ao acesso de páginas do usuário no nosso site, ou seja, se ele acessou a página home, como funciona e contato, por fim, informamos se ele comprou ou não. Nesse arquivo temos diversos dados do nosso usuário no passado, ou seja, suas características (acessos às páginas) e a marcação (se comprou), no total são 100 linhas.

Agora que temos o arquivo em mãos, precisamos fazer a leitura dele. Para ler esse arquivo, criaremos um novo arquivo python chamado dados.py que será o responsável por fazer a leitura desse arquivo e colocar os dados em um array que tenha as informações dos nossos elementos, porém nós precisamos apenas de 1 array ou 2 arrays? Lembra que utilizamos um array para os dados:

porco1 = [1, 1, 0]

porco2 = [1, 1, 0]

porco3 = [1, 1, 0]

cachorro4 = [1, 1, 1]

cachorro5 = [0, 1, 1]

cachorro6 = [0, 1, 1]

dados = [porco1, porco2, porco3, cachorro4, cachorro5, cachorro6]

E também um outro array para as marcações:

marcacoes = [1, 1, 1, -1, -1, -1]

Isso significa que precisaremos de 2 arrays também! Então o nosso arquivo será divido da seguinte forma:

acessou\_home,acessou\_como\_funciona, acessou\_contato,comprou

1,1,0,0

Os valores referentes a acessou\_home, acessou\_como\_funciona, acessou\_contato (1,1,0) serão o nosso primeiro array que representará os nossos dados. Porém, o valor referente ao comprou (0), será o nosso segundo array que representará as nossas marcações.

Observe que os valores dos nossos dados estão em função das marcações, então é comum chamarmos esse tipo de array de X, pois são os dados misteriosos que nós temos, e o que queremos calcular, nesse caso a marcação, chamamos de Y. É comum usar esses nomes para classificar dados, por isso utilizamos o X para indicar os nossos dados e o Y para dados que iremos prever.

**Importando os dados**

Agora que sabemos a forma que iremos representar os dados do nosso arquivo csv, precisamos importar esse arquivo. Dentro do arquivo dados.py, escreveremos o código para importar o csv:

import csv

Agora precisamos carregar as informações desse csv, então vamos definir a função carregar\_acessos():

import csv

def carregar\_acessos():

Mas o que essa função precisa fazer? Ela vai abrir o arquivo csv para leitura e ler todas as linhas. A cada linha que for lida, ela irá associar os valores para o array de dados e os valores para o array de marcações:

import csv

def carregar\_acessos():

dados = []

marcacoes = []

Agora que definimos os nossos arrays, precisamos abrir o arquivo utilizando a função open() e enviando o nome do arquivo como parâmetro:

import csv

def carregar\_acessos():

dados = []

marcacoes = []

arquivo = open('acesso.csv', 'rb')

Observe que utilizamos também o parâmetro rb indicando que queremos ler o arquivo.

Se estiver utilizando Python3, no segundo parâmetro da função open() pode passar apenas a string 'r'

import csv

def carregar\_acessos():

dados = []

marcacoes = []

arquivo = open('acesso.csv', 'r')

Agora precisamo ler esse arquivo csv, para isso iremos utilizar a função reader() da biblioteca que importamos:

import csv

def carregar\_acessos():

dados = []

marcacoes = []

arquivo = open('acesso.csv', 'rb')

csv.reader(arquivo)

A função reader devolve um leitor:

import csv

def carregar\_acessos():

dados = []

marcacoes = []

arquivo = open('acesso.csv', 'rb')

leitor = csv.reader(arquivo)

Esse leitor irá passar por cada uma das linhas do arquivo csv e adicionar os valores dos dados e a marcação:

import csv

def carregar\_acessos():

dados = []

marcacoes = []

arquivo = open('acesso.csv', 'rb')

leitor = csv.reader(arquivo)

for linha in leitor:

Porém, perceba que precisamos pegar os valores das colunas e não uma linha inteira! Então podemos copiar o nome de cada coluna do nosso arquivo e adicionar no for:

for acessou\_home, acessou\_como\_funciona,

acessou\_contato, comprou in leitor:

Nesse instante, estamos dizendo que, para cada uma dessas colunas (for acessou\_home, acessou\_como\_funciona, acessou\_contato, comprou) dentro do leitor (in leitor). Mas e agora? O que precisamos fazer? Precisamos adicionar ao nosso array dados as colunas: acessou\_home, acessou\_como\_funciona, acessou\_contato utilizando a função append():

import csv

def carregar\_acessos():

dados = []

marcacoes = []

arquivo = open('acesso.csv', 'rb')

leitor = csv.reader(arquivo)

for acessou\_home, acessou\_como\_funciona,

acessou\_contato, comprou in leitor:

dados.append([acessou\_home,

acessou\_como\_funciona,

acessou\_contato])

Por fim, adicionamos as marcações:

import csv

def carregar\_acessos():

dados = []

marcacoes = []

arquivo = open('acesso.csv', 'rb')

leitor = csv.reader(arquivo)

for acessou\_home, acessou\_como\_funciona,

acessou\_contato, comprou in leitor:

dados.append([acessou\_home,

acessou\_como\_funciona,

acessou\_contato])

marcacoes.append(comprou)

Agora que adicionamos os dados e as marcações podemos retorná-los:

import csv

def carregar\_acessos():

dados = []

marcacoes = []

arquivo = open('acesso.csv', 'rb')

leitor = csv.reader(arquivo)

for acessou\_home, acessou\_como\_funciona,

acessou\_contato, comprou in leitor:

dados.append([acessou\_home,

acessou\_como\_funciona,

acessou\_contato])

marcacoes.append(comprou)

return dados, marcacoes

Vamos testar o nosso código? Vá no terminal e abra o interpretador do python:

> python

Python 2.7.10 (default, Oct 14 2015, 16:09:02)

[GCC 5.2.1 20151010] on linux2

Type "help", "copyright", "credits" or "license" for more information.

>>>

Primeiro precisamos importa a função carregar\_acessos do arquivo dados.py:

>>> from dados import carregar\_acessos

>>>

Agora podemos utilizar a nossa função! Vamos retorna o valor dessa função:

>>> dados, marcacoes = carregar\_acessos()

>>>

Será que funcionou? Será que não? Vamos dar uma olhada no valor das nossas marcações:

>>> marcacoes

['comprou', '0', '0', '0', '0', '0', '1', '0', '1', '0', '1', '1', '1', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '1', '1', '1', '1', '0', '0', '0', '1', '0', '1', '1', '1', '1', '0', '0', '1', '0', '0', '0', '1', '0', '0', '1', '1', '0', '0', '0', '0', '1', '0', '1', '0', '1', '0', '1', '0', '1', '0', '0', '0', '1', '1', '0', '1', '1', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '1', '0', '0', '0', '1', '1', '0', '0', '0', '0', '1', '1', '0', '0', '0', '0', '1', '0', '0', '0']

>>>

Ele retornou um array com a primeira linha informando ao que se refere, nesse caso, se comprou ou não. Vamos observar as nossas primeiras marcações no arquivo acesso.csv:

1,1,0,0

1,1,0,0

1,1,0,0

1,1,0,0

1,1,0,0

1,0,1,1

...

Ao compararmos os primeiros resultados funcionou conforme o esperado! Agora vamos dar uma olhada nos nossos dados:

>>> dados

[['acessou\_home', 'acessou\_como\_funciona', 'acessou\_contato'], ['1', '1', '0'], ['1', '1', '0'], ['1', '1', '0'], ['1', '1', '0'], ['1', '1', '0'], ['1', '0', '1'], ['1', '1', '0'], ['1', '0', '1'], ['1', '1', '0'], ['1', '0', '1'], ['1', '1', '0'], ['1', '0', '1'], ['0', '0', '1'], ['0', '0', '1'], ['0', '0', '1'], ['0', '0', '1'], ['0', '0', '1'], ['0', '1', '0'], ['0', '0', '1'], ['0', '1', '0'], ['0', '0', '1'], ['0', '1', '0'], ['0', '0', '1'], ['0', '1', '0'], ['1', '0', '1'], ['1', '1', '1'], ['1', '1', '1'], ['1', '0', '1'], ['0', '1', '0'], ['0', '0', '1'], ['0', '1', '0'], ['1', '0', '1'], ['0', '0', '1'], ['1', '0', '1'], ['1', '0', '1'], ['1', '0', '1'], ['1', '0', '1'], ['0', '0', '1'], ['1', '1', '0'], ['1', '0', '1'], ['0', '1', '0'], ['0', '0', '1'], ['0', '1', '0'], ['1', '0', '1'], ['0', '0', '1'], ['0', '0', '1'], ['1', '0', '1'], ['1', '0', '1'], ['1', '1', '0'], ['1', '1', '0'], ['1', '1', '0'], ['1', '1', '0'], ['1', '0', '1'], ['1', '1', '0'], ['1', '0', '1'], ['1', '1', '0'], ['1', '0', '1'], ['1', '1', '0'], ['1', '0', '1'], ['0', '0', '1'], ['1', '0', '1'], ['0', '1', '0'], ['0', '0', '1'], ['0', '1', '0'], ['1', '0', '1'], ['0', '1', '1'], ['0', '0', '1'], ['1', '0', '1'], ['1', '0', '1'], ['1', '1', '0'], ['1', '1', '0'], ['0', '0', '1'], ['0', '0', '1'], ['0', '0', '1'], ['0', '0', '1'], ['0', '1', '0'], ['0', '0', '1'], ['0', '1', '0'], ['0', '0', '1'], ['1', '1', '0'], ['0', '0', '1'], ['0', '0', '1'], ['0', '0', '1'], ['1', '0', '1'], ['1', '0', '1'], ['1', '1', '0'], ['1', '1', '0'], ['0', '0', '1'], ['0', '0', '1'], ['1', '0', '1'], ['1', '0', '1'], ['1', '1', '0'], ['1', '1', '0'], ['0', '0', '1'], ['0', '0', '1'], ['1', '1', '1'], ['0', '0', '1'], ['0', '1', '0'], ['0', '0', '0']]

>>>

Os dados batem, conforme o esperado! Observe que agora nós separamos os nossos dados das nossas marcações, ou seja, separamos todos os dados que estão em função das marcações! Lembra que para esses casos de classificação é tradicional chamarmos os nossos dados de X, ou seja, os valores que conhecemos do nosso usuário como por exemplo, se ele acessou uma página, se ele tem perninha curta ou qualquer informação que nós já sabemos!

def carregar\_acessos():

X = []

marcacoes = []

# restante do código

Mas e as nossas marcações? Como é mesmo que chamávamos? Lembre-se que as marcações são os valores que não sabemos, ou seja, os valores que queremos calcular, prever! Tradicionalmente, esses valores, essas marcações que queremos prever, chamamos de Y:

def carregar\_acessos():

X = []

Y = []

# restante do código

Vamos substituir todos os dados para X e marcacoes para Y:

import csv

def carregar\_acessos():

X = []

Y = []

arquivo = open('acesso.csv', 'rb')

leitor = csv.reader(arquivo)

for acessou\_home,acessou\_como\_funciona,acessou\_contato, comprou in leitor:

X.append([acessou\_home,acessou\_como\_funciona,acessou\_contato])

Y.append(comprou)

return X, Y

**Analisando os valores adicionados**

Vamos dar uma olhada novamente no resultado das nossas marcações:

>>> marcacoes

['comprou', '0', '0', '0', '0', '0', '1', '0', '1', '0', '1', '1', '1', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '1', '1', '1', '1', '0', '0', '0', '1', '0', '1', '1', '1', '1', '0', '0', '1', '0', '0', '0', '1', '0', '0', '1', '1', '0', '0', '0', '0', '1', '0', '1', '0', '1', '0', '1', '0', '1', '0', '0', '0', '1', '1', '0', '1', '1', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '1', '0', '0', '0', '1', '1', '0', '0', '0', '0', '1', '1', '0', '0', '0', '0', '1', '0', '0', '0']

>>>

Repare que veio também o array com a informação do cabeçalho, ou seja, a palavra 'comprou', porém nós não queremos isso no nosso array, lembra como era as nossas marcações no primeiro exemplo?

marcacoes = [1, 1, 1, -1, -1, -1]

Como podemos ver, só tem números! Em nenhum momento especificamos o que significa por meio de uma palavra na primeira posição:

marcacoes = ['é\_um\_porco', 1, 1, 1, -1, -1, -1]

Então precisamos descartar esses cabeçalhos simplesmente não lendo a primeira linha! Como podemos fazer isso? Simples, basta apenas, após o momento em que criamos o leitor(leitor = csv.reader(arquivo)), adicionarmos o comando next(leitor) que pulará para a linha a seguir:

import csv

def carregar\_acessos():

X = []

Y = []

arquivo = open('acesso.csv', 'rb')

leitor = csv.reader(arquivo)

next(leitor)

for acessou\_home,acessou\_como\_funciona,acessou\_contato, comprou in leitor:

X.append([acessou\_home,acessou\_como\_funciona,acessou\_contato])

Y.append(comprou)

return X, Y

Vamos testar o nosso código novamente? Antes de importar a nossa função carregar\_acessos, feche o interpretador do python e abra novamente para que ele carregue a nova versão da nossa função. Vejamos o resultado:

>>> from dados import carregar\_acessos

>>>

Importou sem nenhum problema! Porém, ao invés de chamar de dados e marcacoeschamaremos de X e Y:

>>> from dados import carregar\_acessos

>>> X, Y = carregar\_acessos()

>>>

Vamos verificar o valor de Y:

>>> from dados import carregar\_acessos

>>> X, Y = carregar\_acessos()

>>> Y

['0', '0', '0', '0', '0', '1', '0', '1', '0', '1', '1', '1', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '1', '1', '1', '1', '0', '0', '0', '1', '0', '1', '1', '1', '1', '0', '0', '1', '0', '0', '0', '1', '0', '0', '1', '1', '0', '0', '0', '0', '1', '0', '1', '0', '1', '0', '1', '0', '1', '0', '0', '0', '1', '1', '0', '1', '1', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '1', '0', '0', '0', '1', '1', '0', '0', '0', '0', '1', '1', '0', '0', '0', '0', '1', '0', '0', '0']

Aparentemente tudo está funcionando como o esperado! Porém, '0'? '1'? Isso é uma *string*! Não queremos *string*s, nós queremos números! Afinal estamos lendo números... E como resolver isso? Podemos converter as *string*s para números, mas que tipo de números? Nesse caso estamos trabalhando com números inteiros, então converteremos para números inteiros! Para converter uma string para inteiro, usaremos a maneira mais tradicional que é enviar, por parâmetro, a coluna desejada na instrução int():

import csv

def carregar\_acessos():

X = []

Y = []

arquivo = open('acesso.csv', 'rb')

leitor = csv.reader(arquivo)

next(leitor)

for acessou\_home,acessou\_como\_funciona,acessou\_contato, comprou in leitor:

X.append([int(acessou\_home),int(acessou\_como\_funciona)

,int(acessou\_contato)])

Y.append(int(comprou))

return X, Y

Testando novamente o nosso código:

>>> from dados import carregar\_acessos

>>> X, Y = carregar\_acessos()

>>>

Vejamos agora o valor do Y:

>>> from dados import carregar\_acessos

>>> X, Y = carregar\_acessos()

>>> Y

[0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0]

E o X?

>>> X

[[1, 1, 0], [1, 1, 0], [1, 1, 0], [1, 1, 0], [1, 1, 0], [1, 0, 1], [1, 1, 0], [1, 0, 1], [1, 1, 0], [1, 0, 1], [1, 1, 0], [1, 0, 1], [0, 0, 1], [0, 0, 1], [0, 0, 1], [0, 0, 1], [0, 0, 1], [0, 1, 0], [0, 0, 1], [0, 1, 0], [0, 0, 1], [0, 1, 0], [0, 0, 1], [0, 1, 0], [1, 0, 1], [1, 1, 1], [1, 1, 1], [1, 0, 1], [0, 1, 0], [0, 0, 1], [0, 1, 0], [1, 0, 1], [0, 0, 1], [1, 0, 1], [1, 0, 1], [1, 0, 1], [1, 0, 1], [0, 0, 1], [1, 1, 0], [1, 0, 1], [0, 1, 0], [0, 0, 1], [0, 1, 0], [1, 0, 1], [0, 0, 1], [0, 0, 1], [1, 0, 1], [1, 0, 1], [1, 1, 0], [1, 1, 0], [1, 1, 0], [1, 1, 0], [1, 0, 1], [1, 1, 0], [1, 0, 1], [1, 1, 0], [1, 0, 1], [1, 1, 0], [1, 0, 1], [0, 0, 1], [1, 0, 1], [0, 1, 0], [0, 0, 1], [0, 1, 0], [1, 0, 1], [0, 1, 1], [0, 0, 1], [1, 0, 1], [1, 0, 1], [1, 1, 0], [1, 1, 0], [0, 0, 1], [0, 0, 1], [0, 0, 1], [0, 0, 1], [0, 1, 0], [0, 0, 1], [0, 1, 0], [0, 0, 1], [1, 1, 0], [0, 0, 1], [0, 0, 1], [0, 0, 1], [1, 0, 1], [1, 0, 1], [1, 1, 0], [1, 1, 0], [0, 0, 1], [0, 0, 1], [1, 0, 1], [1, 0, 1], [1, 1, 0], [1, 1, 0], [0, 0, 1], [0, 0, 1], [1, 1, 1], [0, 0, 1], [0, 1, 0], [0, 0, 0]]

Agora sim nós carregamos os nossos números!

**Melhorando a leitura do código**

Observe que o nosso código ficou um pouco extenso e repetitivo:

import csv

def carregar\_acessos():

X = []

Y = []

arquivo = open('acesso.csv', 'rb')

leitor = csv.reader(arquivo)

next(leitor)

for acessou\_home,acessou\_como\_funciona,acessou\_contato, comprou in leitor:

X.append([int(acessou\_home),int(acessou\_como\_funciona)

,int(acessou\_contato)])

Y.append(int(comprou))

return X, Y

Veja que utilizamos os nomes acessou\_home, acessou\_contato, acessou\_alguma\_coisa! Vamos retirar esse acessou, pois já sabemos que nossos dados refere-se a acesso de página do usuário, então não precisamos repetir a mesma palavra em todos os nossos dados! Veja o resultado do nosso código:

import csv

def carregar\_acessos():

X = []

Y = []

arquivo = open('acesso.csv', 'rb')

leitor = csv.reader(arquivo)

next(leitor)

for home, como\_funciona, contato, comprou in leitor:

X.append([int(home),int(como\_funciona)

,int(contato)])

Y.append(int(comprou))

return X, Y

Também alteraremos no nosso arquivo csv:

home, como\_funciona, contato, comprou

1,1,0,0

1,1,0,0

1,1,0,0

1,1,0,0

1,1,0,0

1,0,1,1

...

Melhoramos uma boa parte do nosso código e deixamos mais limpo, ou seja, com uma leitura mais clara, mas perceba que ainda existe alguns detalhes, vejamos esse trecho de código:

X.append([int(home),int(como\_funciona)

,int(contato)])

Observe que enviamos um monte de informação para o X que, a primeira vista, não da pra saber a que se refere, nesses casos, podemos extrair todas essas informações para uma variável que deixará mais claro o significado desses valores:

dado = [int(home),int(como\_funciona)

,int(contato)]

Observe agora o resultado final do nosso código:

import csv

def carregar\_acessos():

X = []

Y = []

arquivo = open('acesso.csv', 'rb')

leitor = csv.reader(arquivo)

next(leitor)

for home,como\_funciona,contato, comprou in leitor:

dado = [int(home),int(como\_funciona)

,int(contato)]

X.append(dado)

Y.append(int(comprou))

return X, Y

Bem mais limpo e de fácil compreensão! Vamos verificar se o nosso código ainda funciona? Porém, dessa vez, ao invés de usar o interpretador python, vamos criar um novo arquivo chamado classifica\_acessos.py e importar a função carregar\_acessos()dentro desse arquivo:

from dados import carregar\_acessos

X, Y = carregar\_acessos()

Por fim, vamos imprimir o X e o Y para verificar se está funcionando corretamente:

from dados import carregar\_acessos

X, Y = carregar\_acessos()

print(X)

print(Y)

Vamos testar o nosso código:

> python classifica\_acessos.py

[[1, 1, 0], [1, 1, 0], [1, 1, 0], [1, 1, 0], [1, 1, 0], [1, 0, 1], [1, 1, 0], [1, 0, 1], [1, 1, 0], [1, 0, 1], [1, 1, 0], [1, 0, 1], [0, 0, 1], [0, 0, 1], [0, 0, 1], [0, 0, 1], [0, 0, 1], [0, 1, 0], [0, 0, 1], [0, 1, 0], [0, 0, 1], [0, 1, 0], [0, 0, 1], [0, 1, 0], [1, 0, 1], [1, 1, 1], [1, 1, 1], [1, 0, 1], [0, 1, 0], [0, 0, 1], [0, 1, 0], [1, 0, 1], [0, 0, 1], [1, 0, 1], [1, 0, 1], [1, 0, 1], [1, 0, 1], [0, 0, 1], [1, 1, 0], [1, 0, 1], [0, 1, 0], [0, 0, 1], [0, 1, 0], [1, 0, 1], [0, 0, 1], [0, 0, 1], [1, 0, 1], [1, 0, 1], [1, 1, 0], [1, 1, 0], [1, 1, 0], [1, 1, 0], [1, 0, 1], [1, 1, 0], [1, 0, 1], [1, 1, 0], [1, 0, 1], [1, 1, 0], [1, 0, 1], [0, 0, 1], [1, 0, 1], [0, 1, 0], [0, 0, 1], [0, 1, 0], [1, 0, 1], [0, 1, 1], [0, 0, 1], [1, 0, 1], [1, 0, 1], [1, 1, 0], [1, 1, 0], [0, 0, 1], [0, 0, 1], [0, 0, 1], [0, 0, 1], [0, 1, 0], [0, 0, 1], [0, 1, 0], [0, 0, 1], [1, 1, 0], [0, 0, 1], [0, 0, 1], [0, 0, 1], [1, 0, 1], [1, 0, 1], [1, 1, 0], [1, 1, 0], [0, 0, 1], [0, 0, 1], [1, 0, 1], [1, 0, 1], [1, 1, 0], [1, 1, 0], [0, 0, 1], [0, 0, 1], [1, 1, 1], [0, 0, 1], [0, 1, 0], [0, 0, 0]]

[0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0]

Está funcionando perfeitamente! Agora podemos tirar as impressões do X e Y:

from dados import carregar\_acessos

X, Y = carregar\_acessos()