

Introduction à la Programmation des Algorithmes

6.3. Python – Fichiers et modules

François Fleuret

<https://fleuret.org/11x001/>



Fichiers

Python offre de manière native, c'est à dire intégrées au langage, des fonctions de manipulation de fichiers similaires à `stdio.h` en C.

Un fichier Python ouvert en lecture est itérable, et parcours les lignes du fichier les unes après les autres.

```
1  f = open('machin.txt', 'w')
2  f.write('Une première ligne\n')
3  f.write('Une seconde ligne\n')
4  f.close()
5
6  g = open('machin.txt', 'a')
7  for k in range(3, 6):
8      g.write(f'ligne {k}\n')
9  g.close()
10
11 f = open('machin.txt', 'r')
12 for l in f:
13     print(l.strip())
14 f.close()

affiche

Une première ligne
Une seconde ligne
ligne 3
ligne 4
ligne 5
```

Les erreurs de manipulations de fichiers sont gérées avec des exceptions.

```
1  >>> f = open('/truc', 'w')
2  Traceback (most recent call last):
3      File "<stdin>", line 1, in <module>
4  PermissionError: [Errno 13] Permission denied: '/truc'
5  >>> f = open('nexistepas.dat', 'r')
6  Traceback (most recent call last):
7      File "<stdin>", line 1, in <module>
8  FileNotFoundError: [Errno 2] No such file or directory: 'nexistepas.dat'
```

Il arrive souvent qu'une partie d'un programme puisse être structurée avec la notion de **contexte**, qui correspond à un certain état dans lequel des opérations sont possibles, et dans lequel on entre et d'où on sort avec certaines opérations.

La manipulation d'un fichier correspond bien à cette situation: l'ouverture du fichier permet de se placer dans un contexte où on peut faire de opérations de lecture ou d'écriture, et on sort du contexte en fermant le fichier.

Python offre le mot clé **with** qui permet de se placer dans un contexte. Ce mot clé doit être suivi d'un objet capable de traiter l'entrée et la sortie du contexte.

La seule utilisation que nous en ferons dans ce cours sera pour manipuler des fichiers.

```

1 with open('machin.txt', 'w') as f:
2     f.write('Une première ligne\n')
3     f.write('Une seconde ligne\n')
4
5 with open('machin.txt', 'a') as f:
6     for k in range(3, 6):
7         f.write(f'ligne {k}\n')
8
9 with open('machin.txt', 'r') as f:
10    for l in f:
11        print(l.strip())

```

affiche

```

Une première ligne
Une seconde ligne
ligne 3
ligne 4
ligne 5

```

On peut combiner l'utilisation des contextes et la gestion des exceptions:

```

1 try:
2     with open('nexistepas.txt', 'r') as f:
3         for l in f:
4             print(l.strip())
5
6 except FileNotFoundError as e:
7     print(f'Le fichier {e.filename} est introuvable.')

```

affiche

```
Le fichier nexistepas.txt est introuvable.
```

Modules

Il est possible de structurer un long programme Python en plusieurs fichiers, et plus généralement de regrouper des classes, des variables, et des fonctions relatives à un type de données ou de problèmes spécifiques.

On peut utiliser des éléments définis dans un autre fichier à l'aide de l'instruction `import`, qui exécute un autre fichier.

Les identifiants créés durant cette exécution sont mémorisés de façon à pouvoir y accéder en les préfixant avec le nom du module, et à éviter ainsi les collisions de noms.

Si nous avons un fichier `truc.py` dont le contenu est

```
1 def bidule(n):
2     for k in range(n):
3         print(f'bidule-{k}')
```

alors le programme Python dans le même répertoire

```
1 import truc
2
3 truc.bidule(4)
```

affiche

```
bidule-0
bidule-1
bidule-2
bidule-3
```

Ici l'identifiant `bidule` est défini dans l'**espace de nommage** du module `truc`.

Dans la forme la plus simple, la syntaxe est donc

```
import nom_1, nom_2, ..., nom_K
```

où les `nom_k` sont des noms de **modules**, qui sont des noms de fichiers sans l'extension `.py`.

Il est ensuite possible de faire référence aux fonctions, classes et variables définies dans ces fichiers en préfixant leurs identifiants avec le nom du module.

On peut également importer un élément particulier. Il est alors dans l'espace de nommage normal et on peut y faire référence sans le nom du module.

```
1 from truc import bidule  
2  
3 bidule(4)
```

Et il est possible d'importer des modules ou des éléments de modules en aliasant leur noms, dans un but de simplicité ou pour éviter des collisions d'identifiants.

```
1 import truc as tr  
2  
3 tr.bidule(4)
```

ou

```
1 from truc import bidule as chose  
2  
3 chose(4)
```

En général un module ne fait que définir des classes, variables et fonctions, mais n'exécute pas de code.

Il peut arriver que l'on veuille qu'un module puisse être exécuté, par exemple pour faire des tests.

On peut tester la variable `__name__` qui contient le nom du module s'il est exécuté via un `import` et `__main__` s'il est exécuté directement.

```
1 def bidule(n):  
2     for k in range(n):  
3         print(f'bidule-{k}')  
4  
5 if __name__ == '__main__':  
6     bidule(3)
```

affiche

```
bidule-0  
bidule-1  
bidule-2
```

Il existe un très (TRÈS!) grand nombre de modules Python disponibles pour réaliser des tâches de manipulations de fichiers, créations graphiques, opérations réseaux, calculs mathématiques, etc.

Un exemple simple est la récupération d'un fichier sur le web.

 Récupérer des données depuis un site distant est souvent pratique mais pose un certains nombre de problèmes.

- Le programme ne fonctionne qu'avec une connexion réseau.
- Cela peut entraîner des coûts et de l'engorgement pour le serveur web correspondant.
- La durée de vie du programme est limitée par celle du document web utilisé.

Accès et manipulations de fichiers distants

```

1 from urllib.request import urlopen
2
3 url = 'https://api.blockchain.info/charts/market-price?format=csv'
4
5 with urlopen(url) as f:
6     l = list(f)
7
8 for s in l[-5:]:
9     print(s)

affiche

b'2021-10-19 00:00:00,61971.59\n'
b'2021-10-20 00:00:00,64287.64\n'
b'2021-10-21 00:00:00,66063.56\n'
b'2021-10-22 00:00:00,62354.86\n'
b'2021-10-23 00:00:00,60697.06\n'

```

Ce sont des chaînes binaires, qu'il faut convertir en chaînes de caractères avec `decode()`. On peut également supprimer le `\n` final avec `strip()` et les découper en champ séparés par des virgules avec `split(',',')`.

```

1 from urllib.request import urlopen
2
3 url = 'https://api.blockchain.info/charts/market-price?format=csv'
4
5 with urlopen(url) as f:
6     l = list(f)
7
8 for s in l[-5:]:
9     x = s.decode().strip().split(',')
10    print(f'{x[0]} -> ${x[1]}')

affiche

2021-10-19 00:00:00 -> $61971.59
2021-10-20 00:00:00 -> $64287.64
2021-10-21 00:00:00 -> $66063.56
2021-10-22 00:00:00 -> $62354.86
2021-10-23 00:00:00 -> $60697.06

```

La gestion des dates est complexe à faire correctement. On peut pour cela utiliser le module `datetime`.

```
1  >>> import datetime
2  >>> datetime.date(2019, 4, 15)
3  datetime.date(2019, 4, 15)
4  >>> t = datetime.datetime.now()
5  >>> t
6  datetime.datetime(2021, 10, 24, 13, 24, 21, 636715)
7  >>> t.year
8  2021
9  >>> t.month
10 10
11 >>> t.timetuple()
12 time.struct_time(tm_year=2021, tm_mon=10, tm_mday=24, tm_hour=13,
13     tm_min=24, tm_sec=21, tm_wday=6, tm_yday=297, tm_isdst=-1)
14 >>> t0 = datetime.datetime.fromisoformat('2011-11-04 14:05:00')
15 >>> t - t0
16 datetime.timedelta(days=3641, seconds=83961, microseconds=636715)
17 >>> datetime.datetime.strptime('Sun Oct 24 13:18:13 CEST 2021',
18                               '%a %b %d %H:%M:%S %Z %Y')
19 datetime.datetime(2021, 10, 24, 13, 18, 13)
```

```
1  from datetime import datetime
2  from urllib.request import urlopen
3
4  url = 'https://api.blockchain.info/charts/market-price?format=csv'
5
6  with urlopen(url) as f:
7      l = list(f)
8
9  for s in l[-5:]:
10     x = s.decode().strip().split(',')
11     date = datetime.fromisoformat(x[0]).strftime('%d %b %Y')
12     prix = x[1]
13     print(f'{date} -> ${prix}')
```

affiche

```
20 Oct 2021 -> $64287.64
21 Oct 2021 -> $66063.56
22 Oct 2021 -> $62354.86
23 Oct 2021 -> $60697.06
24 Oct 2021 -> $61277.28
```


Manipulations d'images

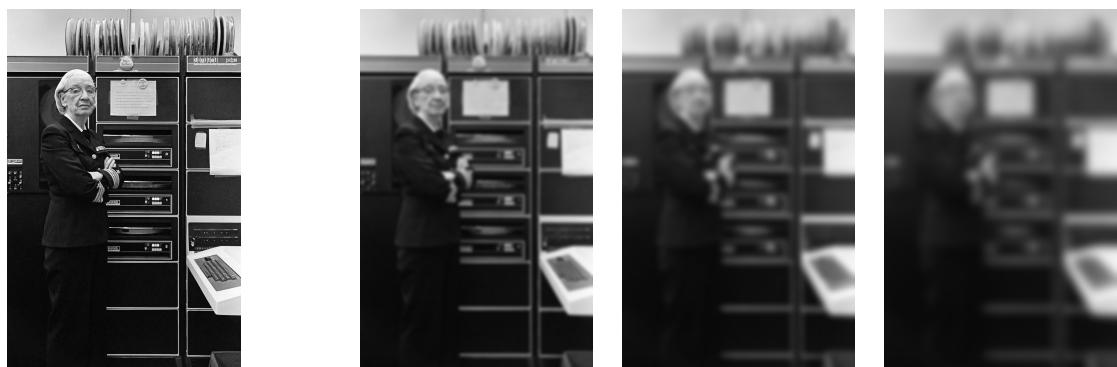
```
1 from PIL import Image, ImageFilter  
2  
3 im = Image.open('GraceHopper.png')  
4  
5 new_im = im.transpose(Image.ROTATE_90)  
6 new_im.save('GraceHopper_r90.png')
```



```
1 from PIL import Image, ImageFilter  
2  
3 im = Image.open('GraceHopper.png')  
4  
5 new_im = im.resize((im.width // 4, im.height // 4))  
6 new_im.save('GraceHopper_resize.png')
```



```
1 from PIL import Image, ImageFilter  
2  
3 im = Image.open('GraceHopper.png')  
4  
5 for k in { 5, 10, 15 }:  
6     new_im = im.filter(ImageFilter.GaussianBlur(k))  
7     new_im.save(f'GraceHopper_blurry_{k}.png')
```



Numpy, SciPy et Matplotlib

Grâce à sa souplesse d'utilisation et aux notebooks interactifs, Python a remplacé en grande partie des outils utilisés en sciences expérimentales tels que R ou Matlab.

Trois importantes librairies qui contribuent à ce succès sont Numpy pour le calcul algébrique rapide, SciPy pour le calcul scientifique et Matplotlib pour les représentations graphiques.

Python est un langage très lent et programmer une opération telle qu'un produit de matrice serait horriblement inefficace.

NumPy offre une très large palette d'opérations qui manipulent des matrices multi-dimensionnelles `ndarray`.

Cette structure permet en particulier de représenter des vecteurs classiques, dont les coefficients sont indexés par un seul indexe, et les matrices, avec deux indexes.

```
1 import numpy  
2  
3 m = numpy.array([[ 1,  1 ], [ -1,  1 ]])  
4  
5 v = numpy.array([ 2,  3 ])  
6  
7 print(v, m @ v)  
  
affiche  
  
[2 3] [5 1]
```

```
1 >>> m = numpy.random.normal(0, 1, (4, 4))  
2 >>> m  
3 array([[-0.82999735, -0.91818592, -0.96424688,  2.14184176],  
4        [-1.92361767, -0.66849122, -0.96641268, -0.14764259],  
5        [ 1.31496295, -0.20098614,  0.08986932, -0.1929336 ],  
6        [-0.42590301,  1.00492111,  0.13936688, -1.35530879]])  
7 >>> m[2]  
8 array([ 1.31496295, -0.20098614,  0.08986932, -0.1929336 ])  
9 >>> m[:, 0]  
10 array([-0.82999735, -1.92361767,  1.31496295, -0.42590301])  
11 >>> m = m @ m.transpose()  
12 >>> numpy.linalg.eig(m)[0]  
13 array([10.76422836,  5.53350223,  0.1160834 ,  0.60071016])
```

```

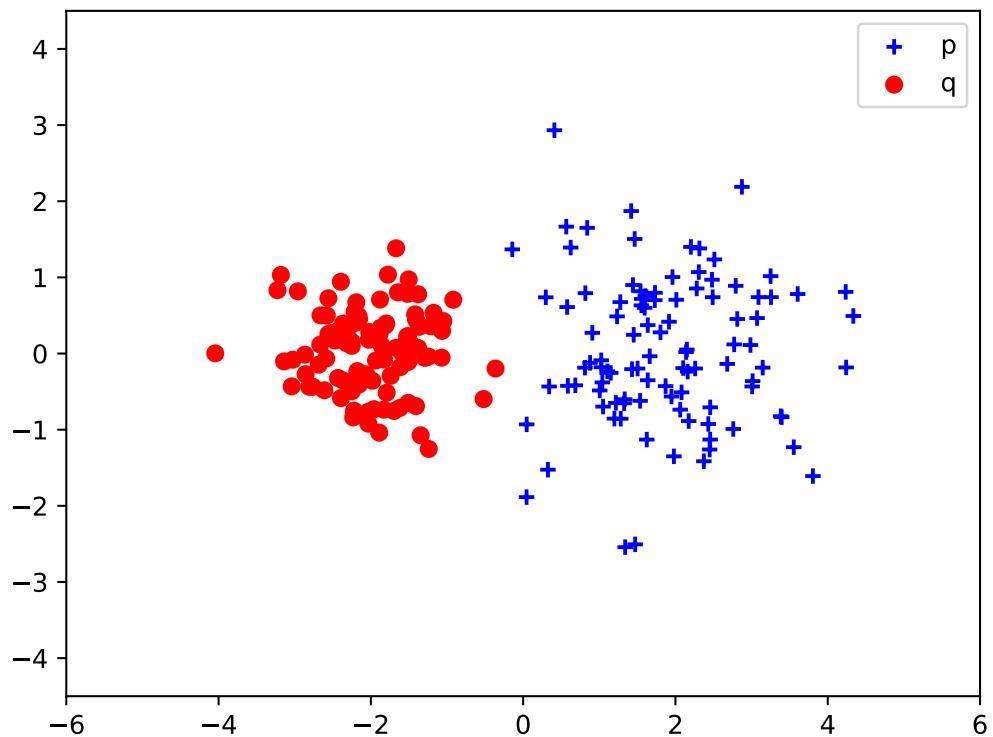
1  >>> p = numpy.random.normal(0, 1, (1000, 1000))
2  >>> q = numpy.random.normal(0, 1, (1000, 1000))
3  >>> p @ q
4  array([[ -6.59884453,   4.33408239, -12.62841631, ..., -15.51141445,
5          48.06933436,  41.36263837],
6          [  6.89179897,  14.16530801,  11.06033872, ...,  16.96517321,
7          50.21274734,  14.58360869],
8          [-16.24756087,  42.41229225,  18.44074437, ..., -23.76344462,
9          46.5083391 , -10.35787684],
10         ...,
11          [  7.85340925,  12.83766931,  24.97148857, ...,  39.95634369,
12          -13.55957854, -52.46306412],
13          [-16.55927574, -20.93303248,  19.87050555, ...,  9.04560522,
14          9.17857406, -17.38930732],
15          [ 17.89102294, -43.35488196, -33.76625869, ..., -4.99645913,
16          25.33190269,   8.47263572]])
```

Ce calcul qui demande 10^9 multiplications est exécuté sur un ordinateur portable standard (Intel i7-8650U 1.90GHz) en 16ms. Soit 62.5 milliards de multiplications par seconde.

On peut combiner Numpy avec Matplotlib pour tracer des graphiques.

```

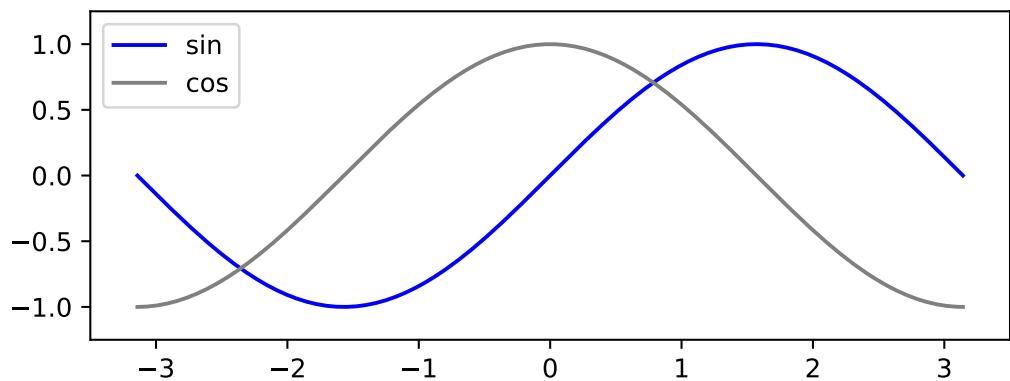
1  import matplotlib.pyplot as plt
2  import numpy
3
4  p = numpy.random.normal(0.0, 1.0, (100, 2)) + numpy.array([ 2, 0 ])
5  q = numpy.random.normal(0.0, 0.5, (100, 2)) + numpy.array([-2, 0 ])
6
7  fig = plt.figure()
8  subfig = fig.add_subplot()
9  subfig.set(xlim = (-6.0, 6.0), ylim = (-4.5, 4.5), aspect = 1)
10
11 subfig.scatter(p[:, 0], p[:, 1], marker = '+', color = 'blue', label = 'p')
12 subfig.scatter(q[:, 0], q[:, 1], marker = 'o', color = 'red', label = 'q')
13 subfig.legend()
14
15 plt.savefig('figure.pdf')
```



```

1 import matplotlib.pyplot as plt
2 import numpy, math
3
4 x = numpy.linspace(-math.pi, math.pi, 250)
5
6 fig = plt.figure()
7 subfig = fig.add_subplot()
8 subfig.set(xlim = (-3.5, 3.5), ylim = (-1.25, 1.25), aspect = 1)
9
10 subfig.plot(x, numpy.sin(x), color = 'blue', label = 'sin')
11 subfig.plot(x, numpy.cos(x), color = 'gray', label = 'cos')
12 subfig.legend()
13
14 plt.savefig('figure.pdf')

```



```

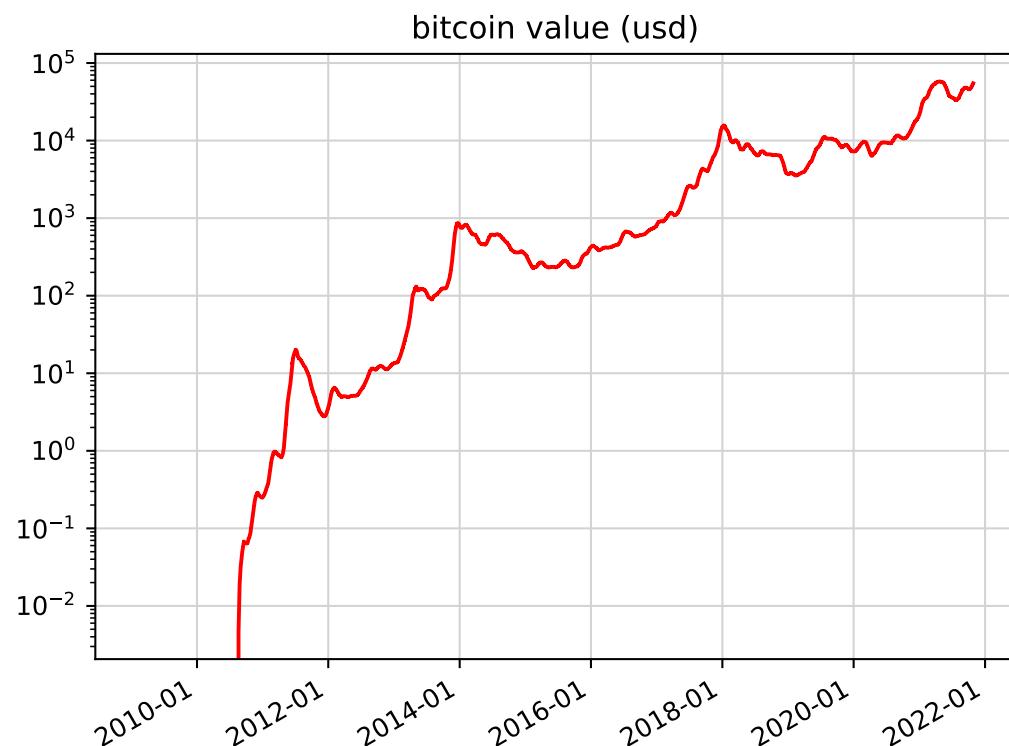
1  from datetime import datetime
2  from urllib.request import urlopen
3
4  url = 'https://api.blockchain.info/charts/market-price?\n'
5  daysAverageString=30D&timespan=all&sampled=true&\n
6  metadata=false&cors=true&format=csv'
7
8  with urlopen(url) as f:
9      l = list(f)
10
11 dates, prix = [], []
12
13 for s in l:
14     x = s.decode().strip().split(',')
15     dates.append(datetime.fromisoformat(x[0]))
16     prix.append(float(x[1]))

```

```

1 import matplotlib.pyplot as plt
2 import matplotlib.dates as dts
3
4 fig, subfig = plt.subplots()
5
6 subfig.set_title('bitcoin value (usd)')
7
8 date_form = dts.DateFormatter("%Y-%m")
9 subfig.xaxis.set_major_formatter(date_form)
10 fig.autofmt_xdate()
11 subfig.set_yscale('log')
12
13 subfig.grid(color = 'lightgray')
14 subfig.plot_date(dates, prix, fmt = ', -r')
15
16 plt.savefig('figure.pdf')

```



PyTorch et IA

Python est le langage le plus utilisé en Intelligence Artificielle.

	Language(s)	License	Main backer
PyTorch	Python, C++	BSD	Facebook
TensorFlow	Python, C++	Apache	Google
JAX	Python	Apache	Google
MXNet	Python, C++, R, Scala	Apache	Amazon
CNTK	Python, C++	MIT	Microsoft
Torch 7	Lua	BSD	Facebook
Theano	Python	BSD	U. of Montreal
Caffe	C++	BSD 2 clauses	U. of CA, Berkeley

MNIST data-set

```
1 1 8 3 6 1 0 3 1 0 0 1 1 2 7 3 0 4 6 5  
2 6 4 7 1 8 9 9 3 0 7 1 0 2 0 3 5 4 6 5  
8 6 3 7 5 8 0 9 1 0 3 1 2 2 3 3 6 4 7 5  
0 6 2 7 9 8 5 9 2 1 1 4 4 5 6 4 1 2 5 3  
9 3 9 0 5 9 6 5 7 4 1 3 4 0 4 8 0 4 3 6  
8 7 6 0 9 7 5 7 2 1 1 6 8 9 4 1 5 2 2 9  
0 3 9 6 7 2 0 3 5 4 3 6 5 8 9 5 4 7 4 2  
1 3 4 8 9 1 9 2 8 7 9 1 8 7 4 1 3 1 1 0  
2 3 9 4 9 2 1 6 8 4 1 7 4 4 9 2 8 7 2 4  
4 2 1 9 7 2 8 7 6 9 2 2 3 8 1 6 5 1 1 0  
4 0 9 1 1 2 4 3 2 7 3 8 6 9 0 5 6 0 7 6  
2 6 4 5 8 3 1 5 1 9 2 7 4 4 8 1 5 8 9  
5 6 7 9 9 3 7 0 9 0 6 6 2 3 9 0 7 5 4 8  
0 9 4 1 2 8 7 1 2 6 1 0 3 0 1 1 8 2 0 3  
9 4 0 5 0 6 1 7 7 8 1 9 2 0 5 1 2 2 7 3  
5 4 9 7 1 8 3 9 6 0 3 1 1 2 6 3 5 7 6 8  
3 9 5 8 5 7 4 1 1 3 1 7 5 5 5 2 5 8 7 0  
9 7 7 5 0 9 0 0 8 9 2 4 8 1 6 1 6 5 1 8  
3 4 0 5 5 8 3 6 2 3 9 2 1 1 5 2 1 3 2 8  
7 3 7 2 4 6 9 7 2 4 2 8 1 1 3 8 4 0 6 5
```

Images 28×28 , 60K exemples d'apprentissage, 10K exemples de test.

```
model = nn.Sequential(  
    nn.Conv2d( 1, 32, 5), nn.MaxPool2d(3), nn.ReLU(),  
    nn.Conv2d(32, 64, 5), nn.MaxPool2d(2), nn.ReLU(),  
    nn.Flatten(),  
    nn.Linear(256, 200), nn.ReLU(),  
    nn.Linear(200, 10)  
)  
  
nb_epochs, batch_size = 10, 100  
criterion = nn.CrossEntropyLoss()  
optimizer = torch.optim.SGD(model.parameters(), lr = 0.1)  
  
model.to(device)  
criterion.to(device)  
train_input, train_targets = train_input.to(device), train_targets.to(device)  
  
for e in range(nb_epochs):  
    for input, targets in zip(train_input.split(batch_size),  
                             train_targets.split(batch_size)):  
        output = model(input)  
        loss = criterion(output, targets)  
        optimizer.zero_grad()  
        loss.backward()  
        optimizer.step()
```

Pytorch offre un grand nombre de réseaux de neurones déjà entraînés.

Nous pouvons tester un de ces réseaux de classification d'image sur cet exemple:



```
import PIL, torch, torchvision

# Load and normalize the image
to_tensor = torchvision.transforms.ToTensor()
img = to_tensor(PIL.Image.open('blacklab.jpg'))
img = img.unsqueeze(0)
img = 0.5 + 0.5 * (img - img.mean()) / img.std()

# Load and evaluate the network
alexnet = torchvision.models.alexnet(weights = 'IMAGENET1K_V1')
alexnet.eval()

output = alexnet(img)

# Prints the classes
scores, indexes = output.view(-1).sort(descending = True)

class_names = eval(open('imagenet1000_clsid_to_human.txt', 'r').read())

for k in range(12):
    print(f'#{k+1} {scores[k].item():.02f} {class_names[indexes[k].item()]}')
```



12.26 Weimaraner
10.95 Chesapeake Bay retriever
10.87 Labrador retriever
10.10 Staffordshire bullterrier, Staffordshire bull terrier
9.55 flat-coated retriever
9.40 Italian greyhound
9.31 American Staffordshire terrier, Staffordshire terrier, American pit bull terrier, pit bull terrier
9.12 Great Dane
8.94 German short-haired pointer
8.53 Doberman, Doberman pinscher
8.35 Rottweiler
8.25 kelpie
8.24 barrow, garden cart, lawn cart, wheelbarrow
8.12 bucket, pail
8.07 soccer ball



Weimaraner



Chesapeake Bay retriever