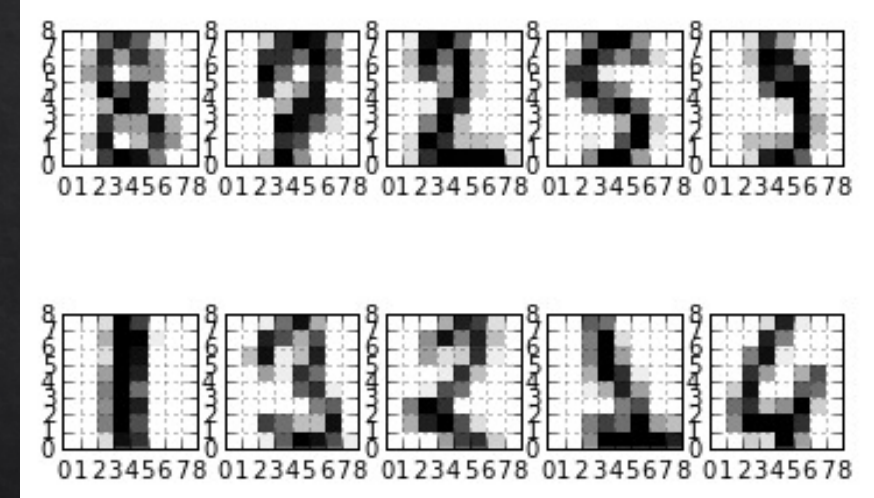


Python for Data Analysis

Dean Chérif – IBO 1 – A5

Contexte de la base de données

- ◆ Images de chiffres écrits à la main (normalisées et transformées en images 8x8 pixels)
- ◆ L'objectif est d'entraîner le programme à reconnaître ces chiffres écrits à la main.
- ◆ Exemples d'images:



Récupération des données

- ◆ Le jeu de données est disponible par défaut dans la librairie Scikit Learn.

```
from sklearn import datasets  
dataset = datasets.load_digits()
```


Quel algorithme de classification ?

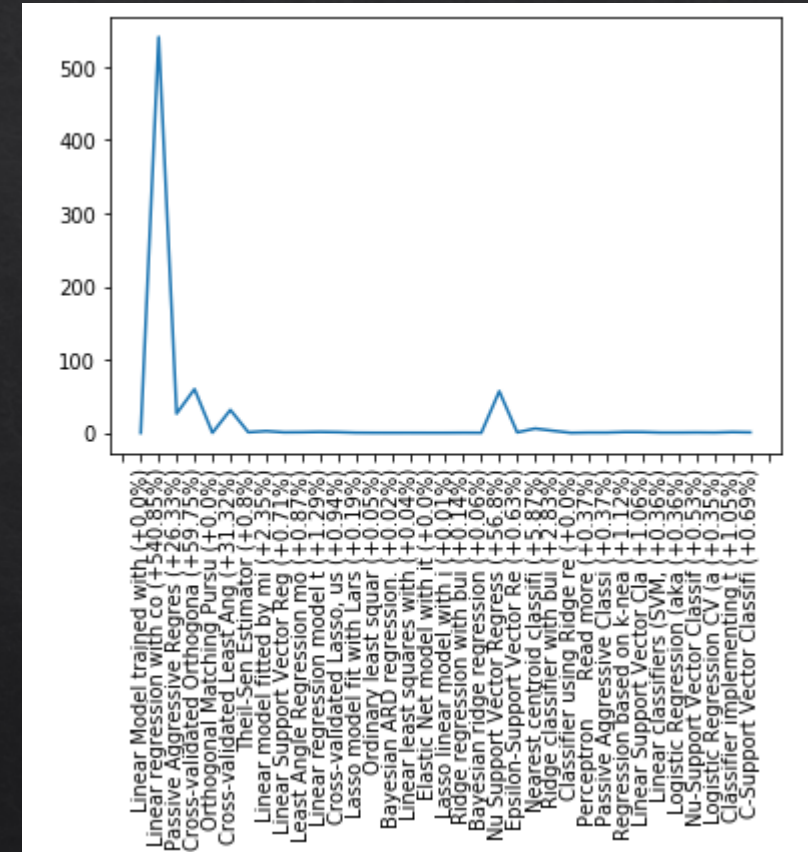
- ◆ Un test de performance a été effectué pour savoir quel algorithme était le plus performant pour la classification des données.

Meilleur algorithme: SVC



Gain de performance

- ◆ Gain de performance d'un algorithme par rapport au précédent (du moins performant au plus)



Quels hyperparamètres pour SVC ?

◇ Possibilités:

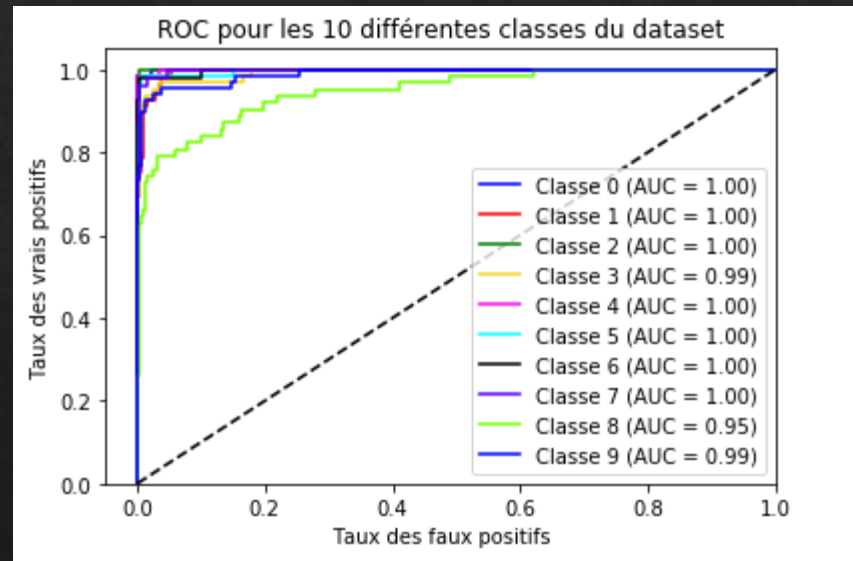
```
{  
  'kernel': ['linear', 'rbf'],  
  'C': [1, 2, 3, 4, 5],  
  'gamma': [1, 0.1, 0.01, 0.001, 0.0001],  
  'probability': [True]  
}
```

◇ Meilleurs paramètres:

```
{  
  'C': 3,  
  'gamma': 0.001,  
  'kernel': 'rbf',  
  'probability': True  
}
```


Courbes ROC

- ◇ Étant donné qu'une courbe ROC ne se fait qu'entre deux classes, ici nous comparons chaque classe aux autres une à une.
- ◇ Par exemple, on fait une courbe ROC de « cette image est un 1 : oui ou non ? »



On remarque qu'en général, l'algorithme ne se trompe que très peu, mais a une faiblesse au niveau de la classification des 8.