

TD N° 1 Machine Learning & Text Mining

Application ML en Reconnaissance des nombres manuscrits

Considérons l'application de machine learning suivante :

```
1. import numpy as np
2. import matplotlib.pyplot as plt
3. import pandas
4. from sklearn.tree import DecisionTreeClassifier
5. data=pandas.read_csv("d:\\train.csv").as_matrix()
6. clf=DecisionTreeClassifier()
7. x=data[0:21000,1:]
8. label=data[0:21000,0]
9. clf.fit(x,label)
10. xtest=data[21000:,1:]
11. actual_label=data[21000:,0]
12. p=clf.predict(xtest)
13. count=0
14. for i in range(0,21000):
15.     count+=1 if p[i]==actual_label[i] else 0
16. print("Accuracy=", (count/21000)*100)
17. d=xtest[5]
18. Nombre_de_pixels_errones=100
19. for i in range(Nombre_de_pixels_errones):
20.     position=np.random.randint(0,784,1)[0]
21.     bruit=np.random.randint(-200,200,1)[0]
22.     d[position]+=bruit
23.     d[position]=d[position]%255
24. print(clf.predict([d]))
25. d.shape=(28,28)
26. plt.imshow(255-d,cmap='gray')
27. plt.show()
```

Travail à faire :

1. Analyser le programme ci-dessus en identifiant (dataset, input, output, technique utilisée,...)
2. Identifier les étapes nécessaires à une application de Machine Learning (training data, testing data, testing....)
3. Quelle est la mesure de performance utilisée ?
4. Quelle est l'objectif des lignes de code (18 à 23)
5. Exécuter et tester le programme