INCERTITUDE DE PRÉDICTION POUR UN PROBLÈME DE CLASSIFICATION SUPERVISÉE

Sommaire

Présentation du sujet 2

Développement d'un réseau de neurones 3

Détection des mauvaises prédictions 4

Méthode ensembliste appliquée pour 3 réseaux 5 – 6

Méthode ensembliste appliquée pour 4 réseaux 7 – 8

PRÉSENTATION DU SUJET

Les réseaux de neurones ont permis de nombreuses avancées dans bien des domaines dépassant les techniques classiques, jusqu'alors utilisées.

Cependant, cette performance est venue à un prix : les décisions prises par ces systèmes manquent d'explicabilité, et l'on peine à associer un degré de confiance à chaque prédiction.

Dès lors, il devient difficile d'appliquer ces méthodes à des problématiques réelles, où chaque décision peut avoir des conséquences dramatiques, par exemple en robotique collaborative.

L'enjeu est donc de trouver des moyens pour garder les performances élevées de ces techniques, tout en associant une mesure de l'incertitude, pour les utiliser plus sereinement.

L'objectif de ce projet est donc de détecter les mauvaises prédictions de réseaux de neurones, préalablement entraînés.

Dans un premier temps, nous développerons un réseau de neurones capable de classifier les images de « fashion_mnist », une base de données que nous nous approprierons.

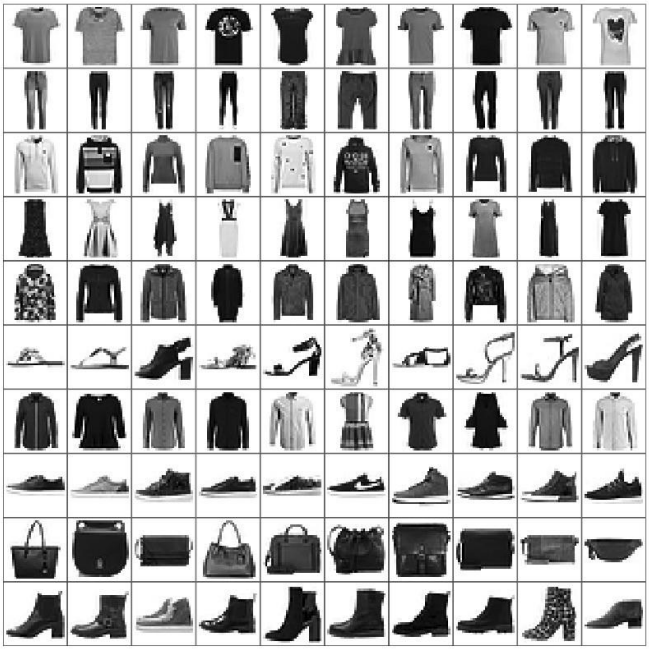
Ensuite, nous évaluerons les performances de ce réseau par divers moyens, tels que les matrices de confusion.

Puis, nous analyserons les résultats de la méthode ensembliste, appliquée ici à 3 réseaux neuronaux.

Enfin, nous développerons la méthode ensembliste pour 4 réseaux neuronaux.

DÉVELOPPEMENT D'UN RÉSEAU DE NEURONES

Nous étudierons la base de données « fashion_mnist », composée de 70 000 images de vêtements : 60 000 images d'entraînement pour 10 000 images de test. Ces images sont en niveaux de gris, de dimensions 28 x 28 pixels, et chacune associée à une étiquette parmi 10 classes. La fonction du réseau neuronal développé est de prédire la classe du vêtement de l'image :



Cette base de données a pour avantages son très grand nombre d'échantillons exploitables par Keras, ainsi que la possible grande précision des réseaux entraînés (précision de 84.6%) :

```
Test loss: 0.40292209386825556
Test accuracy : 0.8470000028610229
```

Les performances du réseau développé peuvent être quantifiées par une matrice de confusion :

```
[[748   1   7  12   5   0 222   0   5   0]
 [  2 968   1  17   8   0   3   0   1   0]
 [  9   0 596   2 308   0  85   0   0   0]
 [ 30   7   7 770 136   0  47   0   3   0]
 [  0   0  26   1 924   0  49   0   0   0]
 [  0   0   0   0   0 929   0  61   1   9]
 [ 71   1  54  13 153   0 704   0   4   0]
 [  0   0   0   0   0   5   0 989   0   6]
 [  4   0   2   2  12   1  15   7 957   0]
 [  0   0   0   0   0   3   1 111   0 885]]
```

DÉTECTION DES MAUVAISES PRÉDICTIONS

Une première méthode de détection des mauvaises prédictions est fondée sur les retours des neurones de sortie (un par classe, donc 10) :

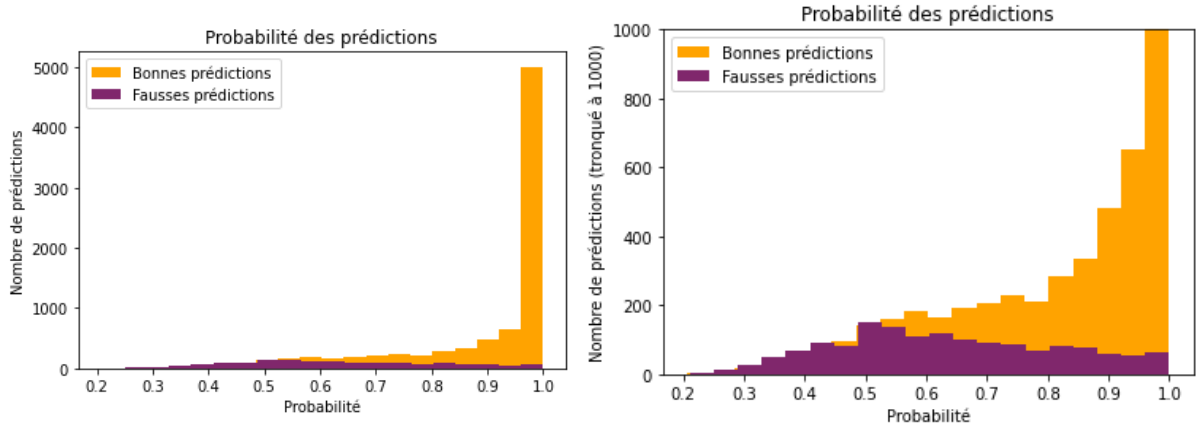
```
[17] print(model.predict(X_test,batch_size=10000,verbose=0)[0])

     #Probabilités estimées par les neurones, d'appartenir à telle ou telle classe

     [3.5699664e-05 2.4015302e-05 6.3317770e-05 2.3991586e-05 1.2837345e-04
      1.5483562e-02 8.1111270e-05 4.0628958e-01 1.8308067e-04 5.7768720e-01]
```

Chacun renvoie une valeur, entre 0 et 1, et la somme des dix valeurs vaut 1. Ces valeurs peuvent être interprétées comme des probabilités d'appartenir à telle ou telle classe (formule des probabilités totales). En fixant une valeur seuil, on peut minimiser (voir supprimer) les mauvaises prédictions.

Ainsi, nous avons visualisé le nombre de bonnes / mauvaises prédictions, selon la probabilité d'appartenir à la classe prédite :



Cependant, il existe des mauvaises prédictions données avec une très grande certitude. Alors, même en utilisant une valeur seuil très élevée (par exemple 0.99), des mauvaises prédictions ne sont pas éliminées, et de nombreuses bonnes prédictions sont rejetées à tort :

```
Matrice de confusion pour seuil de confiance à 0.99 :

[[3889 4581]
 [  20 1510]]
```

Par exemple, pour un seuil de confiance de 99%, 20 mauvaises prédictions n'ont pas été rejetées (soit 1.3 % des mauvaises prédictions), alors que 4581 bonnes prédictions l'ont été (soit 54.1 % des bonnes prédictions).

MÉTHODE ENSEMBLISTE POUR 3 RÉSEAUX NEURONAUX

L'apprentissage ensembliste est l'utilisation de plusieurs algorithmes pour un même problème, afin d'obtenir de meilleures prédictions.

Pour cette étude, nous utiliserons la règle de prédiction suivante, basée sur le fait que la majorité a raison :

a) Si les 3 réseaux donnent la même prédiction, alors cette prédiction sera une prédiction « confiante ».

b) Si 2 des 3 réseaux donnent la même prédiction (l'autre réseau faisant une prédiction différente), alors cette prédiction sera « avec réserve ».

c) Si les 3 réseaux donnent des prédictions différentes, il y a « indécision ».

Les matrices de confusion que nous avons obtenu pour les différents cas sont :

```
[[734   0   4  20   0   0  48   0   4   0]
 [  1 936   0  19   1   0   1   0   1   0]
 [  7   0 673   9  29   0  50   0   1   0]
 [ 10   1   4 883   6   0  16   0   2   0]
 [  0   0  55  35 472   0  61   0   0   0]
 [  0   0   0   0   0 870   0  12   1   3]
 [ 62   0  42  25  16   0 577   0   3   0]
 [  0   0   0   0   0   1   0 896   0  15]
 [  2   0   0   3   0   1   2   3 959   0]
 [  0   0   0   0   0   0   1  17   0 925]]
```

Cas a)

```
[[ 76   0   3  16   0   1  85   0   2   0]
 [  1  16   0  10   2   0   7   0   0   0]
 [  4   0  92   6  33   0  66   0   2   0]
 [  9   1   0  34  13   0  14   0   2   0]
 [  1   0  65  14 226   0  38   0   5   0]
 [  0   0   0   1   0  86   0  10   0  14]
 [ 51   0  23   9  38   0 138   0   7   0]
 [  0   0   0   0   0  15   0  63   0   7]
 [  0   0   2   2   1   1   1   0  21   0]
 [  0   0   0   0   0   2   0  28   0  24]]
```

Cas b)

```
[[ 7]
 [ 5]
 [28]
 [ 5]
 [28]
 [ 3]
 [ 9]
 [ 3]
 [ 2]
 [ 3]]
```

Cas c)

```
[[810   0   7  36   0   1 133   0   6   0   7]
 [  2 952   0  29   3   0   8   0   1   0   5]
 [ 11   0 765  15  62   0 116   0   3   0  28]
 [ 19   2   4 917  19   0  30   0   4   0   5]
 [  1   0 120  49 698   0  99   0   5   0  28]
 [  0   0   0   1   0 956   0  22   1  17   3]
 [113   0  65  34  54   0 715   0  10   0   9]
 [  0   0   0   0   0  16   0 959   0  22   3]
 [  2   0   2   5   1   2   3   3 980   0   2]
 [  0   0   0   0   0   2   1  45   0 949   3]
 [  7   5  28   5  28   3   9   3   2   3   0]]
```

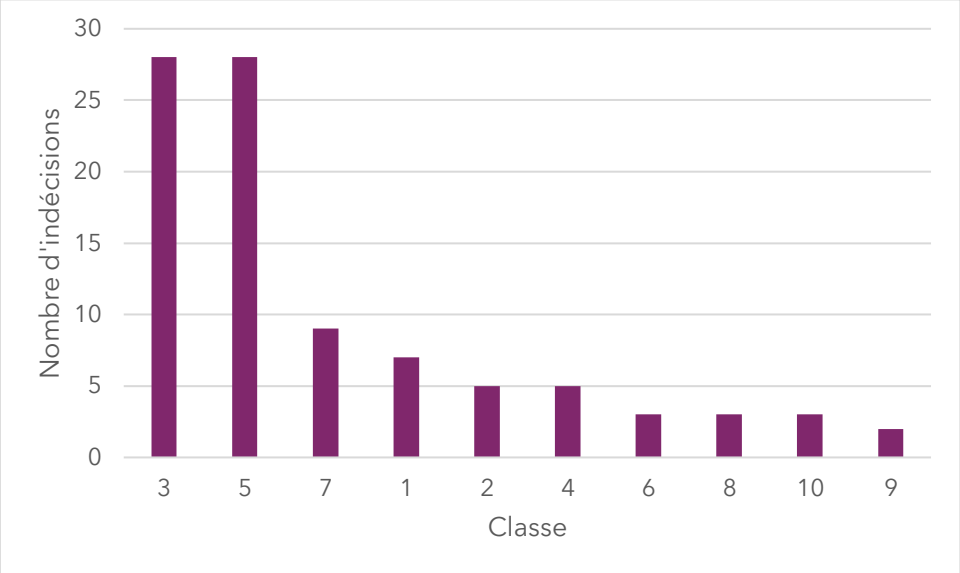
Matrice de confusion globale (avec indécisions)

La somme des cœfficients non diagonaux de la matrice de confusion globale (hors indécisions), et des indécisions est 1299, alors que la somme des cœfficients non diagonaux de la matrice de confusion du premier réseau développé est 1530 > 1299, donc les prédictions de la méthode ensembliste sont plus précises que celles du premier réseau.

La matrice de confusion du cas a) montre qu'il arrive que les 3 réseaux se trompent de la même façon, même si les prédictions sont très majoritairement bonnes (93.03% de précision).

La matrice de confusion du cas b) montre que les prédictions « avec réserve » sont majoritairement bonnes (55.9% de précision).

La répartition des indécisions selon les classes est :



Nous constatons que les classes où les indécisions sont plus nombreuses sont les classes 3 et 5, soit respectivement les classes « Dress » (Robe) et « Sandal » (Sandales).

MÉTHODE ENSEMBLISTE POUR 4 RÉSEAUX NEURONAUX

Nous pouvons également appliquer la méthode ensembliste pour 4 réseaux neuronaux, en utilisant la règle suivante :

a) Si les 4 réseaux donnent la même prédiction, alors cette prédiction sera une prédiction « confiante ».

b) Si 3 réseaux donnent une prédiction commune (le dernier réseau faisant une prédiction différente), la prédiction sera « avec réserve ».

c) Si 2 réseaux donnent une même prédiction, les 2 autres réseaux faisant des prédictions chacune différentes, la prédiction sera « avec réserve ».

d) Si 2 prédictions sont données chacune par 2 réseaux, ou si chaque réseau donne une prédiction différente, il y a « indécision ».

Les matrices de confusion obtenues pour les différents cas sont :

```
[[727   0   4  20   0   0  47   0   4   0]
 [  1 936   0  18   1   0   1   0   1   0]
 [  6   0 673   6  27   0  27   0   1   0]
 [  9   1   4 877   6   0  15   0   2   0]
 [  0   0  54  30 472   0  31   0   0   0]
 [  0   0   0   0   0 870   0   9   1   3]
 [ 62   0  42  22  15   0 523   0   3   0]
 [  0   0   0   0   0   1   0 894   0  15]
 [  2   0   0   3   0   1   0   3 959   0]
 [  0   0   0   0   0   0   1  17   0 925]]
```

Cas a)

```
[[ 77   0   3   9   0   1  39   0   1   0]
 [  1  16   0   1   2   0   0   0   0   0]
 [  5   0  82   3  35   0  28   0   0   0]
 [  8   1   0  25  13   0  10   0   2   0]
 [  0   0  52   7 220   0  30   0   3   0]
 [  0   0   0   0   0  84   0   7   0   8]
 [ 39   0  23   8  38   0  93   0   7   0]
 [  0   0   0   0   0  15   0  49   0   6]
 [  0   0   2   1   1   1   2   0  20   0]
 [  0   0   0   0   0   2   0  19   0  22]]
```

Cas b)

```
[[ 1  0  2  2  1  0  7  0  1  0]
 [ 0  2  2  1  0  0  3  0  0  0]
 [ 0  0 22  3  5  0  7  0  0  0]
 [ 0  0  2  6  0  0  3  0  0  0]
 [ 0  0 15  0 18  0  4  0  1  0]
 [ 0  0  0  0  0  3  0  1  0  0]
 [ 4  0  4  3  2  0 10  0  0  0]
 [ 0  0  0  0  0  2  0  1  0  1]
 [ 0  0  1  0  1  0  1  0  1  0]
 [ 0  0  0  0  0  2  0  0  0  1]]
```

Cas c)

```
[[ 54]
 [ 14]
 [ 70]
 [ 16]
 [ 63]
 [ 14]
 [102]
 [ 16]
 [  1]
 [ 11]]
```

Cas d)

```
[[805   0   9  31   1   1  93   0   6   0  54]
 [  2 954   2  20   3   0   4   0   1   0  14]
 [ 11   0 777  12  67   0  62   0   1   0  70]
 [ 17   2   6 908  19   0  28   0   4   0  16]
 [  0   0 121  37 710   0  65   0   4   0  63]
 [  0   0   0   0   0 957   0  17   1  11  14]
 [105   0  69  33  55   0 626   0  10   0 102]
 [  0   0   0   0   0  18   0 944   0  22  16]
 [  2   0   3   4   2   2   3   3 980   0   1]
 [  0   0   0   0   0   4   1  36   0 948  11]
 [ 54  14  70  16  63  14 102  16   1  11   0]]
```

Matrice de confusion globale (avec indécisions)

La somme des cœfficients non diagonaux de la matrice de confusion globale (hors indécisions), et des indécisions est **1391**.

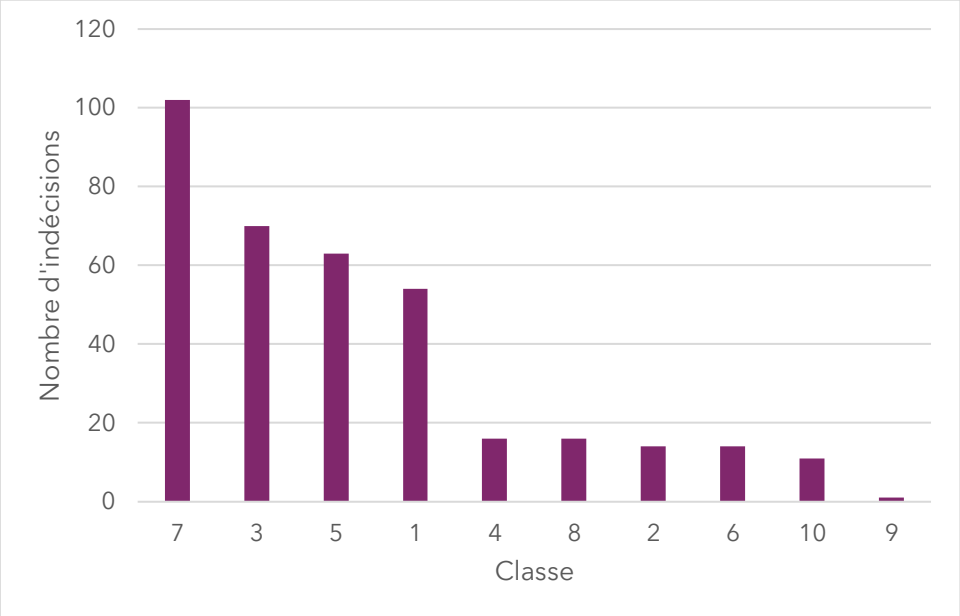
Les prédictions de la méthode ensembliste appliquée à 4 réseaux sont moins précises que celles de la méthode ensembliste appliqué à 3 réseaux (1391 > 1299).

Cependant, elles sont plus précises que celles du premier réseau (1530 > 1391).

La matrice de confusion du cas a) montre qu'il arrive que les 4 réseaux se trompent de la même façon.

Les matrices de confusion des cas b) et c) montrent que les prédictions « avec réserve » sont **majoritairement bonnes** pour les prédictions communes à **3** réseaux, mais **imprécises** pour les prédictions communes à **2** réseaux : la somme des cœfficients non diagonaux représente respectivement **38.6%** et **55.5%** de la somme des cœfficients de la matrice de confusion.

La répartition des indécisions selon les classes est :



Nous constatons que les classes où les indécisions sont plus nombreuses sont les classes 7, 5 et 3, soit respectivement les classes « Sneaker » (Baskets), « Sandal » (Sandales) et « Dress » (Robe).

Les nombreuses indécisions pour les classes 3 et 5 corroborent les indécisions de l'application de la méthode ensembliste pour 3 réseaux, tandis que beaucoup d'indécisions sur la classe 7 semblent s'être créées par l'ajout d'un réseau.