

TP: Classification - BI

Durand JF & Amara Antoine

Exercice : ZOO

Q2.1

Nous avons commencé par construire le tableau suivant :

Caractéristiques	Classification Animale
poils	mammifères
plumes	oiseaux
ponte d'oeufs	oiseaux, insecte, amphibiens, reptiles, poissons, invertébrés
production de lait	mammifères
capacité de voler	insectes, amphibiens, mammifères, oiseaux
Respiration à l'air	Tout sauf poissons
venimosité	reptiles, poissons, insectes, oiseaux
2 pattes	mammifères, oiseaux
4 pattes	mammifères, amphibiens, reptiles
6 pattes	insectes
domesticable	oiseaux, mammifères, amphibiens, poissons, insectes

A partir du tableau précédent nous en avons déduit un arbre pour classer les animaux listés :

- Respire à l'air ?
 - non--> poisson
 - oui--> Plume ?
 - oui--> poule
 - non--> Capacité de voler ?
 - oui--> abeille
 - non--> Poils ?
 - non--> crabe
 - oui--> Carnivore ?
 - non--> antilope
 - oui--> Produit du lait ?
 - non--> ours
 - oui--> chat

Q2.2

Le taux d'erreur est de 100%, le modèle est centré sur un petit échantillon et lui même appliqué sur

un petit échantillon.

Q2.3

De notre point de vue, aucun attributs ne semble gênant à la création du modèle. On parle ici de classification supervisé car nous utilisons des données d'entraînement pour construire le modèle. Une fois le modèle construit grâce à ces données on le test avec de nouvelles instance, qui n'était pas présente lors de la construction du modèle. Cela correspond à un entraînement supervisé car nous avons un oracle(nos données d'entraînement) qui permet d'améliorer au fur et à mesure l'erreur de prédiction de notre modèle. Notre jeu de test permet de valider le modèle. Le principe de séparer données d'entraînement et données de test s'appelle la cross-validation.

Q2.4

Le meilleur algorithme que nous ayons expérimenté est celui de Hoedffing, celui-ci a une très bonne précision: 96% de classification correct et un taux d'erreur de 4%. Nous avons réglé le paramètre de la cross-validation(folds) à 10.

Q2.5

- abeilles
 - plumes = false
 - ponte d'oeufs = true
 - capacité de voler = true
 - respire à l'air = true
- poule
 - plumes = true
 - ponte d'oeufs = true
 - capacité de voler = true
 - respire à l'air = true
- ours
 - carnivore = true
 - poils = true
 - respire à l'air = true
- antilope
 - carnivore = false
 - poils = true
 - respire à l'air = true

Exercice : Titanic

Q3.1

```
Algorithm : J48
```

```
=====
```

```
SEX = male
| CLASS = 1st
| | AGE = adult: no (175.0/57.0)
| | AGE = child: yes (5.0)
| CLASS = 2nd
| | AGE = adult: no (168.0/14.0)
| | AGE = child: yes (11.0)
| CLASS = 3rd: no (510.0/88.0)
| CLASS = crew: no (862.0/192.0)
SEX = female
| CLASS = 1st: yes (145.0/4.0)
| CLASS = 2nd: yes (106.0/13.0)
| CLASS = 3rd: no (196.0/90.0)
| CLASS = crew: yes (23.0/3.0)
```

```
Number of Leaves : 10
```

```
Size of the tree : 15
```

Q3.2 | Q3.3 | Q3.4

Le résultat semble identique quelque soit la méthode de validation choisie.

Q3.5

Titanic Correctly classified instances F-measure

Algo 1	79.055 %	0.765
Algo 2	79.055 %	0.768
Algo 3	78.5552 %	0.683
Algo 4	79.055 %	0.768

Q3.7

On remarque que les résultats présents via les différents algorithmes sont sensiblement identiques.