

# Sorbonne Université

## Examen M2 DAC - REDS

Jeudi 18 février 2021

### Consignes générales

- Documents autorisés.
- Durée : 1h30

## 1 Méthodologie Data Science(8pts)

On s'intéresse dans cet exercice à l'évaluation des performances de différents modèles pour une tâche d'apprentissage statistique. La tâche considérée ici est la classification binaire Homme/Femme à partir d'images. L'évaluation se fera en termes d'accuracy (précision moyenne sur les deux classes) sur une base de test.

### Question 1

D'après le tableau de scores suivants, classer les modèles selon leur intérêt pour réaliser la tâche (en justifiant). Les scores et les écarts-type sont obtenus avec une validation croisée. La difficulté de mise en œuvre est relative au nombre d'hyper-paramètres, à la sensibilité par rapport à leur valeur, à la sensibilité par rapport à l'initialisation d'une descente de gradient. (2 pts)

|                  | Accuracy (%) | Écart-type (%) | Difficulté de mise en œuvre |
|------------------|--------------|----------------|-----------------------------|
| Modèle aléatoire | 50           | 0              | Nulle                       |
| Modèle 1         | 80           | 10             | Élevée                      |
| Modèle 2         | 80           | 10             | Moyenne                     |
| Modèle 3         | 78           | 2              | Moyenne                     |
| Modèle 4         | 52           | 5              | Faible                      |

### Question 2

Voici deux matrices de confusion indiquant les prédictions classe par classe. Quel modèle choisir de préférence ? Expliquer. (2 pts)

| Modèle 5 : | Vérité \ Prédiction | Homme | Femme |
|------------|---------------------|-------|-------|
|            |                     | 70%   | 30%   |
|            | Homme               | 70%   | 30%   |
|            | Femme               | 30%   | 70%   |

| Modèle 6 : | Prédiction |              |
|------------|------------|--------------|
|            | Vérité     |              |
|            | Homme      | 100%         |
|            | Femme      | 40%          |
|            |            | Femme 0% 60% |

**Question 3**

On suppose que la base de train est la suivante. Que peut-on remarquer dans les données ? (1 pt)

**Question 4**

Sur la base de test utilisée lors de la mise au point du modèle les performances sont très bonnes, pourtant, les performances s'effondrent lors d'une utilisation réelle dans une application pratique. Quelles peuvent être les raisons ? (1 pt)

**Question 5**

Commenter les conséquences sociales potentielles de cette situation. (1 pt)

**Question 6**

Que peut-on faire pour éviter ce problème lors de l'apprentissage ? (1 pt)

## 2 Méthodologie Recherche (8pts)

### 2.1 Exercice 1 : Proposition de modèle

Soit l'introduction d'un papier que vous souhaitez écrire : *"Effectively refuting an argument is an important skill in persuasion dialogue, and the first step is to find appropriate points to attack in the argument. In this paper, we aim at detecting attackable sentences in arguments, and more importantly the different parts of sentences which are attackable. This attackability information would help people make persuasive refutations and strengthen an argument by solidifying potentially attackable points."*

**Question 1**

Proposer un modèle neuronal qui pourrait correspondre à cette introduction. Donner les grandes lignes et/ou un schéma associé. Et préciser la fonction de loss associée (de façon formelle ou explicitée avec une phrase). On considérera que vous avez une supervision pour les parties de phrases (attaquables ou non). (3pt)

### 2.2 Exercice 2 : Protocole d'évaluation

Considérons un projet de recherche dont le sujet est le suivant :

*"Our objective is to explore how a driver might complete a productivity task like writing or presentation slide creation while maintaining driving safety".*

Vous êtes en charge de la réalisation de ce projet, tant sur le point de vue des réflexions scientifiques que de la mise en oeuvre pratique. On considère que vous avez tous les moyens possibles à votre disposition. Pas de censure sur les idées!

### Question 1

Quelles sont les questions de recherche associées à ce sujet? (1pt)

### Question 2

Décrire le protocole d'évaluation "idéal" pour ce projet. On attend de vous une description de chaque élément du protocole. (3pt)

### Question 3

Pour chaque question de recherche que vous avez identifié, lister les résultats obtenus (graphiques, tableaux, ...). (1pt)

## 3 Docker (4pts)

Fichier `server.py` :

```
1 import http.server
2
3 PORT = 8888
4 server_address = ("", PORT)
5 server = http.server.HTTPServer
6 handler = http.server.CGIHTTPRequestHandler
7 handler.cgi_directories = ["/app"]
8 httpd = server(server_address, handler)
9 httpd.serve_forever()
```

### Question 1

(1pt) Etablir un fichier Dockerfile afin de construire une image pour un serveur http avec une application Python :

À partir d'une image de base python en version 3.8

Vous indiquerez un nom quelconque de développeur.

Il sera nécessaire :

- d'installer les paquets `python-pip` et `python-dev`
- d'insérer le fichier `server.py` contenant le code
- d'établir un répertoire de travail `app`

avec un fichier `requirements.txt` contenant les modules Python à installer (flask, Flask-Bootstrap, ipython)

le serveur sera accessible sur le port 8888.

Utilisez le template suivant :

```
FROM .....  
MAINTAINER .....  
RUN apt-get update -y && .....  
COPY ./app /app  
WORKDIR ....  
RUN ....  
EXPOSE ....  
ENTRYPOINT [ "python" ]  
CMD [ "server.py" ]
```

**Question 2**

(1pt) Il se trouve que l'ensemble des fichiers établis proviennent d'un git comprenant 1 Go de données et qu'il y a un répertoire data de 10 Go comprenant des textes. Comment restreindre la taille de l'image en question ?

**Question 3**

(0.5pt) Donner la commande très simple en une ligne qui permet de construire l'image exam avec pour tag la version v0.1.

**Question 4**

(0.5pt) Donner la commande très simple en une ligne qui permet de lancer l'image exam version v0.1 en mode daemon avec comme nom exam en exportant le port 8080 de la machine hôte :

**Question 5**

(1pt) Soit la commande docker suivante à traduire en un fichier `docker-compose.yml` :

```
docker run -d --name postgres -e POSTGRES_DB=data -e  
  ➔ POSTGRES_USER=utilisateur -e POSTGRES_PASSWORD=pass -v /  
  ➔ root/bd:/var/lib/postgresql/data postgres
```

Utiliser le template suivant :

```
services :  
  bdd :  
    image : .....  
    container_name : .....  
    environment :  
      .... : ....  
      .... : ....  
      .... : ....  
volumes :  
  - .....
```