

# Examen - Apprentissage Automatique (MASCO 1)

Durée : 2h

## A lire avant de commencer

- L'objet de cet examen est d'évaluer votre capacité à concevoir un problème d'apprentissage automatique ainsi que les conditions permettant de le résoudre.
- Vous serez évalué sur vos connaissances techniques ainsi que sur votre capacité à bien décrire les différents aspects de votre problème et de votre solution. Pensez à être précis, complet et concis.
- Gardez la même organisation que celle qui vous est donnée ci-dessous (même numéros de sections et de sous sections)
- Vous pouvez utiliser le logiciel que vous souhaitez pour rédiger votre examen, pour peu qu'il permette de générer un fichier PDF.
- Vous m'enverrez à la fin de l'épreuve le fichier PDF par courrier électronique, à l'adresse suivante : `alexis.nasr@univ-amu.fr`.  
Le fichier aura pour nom : `examen_apprentissage_votre_prenom_votre_nom.pdf`
- Vous pouvez utiliser les documents que vous désirez pendant l'examen, mais attention au plagiat ! vos copies seront soumises à un logiciel anti-plagiat.
- Il est **STRICTEMENT** interdit de communiquer entre vous lors de l'examen.
- N'oubliez pas de mentionner vos prénom et nom sur votre copie.
- Bonne chance !

## 1 Le problème

- Décrivez le plus précisément possible le problème que vous souhaitez résoudre.
- Précisez en particulier s'il s'agit d'un problème de classification ou de régression.
- Dites pourquoi ce problème a retenu votre attention, vous paraît-il particulièrement important ? intéressant ?
- Est il possible de comparer votre problème avec un des trois projets que vous avez eu à réaliser ?

## 2 Les données

### 2.1 Description Générale

- Décrire de manière générale les données d'apprentissage de développement et de test.
- On ne vous demande pas à ce niveau de détailler la représentation informatique des données.
- Toutes les données que vous utiliserez doivent être décrites ici.

### 2.2 Acquisition

- Décrire comment vous allez acquérir vos données.
- Est ce que vos données sont difficiles à acquérir ?
- Sont elles chères à acquérir ?

- Nécessitent-elles de l'annotation humaine ?
- Décrire les biais<sup>1</sup> possibles pouvant affecter vos données, comment faire pour les atténuer ?
- Pensez vous qu'il vous faut beaucoup de données (d'apprentissage, de test et de développement) pour obtenir des résultats fiables ?

### 2.3 Représentation

- Décrivez maintenant comment vous allez mettre en forme vos données pour les fournir à votre modèle de prédiction.
- Donner un exemple d'entrée-sortie de votre modèle.

## 3 Le modèle

### 3.1 Nature du modèle

- Quel type de modèle allez vous utiliser pour résoudre votre problème ? justifier votre réponse. Dites si plusieurs types de modèles peuvent faire l'affaire et leurs avantages respectifs.
- Quels sont les hyper paramètres de votre modèle (nombre de couches, taille, fonctions d'activation ...). Lesquels vous semblent particulièrement importants ?
- Comment allez vous faire pour déterminer les valeurs de vos hyper-paramètres.

### 3.2 Sur apprentissage

- Votre problème vous paraît-il particulièrement sensible au sur-apprentissage ?
- Quels moyens pensez vous mettre en œuvre pour diminuer le risque de sur-apprentissage ?

#### 3.2.1 Résultats

- Quelle mesure allez vous utiliser pour évaluer vos résultats ?
- Pensez vous obtenir de bons résultats ?
- Pensez vous qu'il s'agit d'un problème difficile ?

---

1. Le terme biais est à prendre ici dans le sens d'un biais statistique, qui apparaît lorsque l'échantillonnage de la population que vous voulez étudier est déséquilibré. Par exemple : vous menez une étude pour étudier l'évolution du régime alimentaire végétarien en France, mais vous choisissez votre population dans une boucherie !