|  | **École nationale Supérieure d'Informatique, Alger**  **2CSSIL2 et 2CSSIQ2 2022/2023**  **Contrôle Final (CF)**  **Machine Learning (ML)** | |
| --- | --- | --- |
| **Dimanche 11 juin 2023 - [9h, 11h] - Partie 1**  *Documents non autorisés (l'enseignant est considéré comme document)*  Il faut répondre sur cette feuille (pages 1 et 2) et la rendre même si elle est vide. | |
|  | |  |
| **Nom + Prénom :** ML team (Aries & Benatchba) | | ⬤ SIQ2 ⬤ SIL2 |

**QCM. Test de connaissance (10pts: 30mn)**

Pour les questions à choix multiple, un choix erroné annule un autre correct. La note minimale par question est 0.

**(1)** Pour chaque description de tâche, choisir le type de son apprentissage le plus adéquat : **(1.25pts = 0.25 \* 5)**

|  | **Supervisé** | **Non supervisé** | **Par renforcement** |
| --- | --- | --- | --- |
| **Un système qui génère un diagramme de classes (DC) à partir d'une description textuelle, ensuite un expert note ce DC. Cette note est utilisée pour régler ce système.** | ◯ | ◯ | ⬤ |
| **Un système qui détecte les attaques réseaux en se basant sur les logs de attaques de plusieurs années précédentes.** | ⬤ | ◯ | ◯ |
| **Un système qui encode un diagramme de classes (DC) sous forme d'un vecteur réduit.** | ◯ | ⬤ | ◯ |
| **Un système qui adapte le routage dans un réseau en se basant sur la latence (délai de transit).** | ◯ | ◯ | ⬤ |
| **Un système qui trouve des groupes similaires sur les réseaux sociaux.** | ◯ | ⬤ | ◯ |

**(2)** Nous avons une petite quantité de données avec deux classes déséquilibrées. Sélectionner toutes les opérations de préparation de données adéquates à ce cas (qui ne causent pas un autre problème) : **(0.75pts = 0.25 \* 3) - err**

| ☒ Collecter plus de données.  ☐ Appliquer sous-échantillonnage.  ☐ Diviser le dataset en 3 (train, val, test). | ☐ Générer plus de caractéristiques.  ☒ Appliquer sur-échantillonnage.  ☒ Diviser le dataset en 2 (train, test) avec validation croisée. |
| --- | --- |

**(3)** Pour chaque algorithme, choisir le type du modèle entraîné : **(1.25pts = 0.25 \* 5)**

|  | **Discriminatif** | **Génératif** | **Dépend de l'architecture** |
| --- | --- | --- | --- |
| **Régression logistique** | ⬤ | ◯ | ◯ |
| **Arbres de décision** | ⬤ | ◯ | ◯ |
| **SVM** | ⬤ | ◯ | ◯ |
| **Naive Bayes** | ◯ | ⬤ | ◯ |
| **Réseaux de neurones** | ◯ | ◯ | ⬤ |

**(4)** Pourquoi la régression logistique est appelée "régression" et pas "classification" ? **(0.5pts)**

| ◯ Puisqu'elle est basée sur la régression linéaire.  ◯ Puisque ces entrées doivent êtres continues. | ⬤ Puisqu'elle estime une probabilité (valeur continue).  ◯ Puisqu'elle calcule une somme pondérée. |
| --- | --- |

**(5)** Nous avons 5 caractéristiques et 4 classes, pour chaque classe 100 échantillons. Calculer le nombre de paramètres (sans biais) et le cumule du nombre des échantillons (si un échantillon est utilisé 2 fois, on le compte 2) pour chaque modèle de régression logistique : **(2pts = 0.25 \* 8)**

|  | **Nombre des paramètres** | **Nombre des échantillons** |
| --- | --- | --- |
| **Binaire (une classe contre les autres)** | 5 | 100 \* 4 = 400 |
| **Multinomiale (Softmax)** | 5 \* 4 = 20 | 100 \* 4 = 400 |
| **One-vs-One** | 5 \* 6 = 30 | 200 \* 6 = 1200 |
| **One-vs-Rest** | 5 \* 4 = 20 | 400 \* 4 = 1600 |

**(6)** Choisir les propositions correctes concernant les arbres de décisions (justes pour ID3 OU CART) :

**(1.25pts = 0.25 \* 5) - err**

| ☒ Peuvent utiliser des attributs nominaux.  ☒ Peuvent utiliser des attributs continus.  ☒ Peuvent avoir un sur-apprentissage. | ☒ Sont représentées comme des conditions SI.  ☒ L'attribut qui divise mieux les échantillons d'un noeud est choisi.  ☐ Les feuilles doivent toujours contenir des classes homogènes. |
| --- | --- |

**(7)** Choisir les propositions correctes concernant les forêts aléatoires : **(1pt = 0.25 \* 4) - err**

| ☒ Basées sur les arbres de décision.  ☒ Basées sur le bootstrapping (bagging).  ☐ Basées sur le boosting. | ☐ Basées sur le stacking.  ☒ Peuvent être puissants même avec de faibles estimateurs.  ☒ Ne sont pas déterministes. |
| --- | --- |

**(8)** Cocher les méthodes qui vérifient ces propositions (FFS=Feature forward selection) : **(1pt = 0.25 \* 4) - err**

|  | **L1** | **ANOVA** | **FFS** | **L2** |
| --- | --- | --- | --- | --- |
| **Considérée comme fonction objectif.** | ☒ | ☐ | ☐ | ☒ |
| **Sélectionne les attributs avant l'entraînement.** | ☐ | ☒ | ☒ | ☐ |
| **Ne peut pas régler le problème de sur-apprentissage.** | ☐ | ☐ | ☐ | ☐ |
| **Réduit le nombre d'échantillons.** | ☐ | ☐ | ☐ | ☐ |

**(9)** Choisir les propositions incorrectes concernant l'algorithme SVM forme duelle : **(0.5pts = 0.25 \* 2) - err**

| ☐ L'estimation SVM-duelle est basée sur les échantillons d'entraînement et leurs classes.  ☐ SVM-duelle ne peut pas être optimisé par la descente du gradient traditionnelle.  ☒ Lors du test, un échantillon situé dans la marge n'a pas de classe.  ☒ Avec soft-margin, un échantillon mal-classé a plus d'erreur lorsqu'il est à l'intérieur de la marge qu'en dehors.  ☐ Le but du SVM-duelle est de trouver une séparation avec une grande marge entre les deux classes. |
| --- |

**(10)** Sélectionner tous les modèles autorégressifs (Vous devez connaître ce terme) : **(0.5pts) - err**

| ☐ Générateur de visages CNN | ☐ Générateur de textes LSTM | ☒ GPT | ☐ BERT |
| --- | --- | --- | --- |

|  | **École nationale Supérieure d'Informatique, Alger**  **2CSSIL2 et 2CSSIQ2 2022/2023**  **Contrôle Final (CF)**  **Machine Learning (ML)** | |
| --- | --- | --- |
| **Dimanche 11 juin 2023 - [9h, 11h] - Partie 2**  *Documents non autorisés (l'enseignant est considéré comme document)*  *Donner plusieurs réponses dont une est juste et le reste n'a pas de sens vaut un 0.* | |

**Exercice 01. Test de reflection et ingénierie (6pts: 1h)**

Un manager d'un restaurant spécialisé dans les plats traditionnels algériens veut améliorer la qualité de service de ce restaurant. Cette dernière est jugée en se basant sur le nombre de clients par jour, le nombre de clients qui reviennent après au moins une semaine et le nombre de clients satisfaits. En plus de ces statistiques, le nombre de plats vendus et de plats retournés sont sauvegardés. Ces nombres dépendent de la saison (un plat chaud peut être plus vendu en hiver qu'en été). Pour avoir des statistiques plus précises, on fournit à chaque client un petit formulaire pour juger chaque plat et son prix : satisfaisant ou non. Il y a des clients qui refusent de participer. De plus, pour chaque plat, le chef sauvegarde des informations sur les ingrédients et la méthode de préparation. Il existe plusieurs types d'ingrédients selon l'unité de mesure : centilitre (cl), gramme (g), nombre (nbr), nombre de cuillères à soupe (c-a-s) et nombre de cuillères à café (c-a-c). Des fois, le chef sauvegarde des ingrédients avec une autre unité. Il existe des ingrédients ayant une mesure max, d'autres non. Le plat possède aussi un/plusieurs mode(s) de cuisson (bouillir, vapeur, frire, etc.).

**A) Préparation de données (2.5pts)**

1. Mentionner les (types de ) caractéristiques qui permettent de prédire si un plat va satisfaire un maximum de clients.

| Saison, ingrédients, mode de cuisson, prix **(0.5pts = 0.25 pour au moins 2 justes + 0.25 pour exactement ces 4 caractéristiques )** |
| --- |

1. Mentionner les opérations de préparation de données sur les différents types de caractéristiques (type de normalisation, représentation et transformations) pour être utilisées avec la régression logistique.

| Saison → OneHot encoding ingrédients → unifier les unités  ingrédients ayant un max → normalisation MinMax  ingrédients (reste) → standardisation (ou coupure puis MinMax)  mode de cuisson → Encodage binaire (chaque mode est utilisé ou non)  **(1.25pts = 0.25 \* 5)** |
| --- |

1. Nous voulons appliquer un classement binaire sur un plat : satisfaisant ou non. Proposer une méthode pour annoter les échantillons en gardant le maximum d'expériences clients.

| * Si somme (Y\_formulaire/client = sat ET clients aime) > seuil → Y = satisfaisant * Sinon Si le plat est très vendu par rapport le nombre total → Y = satisfaisant * Sinon Si le plat est moin vendu par rapport le nombre total → Y = non satisfaisant * Sinon, ignorer l'expérience   **(0.5pts = 0.25 pour le formulaire/condition + 0.25 pour au moins un autre critère)** |
| --- |

1. Le manager veut importer des données sur les plats traditionnels des restaurants génériques. On lui a dit que c'est une mauvaise idée, pourquoi ?

| Les clients d'un restaurant générique ne sont pas là pour déguster des plats traditionnels. Aussi, les méthodes de mesure/statistiques peuvent êtres différentes ce qui va causer des échantillons aberrants.  **(0.25pts)** |
| --- |

**B) Entraînement et validation (2.75pts)**

1. Le manager voulait entraîner un modèle de régression (linéaire, SVM ou arbres de régression) pour estimer le nombre de demandeurs d'un plat. C'est une mauvaise idée, pourquoi ?

| Une de ces raisons : **(0.5pts)**   * La régression nécessite plus de données qu'un classement * Dans ce cas, le but est d'appliquer une interpolation qui est un problème trop complexe. * … |
| --- |

1. En utilisant SVM avec noyau RBF, nous avons remarqué que le modèle a donné de bons résultats par rapport aux autres algorithmes. Proposer deux raisons possibles.

| * Les caractéristiques ne sont pas séparables linéairement (RBF) * Moins d'échantillons et plus de caractéristiques   **(0.5pts = 0.25 \* 2)** |
| --- |

1. Nous avons remarqué que des caractéristiques (comme la quantité d'eau) ne sont pas aussi informatives. Deux modèles de régression logistique ont été entraînés : RL-avec (avec toutes les caractéristiques) et RL-sans (sans les caractéristiques non informatives). J\_train(RL-avec) décroît plus vite que J\_train(RL-sans), J\_val(RL-avec) décroît puis augmente, et J\_val(RL-sans) décroît toujours.
   * Quel est le problème ici ? Pourquoi est-il apparu dans ce cas ? Proposer une solution par rapport à cette expérimentation (des solutions générales ne seront pas acceptées).

| * Problème de sur-apprentissage * Le modèle s'adapte sur des caractéristiques moins importantes * Sélection d'attribut   **(0.75pts = 0.25 \* 3)** |
| --- |

* + Pourquoi J\_train(RL-avec) décroît plus vite que J\_train(RL-avec) ?

| Puisque RL-avec a plus de caractéristiques et donc plus de liberté pour chercher la solution **(0.5pts)** |
| --- |

1. Le modèle Naive Bayes gaussien n'a pas donné de bon résultats. Proposer deux raisons possibles.

| * Les caractéristiques sont dépendantes * Elle ne suivent pas la loi normale   **(0.5pts = 0.25 \* 2)** |
| --- |

**C) Test et déploiement (0.75pt)**

1. Nous sommes intéressés par les plats satisfaisants, est ce que "Accuracy" est bon ? Pourquoi ?

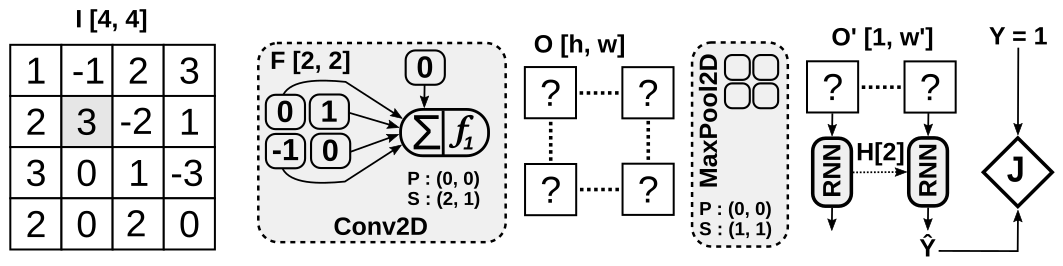
| Non. Puisque dans le classement binaire, nous nous intéressons seulement à la classe positive. Si nous avons trop de vrais négatifs, cette mesure nous donne l'illusion que le modèle est bon.  **(0.5pts = 0.25 + 0.25)** |
| --- |

1. F1 sur le dataset de test était trop bas pour tous les modèles. Mentionner une raison possible.

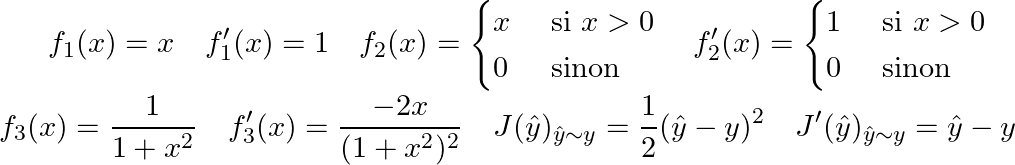
| * Dataset de test non représentatif * Dataset d'entraînement non suffisant ⇒ tous les modèles aurons un sous-apprentissage   **(0.25pts)** |
| --- |

**Exercice 02. Application numérique (4pts: 30mn)**

Voici une architecture composée par une couche Conv2D, une MaxPool2D et une RNN (Many to One).



| **Calculer :**   1. h, w et w' (avec détail). **(0.75)** 2. O (sans détail). **(0.5)** 3. O' (sans détail). **(0.5)** 4. Ŷ et les Ht (sans détail) où H0=[0, 0]. **(0.75)** 5. J(Ŷ)Ŷ∼Y et J'(Ŷ)Ŷ∼Y (∼ veut dire "similaire"). **(0.5)** 6. ; le gradient passé par le vecteur O'. **(0.5)** 7. ; le gradient passé par la cellule I[2, 2]. **(0.5)** |  |
| --- | --- |



**1. Calculer h, w et w' (avec détail)** **(0.75pts = 0.25 \* 3)**

, ,

**2. Calculer O (sans détail)**  **(0.5pts)**

| -3 | -1 | 5 |
| --- | --- | --- |
| -2 | 1 | -5 |

**3. Calculer O' (sans détail)** **(0.5pts)**

| 1 | 5 |
| --- | --- |

**4. Ŷ et les Ht (sans détail) où H0=[0, 0].** **(0.75pts = 0.25 \* 3)**

H1= [0, 1], H2= [1, 1], Ŷ=0.5

**5. J(Ŷ)Ŷ∼Y et J'(Ŷ)Ŷ∼Y (∼ veut dire "similaire").** **(0.5pts = 0.25 \* 2)**

J(Ŷ) = (1-0.5)2/2 = 0.125 J'(Ŷ) = 0.5 - 1 = -0.5

**6.**  **; le gradient passé par le vecteur O'.** **(0.5pts = 0.25 \* 2)**

**7.**  **; le gradient passé par la cellule I[2, 2].** **(0.5pts)**

I[2,2] participe à la génération de O[1,1] et O[1, 2] qui ne participent pas à la génération de O' (ne sont pas des max). Donc, leurs gradients sont tous les deux 0. De ce fait, le gradient passé par la cellule I[2, 2] est aussi 0.

Some nonsense to fill the page:

* **"MLPs propagate errors. Do not be like MLPs."**
* **"CNNs deform others' images. Do not be like CNNs."**
* **"RNNs repeat themselves. Do not be like RNNs."**
* **"Transformers seek attention. Do not be like Transformers."**

