

Nom : UE : IG 201
Prénom : AA : Mathématique
Groupe : Bloc : 2
Nom de l'enseignant : PALERMO O.

Examen de mathématique

/20

Mise en situation

En faisant appel aux différents concepts vus en cours, il sera demandé à l'étudiant de construire un modèle d'apprentissage de type **régression linéaire** sur un jeu de données.

Pour ce faire, plusieurs fichiers lui seront transmis : cet énoncé, un fichier Jupyter Notebook, un fichier de données et un autre le décrivant.

Le thème de l'examen concernera un mollusque marin : [l'abalone](#). L'objectif est de déterminer l'âge de ce mollusque. Pour déterminer son âge, on peut se servir de ces anneaux, mais ce moyen de procéder est fastidieux (cette information est indiquée dans le fichier « description »). Par contre, on peut connaître son âge grâce à ses autres caractéristiques physiques. Le modèle qui sera développé devra utiliser la seconde méthode citée.

Au sujet de l'évaluation

Le fichier qui devra être rendu sera le fichier Jupyter Notebook complété par les soins de l'étudiant. Le nom du fichier devra ressembler à celui-ci « **LR_NOM_PRENOM** » (en l'adaptant, bien sûr). Ce fichier sera considéré comme un rapport où seront notés idées, arguments et résultats. Toute tentative de tricherie ou de recopiage sera pénalisée par une notation de type **fraude**.

Les questions de théorie prendront la forme de justifications. Donc, toutes les instructions qui seront écrites **devront être systématiquement justifiées**. Grâce aux justifications employées, il faudra démontrer le niveau de connaissance de la matière dont il est question. Ci-dessous se trouve la façon dont les justifications impacteront la note des instructions :

- Aucune justification ou justification incorrecte : -60 % ;
- Justification correcte, mais de nombreux détails sont manquants : -40 % ;
- Justification correcte où quelques détails sont manquants : -20 % ;
- Justification correcte et complète : tous les points sont attribués.

Exemple 1 : un étudiant réalise des tests sur son modèle. Cette partie est évaluée sur 2 points. Malheureusement, de nombreux détails manquent, mais son instruction est correcte. Il obtiendra : $2 - 40 \% = 1,2$ point.

Exemple 1 bis : imaginons cette fois que son instruction répond qu'à moitié au problème posé. Il commencera avec la moitié des points. Au niveau de sa justification, c'est cohérent. Toutefois, il manque quelques détails. Il obtiendra : 2 (note de cette partie) - 1 (il répond à moitié) - 20% (quelques détails manquent) = $0,8$ point.

L'étudiant sera évalué sur sa capacité à	Note de la partie
Organiser ses données.	2
Analyser les données grâce aux librairies Pandas et Matplotlib.	4
Déterminer les meilleurs paramètres ou « features » grâce à des tests.	5
Construire un modèle avec les paramètres et/ou les « features » optimaux précédemment trouvés.	2
Expliquer le modèle final obtenu.	4
Prédire des résultats.	3

Travaux à effectuer

Dans le fichier Jupyter Notebook, des instructions seront déjà présentes. Principalement, quelques librairies seront déjà importées¹ et la structure du document sera déjà réalisée.

Ensuite vient l'analyse, il sera nécessaire d'étudier la relation entre les différentes données. Il n'y a absolument aucune restriction pour établir cette partie. Les librairies Pandas et Matplotlib seront des alliées inévitables.

Dès que les données auront été assimilées, le but sera de déterminer les meilleurs paramètres et/ou « features ». Pour y parvenir, des tests devront être opérationnalisés.

Après les avoir trouvés, il sera nécessaire de construire un modèle mathématique et de pleinement l'expliquer.

Enfin, au minimum trois prédictions devront être effectuées.

¹ Si nécessaire, vous pouvez en rajouter d'autres.