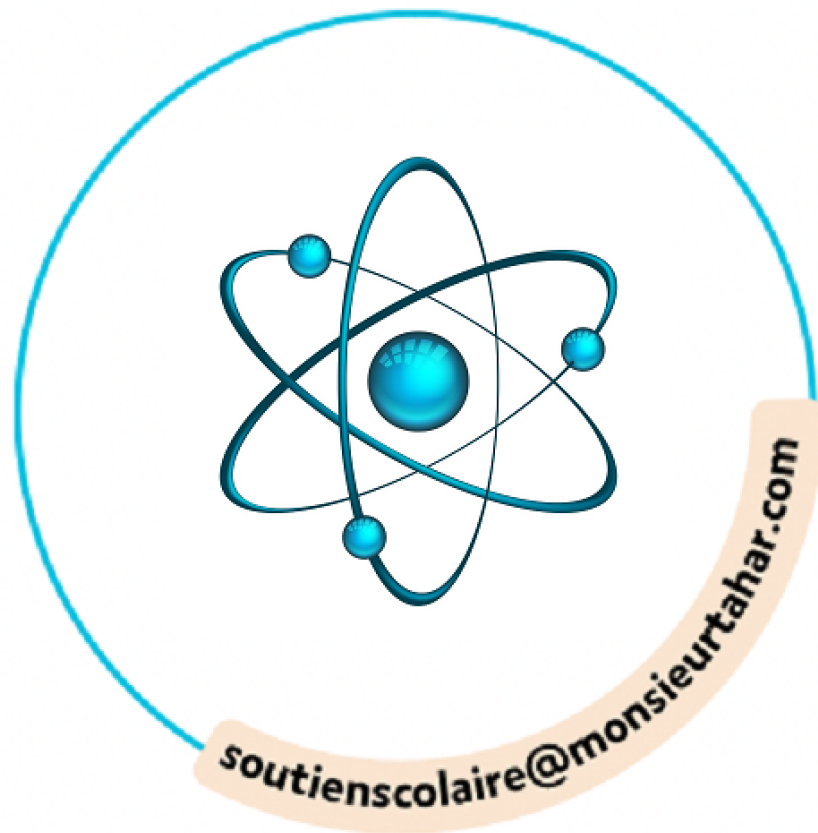


ENSEIGNEMENT SCIENTIFIQUE

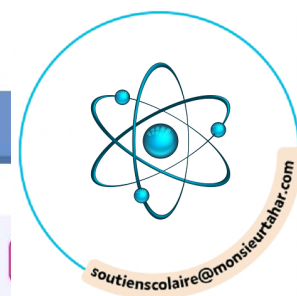


PHYSIQUE- CHIMIE

CHAPITRE 7

Exercices

Tester ses connaissances



QCM

CORRIGÉ p. 246

Pour chaque question, indiquer la (ou les) bonne(s) réponse(s).

	a	b	c	d
1 Au niveau de son processeur, l'ordinateur traite l'information sous forme :	d'un programme de haut niveau.	d'un fichier texte.	de séries de 0 et de 1 (bits).	de langage naturel.
2 En programmation, un bug :	est un dysfonctionnement des composants de l'ordinateur.	est une erreur de logique ou de syntaxe dans un algorithme.	peut aboutir à un résultat incohérent.	engendre systématiquement un message d'erreur.
3 L'intelligence artificielle :	utilise des données d'entraînement pour en extraire un modèle applicable à de nouvelles données.	progresses de manière autonome sans intervention humaine.	est une méthode d'enseignement de l'informatique par des robots.	est basée sur des algorithmes élaborés pour effectuer une tâche complexe.

4 Vrai ou Faux ?

CORRIGÉ p. 246

Indiquer si les affirmations suivantes sont vraies ou fausses.

- « Un fichier son d'un morceau de musique est un fichier exécutable. »
- « Une IA bien entraînée est fiable à 100 %. »
- « Augmenter le nombre de données d'entraînement augmente la fiabilité d'une IA. »
- « Une IA ne peut remplacer l'humain que sur des tâches précises où elle est entraînée ; elle ne peut pas apprendre d'elle-même sans entraînement complémentaire. »

5 Phrases à construire

Construire une phrase à l'aide des mots suivants.

- pixel – octets – couleur – image
- instruction – informations – exécutable – type
- algorithme – coder – IA – modèle – programmation – entraîner

6 Définitions inversées

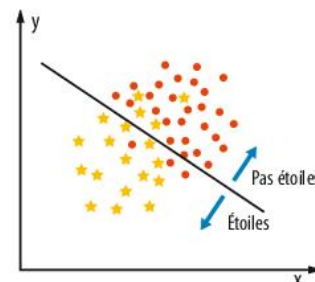
À l'oral !

Retrouver à quels termes correspondent ces définitions.

- Détection erronée d'une condition positive lors d'un test.
- Nombre de prélèvements par seconde lors de la numérisation d'un son.
- Terme anglais désignant l'apprentissage automatisé d'une IA à partir de données.
- Erreur de programmation ou de logique conduisant au mauvais fonctionnement d'un programme.

7 Faux positifs

L'objectif est de sélectionner les étoiles bleues mais pas les ronds rouges. On fixe une droite de séparation. Tous les points situés sous la droite sont considérés par le système comme étant une étoile.



1. Remplir le tableau de contingence suivant :

	Étoile	Rond
Bien classifié		
Mal classifié		

2. Calculer le taux de faux positifs dans cette situation.

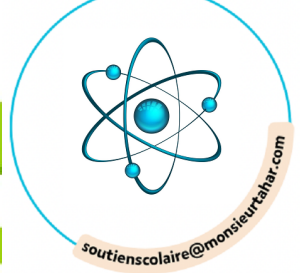
8 Calculs !

La bibliothèque nationale de France (BNF) contient 15 millions de livres. Chacun contient en moyenne l'équivalent de 400 pages de texte numérique.



- Calculer l'espace mémoire (en Go) nécessaire pour stocker l'ensemble des livres de la BNF.
- Calculer la correspondance en heures et minutes pour une vidéo en 720p au format 16/9.
- Calculer la correspondance en heures et minutes d'un enregistrement audio en réalisant l'enregistrement en qualité CD (44 100 Hz – 16 bits).

Exercice résolu



9 Du modèle à son exploitation

Dans une master-class de jazz, on demande aux participants combien de temps ils ont répété un morceau avant de passer devant le jury.

Heures de répétition	Note	Heures de répétition	Note
20	5	5	9
40	20	8	8
30	15	26	14
15	10	32	16
25	11	35	17
22	12	0	2
2,5	6		



1. **Observer** graphiquement s'il y a un lien entre le temps de travail et la note.
2. **Déterminer** l'équation de la courbe de tendance, ici une droite (on la donnera sous la forme $y = ax + b$), avec y représentant la note obtenue d'un participant en fonction du nombre d'heures x de répétition.
3. Lors de la sélection ultérieure, les participants donnent leur temps de répétition. **Prédire** si le musicien réussira la sélection ou pas (on considérera qu'il a échoué pour une note en dessous de 10).
4. Suite à l'examen, on observe si les participants ont été sélectionnés ou non. **Comparer** la prédiction au résultat réel pour remplir un tableau de contingence.
5. **Discuter** la pertinence du modèle pour la prédiction de la réussite.

Participant	Heures de répétition	Prédiction
Marcus	10	échec
Pauline	20	succès
Liam	10	échec
Ginko	30	succès

Conseils

Pour observer une dépendance entre deux variables, on peut tracer un nuage de points et observer leur distribution.

Méthode graphique : on peut tracer une droite au jugé et déterminer graphiquement l'équation.

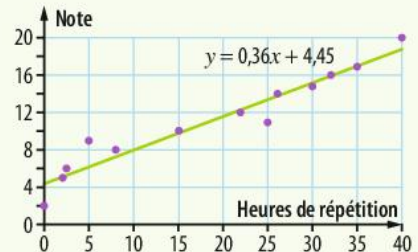
Méthode numérique : à l'aide d'un tableur, on peut déterminer automatiquement la courbe de tendance optimale et son équation (chapitre 12).

On utilise l'équation déterminée précédemment pour faire une prédiction.

Le tableau de contingence permet de connaître la précision d'un modèle, il suffit de compter les prédictions correctes et celles erronées.

Solution

1. On observe que les points sont globalement alignés. Il existe donc une dépendance affine.



2. On obtient l'équation $y = 0,36x + 4,45$.

3. Avec 10 heures de répétition, on prédit un score de $0,36 \times 10 + 4,45 = 8,05$ qui est en dessous de 10. On peut prédire que le participant échouera à la sélection.

Participant	Heures de répétition	Prédiction	Résultat réel	
Marcus	10	échec	échec	vrai négatif
Pauline	20	succès	succès	vrai positif
Liam	10	échec	succès	faux négatif
Ginko	30	succès	échec	faux positif

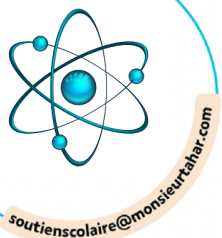
4. Ce qui donne le tableau de contingence suivant :

5. Avec seulement 50 % de résultats pertinents (vrai positifs et négatifs), le test ne peut pas être considéré comme fiable. Cependant la taille de l'échantillon est trop faible et ne permet pas de conclure avec certitude.

	positif	négatif
Vrai	¼	¼
Faux	¼	¼

Exercices

Développer ses compétences



10 Différents symptômes pour une même erreur

Lorsqu'un développeur rédige un programme dans un langage donné, un de ses objectifs est de créer le programme le plus adapté à la finalité souhaitée. Par exemple, le gestionnaire d'une clinique vétérinaire souhaite compter plus facilement le nombre d'animaux présents dans son centre de soin. Il a l'idée de créer une liste L dans laquelle chaque terme correspondra au nombre d'animaux dans une des cages. Il crée ensuite une fonction `total_animaux(L)` qui calcule la somme des termes de cette liste :

```
def total_animaux(L):  
    S=0  
    for k in range(5):  
        S=S+L[k]  
    return S
```

Listes

```
L1 = [8, 4, 2, 0, 3]  
L2 = [5, 2, -3, 6, 0]  
L3 = [1, -2, 6, 3, 2, 4]  
L4 = [4, 2, 9, 5]
```

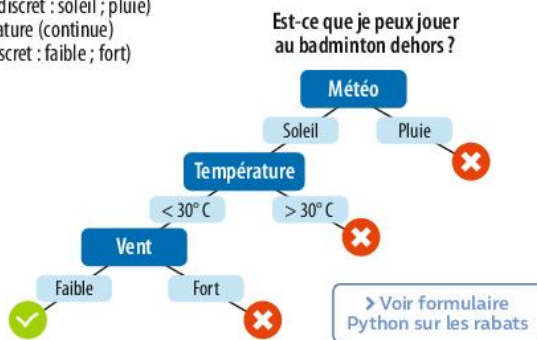
1. **Tester** cet algorithme avec les listes proposées. **Observer** les différentes réponses et **comparer** aux résultats calculés.
2. **Proposer** une explication pour expliquer ces différences.
3. **Proposer** une modification de cet algorithme afin qu'il fonctionne dans tous les cas.

11 Arbre de décision

Algorithme

Un choix est souvent motivé par un ensemble de paramètres qui l'impactent directement. La manière la plus simple de représenter tous les résultats possibles est la représentation par un arbre, chaque embranchement représentant un choix sur la valeur d'un paramètre.

Météo (discret : soleil ; pluie)
Température (continue)
Vent (discret : faible ; fort)



Créer une fonction dans le langage Python permettant de simuler une telle prise de décision.

12 L'apprentissage

À l'oral

Le jeu de société *Mastermind* propose de deviner un code caché de 4 couleurs à l'aide d'une série d'essais/erreurs. Un joueur crée un code, l'autre doit le deviner. À chaque essai, on lui indique si la couleur :

- est bonne et à la bonne position ;
- est bonne mais à la mauvaise position ;
- est mauvaise.



Le joueur doit ajuster sa réponse jusqu'à deviner le code.

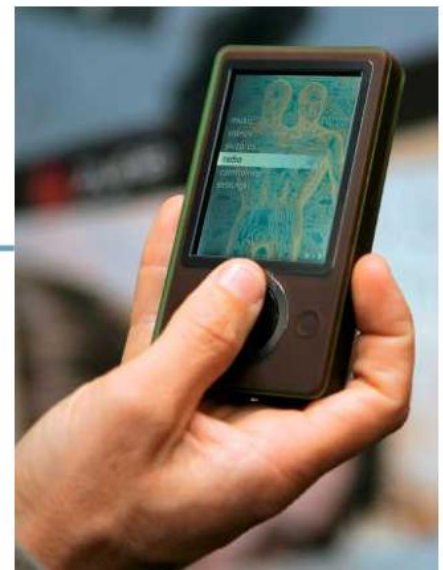
1. **Faire le parallèle** avec le principe de l'apprentissage.
2. **Indiquer** quelles sont les données, quel est l'apprentissage et comment se fait l'ajustement de la réponse.

13 Le bug du Zune

Le *Zune* était un appareil de Microsoft censé concurrencer l'*iPod* de Apple. Le dernier jour de 2008, année bissextile, tous les appareils commercialisés ont soudainement cessé de fonctionner. Le lendemain, après avoir rechargé la batterie, les appareils ont de nouveau pu fonctionner. Le code incriminé avait pour objectif de déterminer l'année en cours en comptant le nombre de jours écoulés depuis le début de l'année. L'algorithme était bloqué dans ce que l'on appelle une boucle infinie. Le code était exécuté perpétuellement mais sans jamais rien produire.

1. **Décrire** ligne par ligne l'extrait de l'algorithme de *Zune*.
2. **Tester** cet algorithme « à la main » avec le 242^e jour d'une année non bissextile, avec le 345^e jour d'une année bissextile.
3. **Expliquer** pourquoi il ne peut pas fonctionner le 366^e jour d'une année bissextile et proposer une modification.

```
Année= année en cours  
Jour= numéro du jour en cours  
Tant que (jour > 365) :  
    Si Année est bissextile :  
        Si Jour > 366 :  
            On enlève 366 à Jour  
            On augmente Année de 1  
        Sinon :  
            On enlève 365 à Jour  
            On augmente Année de 1
```



14 Test de primalité

Prépa E3C

L'étude des nombres premiers joue un rôle très important en arithmétique et en cryptographie. On souhaite ici réaliser un programme qui permet de tester si un nombre est premier ou non.

Les programmes suivants comportent chacun une erreur.

```
def primalite_Lucas(n):
    for k in range(2, n):
        if n%k==0:
            return False
        else:
            return True
```

Programme A

```
def primalite_Jules(n):
    for k in range(1, n):
        if n%k==0:
            return False
        return True
```

Programme B

```
def primalite_Evariste(n):
    for k in range(2, n+1):
        if n%k==0:
            return False
        return True
```

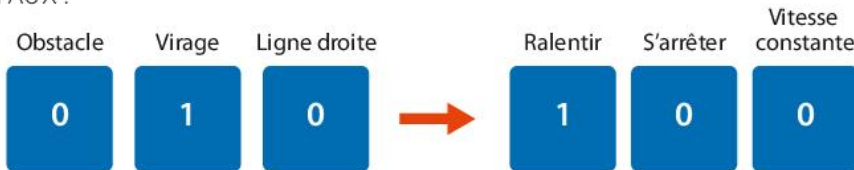
Programme C

1. Créer un jeu de données test et anticiper « à la main » les résultats.
2. Exécuter les programmes à l'aide du jeu de données et identifier puis corriger les bogues éventuels.

15 Entraîner une voiture autonome

On souhaite construire un algorithme capable d'adapter le comportement d'une voiture (ralentir, s'arrêter, maintenir la vitesse), en fonction de différentes situations (obstacle, virage libre, ligne droite libre). On considérera que les situations sont exclusives.

On représente cela par des triplets de bits, avec '1' signifiant VRAI et '0' signifiant FAUX :



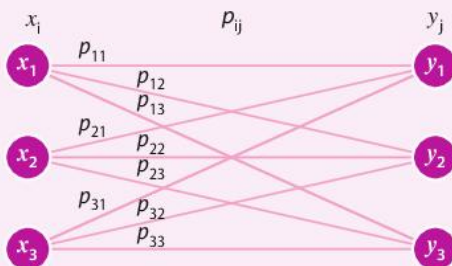
Présence d'un virage

Ralentir



1. À partir de cet exemple, modéliser le comportement de la voiture pour les deux autres situations possibles.
2. On imagine à présent un réseau liant 3 valeurs en entrée à 3 valeurs en sortie tel que :
On a par exemple $y_1 = x_1 \cdot p_{11} + x_2 \cdot p_{21} + x_3 \cdot p_{31}$.

Environnement Poids du réseau Comportement de la voiture



À l'aide du comportement modélisé précédemment, calculer les valeurs manquantes dans le tableau.

Poids	Valeur
P_11	0
P_12	
P_13	
P_21	1
P_22	
P_23	
P_31	0
P_32	
P_33	

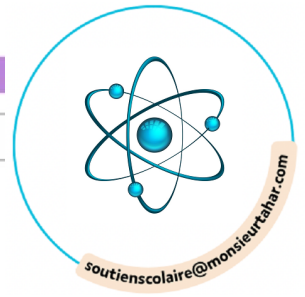
3. Indiquer si pour une situation non prévue : un panneau stop dans un virage libre, on peut utiliser le modèle précédent.
4. Représenter de même un réseau permettant de choisir entre deux activités (pique-nique ou cinéma) en fonction de la météo (soleil, pluie).

16 Précision ou automatisation ?

À l'oral

Tâche	Priorité
Détection de spams	
Diagnostic d'un cancer	
Identification faciale pour déverrouiller un téléphone portable	
Compréhension d'une commande vocale pour une enceinte connectée	

Considérer les cas du tableau et discuter s'il vaut mieux prioriser la précision du système (minimiser les faux-négatifs) ou bien son automatisation (maximiser les vrais-positifs). Justifier.



17 Résultats d'une recherche

Léa souhaite écouter des musiques de Jean-Jacques Goldman. Toutefois, ses musiques sont mélangées, elle doit donc utiliser l'outil de recherche pour retrouver uniquement les musiques de ce chanteur.

Léa fait plusieurs recherches successives à l'aide de différentes chaînes de caractères :

- Jacques ;
- Gold ;
- Goldman.



1. Pour chacune des recherches, **établir** un tableau de contingence.
2. **Déterminer** d'autres chaînes de caractères qui retournent 100 % des musiques de Jean-Jacques Goldman sans celles d'autres artistes. **Indiquer** le tableau de contingence dans ce cas.

Recherche dans : Dossier de ...

Nom



Goldman - Là-bas.mp3



Jacques Brel - Ne me quitte pas.mp3



Jacques Dutronc - L'opportuniste.mp3



Jean-Jacques Goldman - Comme toi.mp3



Jean-Jacques Goldman - Encore un matin.mp3



Jean-Jacques Goldman - Je marche seul.mp3



Neil Young - Heart of Gold.mp3



Sting - Fields of Gold.mp3

18 Efficacité d'un test médical

Prépa E3C



Un laboratoire pharmaceutique élabore un test de dépistage d'une anomalie génétique qui touche environ 5 % d'une population et souhaite mesurer l'efficacité du test. Pour cela, il choisit un échantillon représentatif de 1 000 personnes dont le génome a été donné : 50 sont porteuses de l'anomalie et 950, dites saines, ne le sont pas.

En faisant passer le test aux 50 personnes atteintes, 42 ont été dépistées correctement.

Pour les 950 personnes saines, 810 ont effectivement eu un résultat négatif au test.

1. **Donner** le tableau de contingence (en fréquences) associé à ce test.
2. Une fois ce test commercialisé, un hôpital demande à 500 personnes issues d'une population à risque de se soumettre au test (ces personnes ne font pas partie de l'échantillon d'étude ci-dessus). Parmi elles, 80 % ressortent avec un résultat positif au test et 20 % avec un résultat négatif. D'après le tableau de contingence obtenu à la question 1, **estimer** combien d'entre elles sont effectivement atteintes de l'anomalie.

Indication : ne pas oublier les faux négatifs

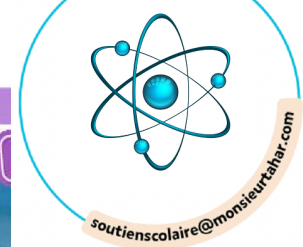
19 L'IA, un modèle évolutif

Prépa E3C

On souhaite entraîner un algorithme d'intelligence artificielle de sorte que, lorsqu'on lui présente une photo, il puisse dire si cette photo représente un chien ou non. Après avoir fourni 20 premières photos d'entraînement, les résultats de l'IA sont les suivants :

- sur 10 photos présentant des chiens, l'IA a eu raison pour 7 d'entre elles ;
- sur 10 photos ne présentant pas de chien, il s'est trompé 2 fois.

1. **Donner** le tableau de contingence associé à cette expérience à cette étape de l'entraînement.
2. On présente une nouvelle photo à l'IA qui prétend alors qu'il y a un chien. **Calculer** la probabilité que ce résultat soit correct (autrement dit, la fiabilité de cette réponse).
3. En réalité cette nouvelle photo ne présentait pas de chien. L'IA s'est donc trompée et adapte alors son modèle grâce à cette 21^e photo. **Donner** le nouveau tableau de contingence obtenu.
4. On présente une 22^e photo à l'IA et celle-ci prétend une nouvelle fois qu'il y a un chien. **Indiquer** la probabilité que ce résultat soit correct.



Une IA judiciaire est-elle forcément impartiale ?

COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) est un logiciel employé dans plusieurs états des États-Unis pour aider aux prises de décision judiciaires. Ce logiciel a été conçu pour déterminer le risque de récidive chez les prévenus.

Utilisant l'intelligence artificielle, COMPAS a subi une phase d'apprentissage à partir d'archives judiciaires. Son algorithme a ainsi mis en relation le taux de récidive de condamnés avec 5 critères : le sexe, l'âge, la nature du crime, les antécédents judiciaires et le degré de préméditation.

Pour tout nouveau prévenu, COMPAS est ainsi en mesure de fournir un risque de récidive en utilisant les 5 critères d'évaluation.

Le système judiciaire américain étant fréquemment accusé de discrimination raciale, une équipe de chercheurs a cherché à déterminer si COMPAS est lui aussi biaisé.

Allégorie de la Justice



1 Risques donnés par COMPAS

Les risques donnés par COMPAS ont été recueillis pour 7 214 anciens prévenus, puis mis en relation avec le fait qu'ils avaient effectivement récidivé ou non, ce qui a permis d'établir des tableaux de contingence selon ce modèle :

		Risque déterminé par COMPAS	
		risque faible	risque élevé
Évolution du prévenu	pas de récidive	VN vrai négatif	FP faux positif
	récidive	FN faux négatif	VP vrai positif

Ces données permettent de tester la fiabilité de l'algorithme au moyen de deux indicateurs :

- le taux de prédiction correcte s'il y a une récidive :
sensibilité = $VP / (VP + FN)$
- le taux de prédiction correcte s'il n'y a pas de récidive :
spécificité = $VN / (VN + FP)$

2 Traitements des données pour deux groupes ethniques

Le recueil et le traitement de données ethniques dans le domaine judiciaire est légal aux États-Unis, aussi des tableaux de contingence ont pu être établis pour deux groupes ethniques :

Groupe A		Risque déterminé par COMPAS	
		risque faible	risque élevé
Évolution du prévenu	pas de récidive	990	805
	récidive	532	1369

Groupe B		Risque déterminé par COMPAS	
		risque faible	risque élevé
Évolution du prévenu	pas de récidive	1139	349
	récidive	461	505

Les résultats de l'étude ont été contestés par les concepteurs de COMPAS, mais des recherches menées par d'autres équipes ont par la suite confirmé que le logiciel est très imparfait.

Calculer l'efficacité de l'algorithme à l'aide de deux indicateurs :

- taux de réussite de prédiction d'une récidive
 $TRP_R = VP / (VP + FN)$
- taux de réussite de prédiction d'une non-récidive
 $TRP_{NR} = VN / (VN + FP)$

À VOTRE AVIS

Voir Esprit scientifique
➤ esprit critique p. 6-11

- L'efficacité est-elle très différente pour les deux groupes ?
- Cela signifie-t-il que l'algorithme a une fiabilité équitable vis-à-vis des deux groupes ?