

MAPSI – Examen final – 30pts

Durée : 2 heures

*Seuls documents autorisés : Calculatrice, antisèche recto-verso, tables de lois.
– Barème indicatif –*

Exercice 1 – Betterave Bio... ou pas (7pts)

Une coopérative de betterave regroupant 100 parcelles fait une expérimentation sur la culture biologique. 10 parcelles sont cultivées en Bio, sans pesticide ni engrais chimique tandis que les autres parcelles restent en agriculture traditionnelle.

	Bio	Traditionnelle
Haut rendement (en moyenne 1.5T / parcelle)	6	80
Bas rendement (en moyenne 1T / parcelle)	4	10

TABLE 1 – Tableau de répartition des 100 parcelles selon leur mode de culture et leur rendement.

Q 1.1 (1pt) Quelles sont les deux variables aléatoires étudiées et leurs modalités ? Y a-t-il plus de chance d'avoir des haut rendements en agriculture biologique ou traditionnelle ? Indiquer les probabilités que vous comparez.

Q 1.2 (1.5pt) Nous prendrons un prix de vente de 1 pour une 1 tonne de Betterave issue de l'agriculture traditionnelle. Le prix de vente de la betterave bio est 25% supérieur à celui de la betterave traditionnelle. En l'état actuel de l'expérimentation et en assimilant le gain au chiffre d'affaire, est-il plus rentable de cultiver de la betterave bio ou traditionnelle sur une parcelle ? Indiquer clairement les étapes de votre raisonnement.

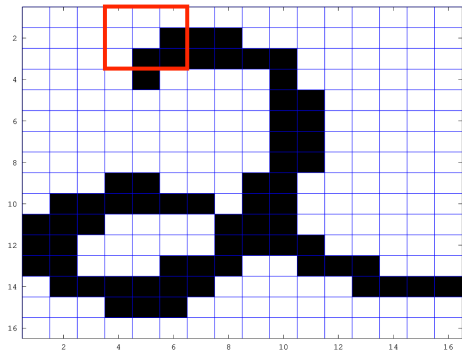
Q 1.3 (2pt) Peut-on affirmer que le rendement d'une parcelle est lié¹ à son mode de culture avec une confiance de 90% ? Le résultat reste-t-il le même à un niveau de confiance de 99% ?

Q 1.4 (0.5pt) Calculer l'espérance du rendement, tous types de parcelles confondus (en T/parcelle). Donner la formule littérale du calcul puis l'application numérique.

Q 1.5 (2pts) Au café du coin, chiffres à l'appui, les anciens affirment que les parcelles donnaient 1.5T en moyenne, avec un écart type $\sigma = 0.2$ T/parcelle. En prenant nos 100 parcelles comme un échantillon iid, peut-on affirmer que le rendement a baissé cette année avec une confiance de 95% ?

1. Lié = non indépendant

Exercice 2 – Reconnaissance de chiffres dans un cadre bruité (8pts)



Notations : \mathbf{x} est une image binaire composée de 16x16 pixels. On note $X = \{\mathbf{x}^1, \dots, \mathbf{x}^N\}$ un ensemble de N images de chiffres manuscrits, où chaque image est décrite par : $\mathbf{x}^i = \{x_1^i, \dots, x_{256}^i\}, x_j \in \{0, 1\}$. On dispose des étiquettes y^i associées à toutes ces images (les dix chiffres sont présents dans la base).

Ces images étant particulièrement bruitées, on décide de travailler par fenêtre de 3x3 pixels glissant sur l'image. Pour chaque fenêtre, on compte le nombre de pixels noirs (3 dans l'exemple ci-contre). Notre image se trouve donc maintenant représentée par des descripteurs $f_j \in \{0, 1, \dots, 9\}$. Chaque image est constituée de 14x14=196 descripteurs f_j .

Q 2.1 (1pt) Modélisation d'un descripteur. On étudie une observation f_j donnant le nombre de pixels noirs dans la fenêtre à une position j . Quelle loi classique permet de modéliser la variable aléatoire F_j dont cette observation est tirée? Donner l'expression de $p(F_j = f_j)$.

Q 2.2 (1pt) Quels vont être les paramètres à apprendre dans ce problème? Combien y aura-t-il de paramètres à apprendre au total pour faire marcher un classifieur de chiffres?

Q 2.3 (1.5pt) Modélisation. On fait l'hypothèse simplificatrice que chaque variable descriptive F_j est indépendante des autres. Donner la vraisemblance d'une image puis la vraisemblance de X . Indiquer brièvement ce que vous pensez de l'hypothèse d'indépendance des F_j .

Notation : le descripteur j de l'image i sera noté f_j^i

Q 2.4 (2.5pts) Donner l'expression des paramètres optimaux et la démonstration du calcul.

Q 2.5 (1pt) Comment faire pour exploiter ces paramètres optimaux pour classer une nouvelle image \mathbf{x} ? Indiquer la procédure détaillée et les calculs à effectuer.

Q 2.6 (1pt) Les fournisseurs de ces bases nous indiquent que les classes de données ne sont pas équiprobables : ils nous donnent 10 paramètres π_0, \dots, π_9 donnant la fréquence de chaque chiffre dans la base... Cette information change-t-elle la définition du classifieur optimal? Dans l'affirmative, réviser la procédure en conséquence –en indiquant toujours le détail des calculs à effectuer–.

Exercice 3 – Prédiction de consommation d'électricité (6pts)

Prédire la consommation électrique est critique pour ajuster au mieux la production et éviter les pertes. Nous disposons d'une base de données de la consommation des 100 jours précédents avec un relevé toutes les heures : $E = \{e_1, \dots, e_{2400}\}$.

Nous choisissons un modèle auto-régressif, qui prédit la consommation à l'instant t en fonction des consommations observées sur les p pas de temps précédents (+ un résidu noté ε) :

$$X_t = \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t, \quad \text{Dans ce problème, on prendra donc : } \hat{e}_t = \sum_{i=1}^p \varphi_i e_{t-i}$$

Q 3.1 (2.5pts) Nous voulons optimiser la prédiction de consommation au sens des moindres carrés. Formaliser cette fonction coût, la dériver par rapport à φ_i et montrer que cette dérivée est de la forme : $\frac{\partial C}{\partial \varphi_i} = \sum_j (\varphi_j \alpha_{i,j}) - \beta_i$. Identifier les coefficients α et β .

Q 3.2 (1.5pt) Montrer que l'optimisation de ce problème correspond à la résolution d'un système d'équations linéaires. Formuler ce problème sous forme matricielle en détaillant les dimensions des matrices et les coefficients en jeu. Vous pourrez simplifier les écritures en introduisant les variables $\gamma_{ij} = \sum_t e_{t-i} e_{t-j}$.

Q 3.3 (1pt) En remarquant que la même translation sur i et j ne change pas la valeur de γ_{ij} (ie $\gamma_{i,j} = \gamma_{i-2,j-2}$), il est possible de simplifier la notation indicielle en introduisant des $\gamma_k, k = i - j$. En remarquant aussi que $\gamma_k = \gamma_{-k}$, combien de termes uniques faut-il calculer pour poser le système d'équations linéaires?

Q 3.4 (1pt) Au terme d'une campagne d'expériences portant sur la détermination du paramètre p , nous avons montré que plus p est grand, plus l'erreur de prédiction sur E est petite... Ce phénomène vous semble-t-il logique? Quelle erreur avons-nous faite? Proposer un protocole expérimental corrige ce défaut.

Exercice 4 – Recommandation. (*Bayesian Personalized Ranking*) (9pts)

Nous allons construire un système de recommandation basé sur l'algorithme *Bayesian Personalized Ranking*². C'est une stratégie assez proche de l'approche Netflix dans laquelle nous devons être capable de prédire si un utilisateur va aimer un film ou pas en fonction de ce qu'il a aimé dans le passé (et de ce que les autres utilisateurs ont aimé dans le passé).

Dans un tel système, chaque utilisateur u est modélisé par un vecteur $\mathbf{z}_u \in \mathbb{R}^K$ et chaque film (ou item) par un vecteur $\mathbf{z}_i \in \mathbb{R}^K$. K donne la dimension des vecteurs (typiquement entre 50 et 300). Les auteurs de l'article original proposent d'utiliser la fonction logistique pour estimer l'affinité entre un utilisateur \mathbf{z}_u et un item \mathbf{z}_i . Ainsi, en notant $A_{ui} \in \{0, 1\}$ la variable binaire modélisant l'affinité, nous obtenons :

$$P(A_{ui} = 1 | \mathbf{z}_u, \mathbf{z}_i) = \frac{1}{1 + \exp(-\mathbf{z}_u \cdot \mathbf{z}_i)}, \quad \mathbf{z}_u \cdot \mathbf{z}_i = \sum_{k=1}^K z_{u,k} z_{i,k}, \text{ produit scalaire classique} \quad (1)$$

Le client (par exemple Netflix) vous fournit une matrice $A \in \{0, 1\}^{N_u \times N_i}$ où chaque utilisateur est décrit sur une ligne et chaque item sur une colonne. Chaque case contient un 1 si l'item a été apprécié par l'utilisateur et 0 sinon (pas vu ou pas aimé). Dans la suite, nous notons a_{ui} les valeurs observées dans A (à la ligne u et la colonne i).

Q 4.1 (0.5pt) Quelle loi peut servir à modéliser A_{ui} ?

Q 4.2 (1pt) Exprimer la vraisemblance de $P(A_{ui} = a_{ui} | \mathbf{z}_u, \mathbf{z}_i)$ en fonction de $P(A_{ui} = 1 | \mathbf{z}_u, \mathbf{z}_i)$ et a_{ui} . Vous utiliserez la *truc* classique permettant de prendre en compte les cas $a_{ui} = 0$ et $a_{ui} = 1$ dans la même expression.

Q 4.3 (2.5pt) Exprimer la log-vraisemblance de l'ensemble des valeurs observées dans la matrice en faisant l'hypothèse qu'elles sont toutes indépendantes.

Montrer que le résultat peut s'écrire sous la forme : $\log \mathcal{L} = \sum_{u,i} \alpha \log(1 + \exp(-\mathbf{z}_u \cdot \mathbf{z}_i)) + \beta_{u,i}(\mathbf{z}_u \cdot \mathbf{z}_i)$ en identifiant α et β .

Q 4.4 (1pt) Rappeler quels sont les paramètres à optimiser et leur nombre.

Q 4.5 (2pts) Calculer le gradient de la log-vraisemblance par rapport au paramètre $z_{u,k}$, correspondant à la k ème composante du vecteur représentant un individu \mathbf{z}_u .

Q 4.6 (1pt) Rappeler comment exploiter ce gradient pour maximiser la vraisemblance. Comment procéder pour optimiser tous les paramètres du systèmes ?

Q 4.7 (0.5pt) Comment évaluer ce système avec une fiabilité satisfaisante ?

Q 4.8 (0.5pt) Pour un client de la base initiale (dont on connaît un certain nombre de préférences), comment exploiter ce système en inférence dans un système opérationnel ?

Q 4.9 (Bonus) Si la dimension K de l'espace de représentation des utilisateurs et des items est faible, à votre avis, serait-il envisageable de faire émerger une signification associée à l'une des dimension k ? Pourquoi ?

2. BPR : Bayesian Personalized Ranking from Implicit Feedback, *S. Rendle, C. Freudenthaler, Z. Gantner and L. Schmidt-Thieme*