Exercice 1 (Rappel sur la régression linéaire)**.** On se propose d'étudier un problème de régression a priori intraitable. La variable observée est Y et on souhaite expliquer cette variable par un ensemble de co-variables X_1,...,X_p. Le modèle linéaire est donc :

$$Y = a_0 + a_1 X_1 + \cdots + a_p X_p + \varepsilon, \tag{1}$$

où ε est une variable gaussienne à valeurs dans $\mathbb{R}$ $\mathcal{N}(0, \sigma^2)$. Il s'agit donc d'estimer les coefficients réels a_0, a_1, ..., a_p. Pour cela, on effectue plusieurs mesures et on obtient autantde variables que d'expériences réalisées, c'est-à-dire :

$$Y_i = a_0 + a_1 X_{i,1} + \cdots + a_p X_{i,p} + \varepsilon_i \qquad i = 1, \ldots, n. \tag{2}$$

On note $Y = [Y_1, \ldots, Y_n]^t$, $\varepsilon = [\varepsilon_1, \ldots, \varepsilon_n]^t$ et $X = [(X_{ij})_{i=1,\ldots,n\ j=0,\ldots,p}]$ avec $X_{i,0} = 1$ quelque soit i et on oublie l'ancien Y et l'ancien ε dont on ne se resservira plus par la suite car ce qui nous interesse est évidemment d'estimer les $a_0, a_1, \ldots, a_p$. Les équations (2) deviennent alors avec ces notations :

$$Y = X\, a + \varepsilon. \tag{3}$$

1. Rappeler la formule donnant l'estimateur au sens des moindres carrés.
2. Cet estimateur au sens des moindres carrés est solution du problème d'optimisation :

$$\min_{a \in \mathbb{R}^{p+1}} \|Y - Xa\|_2^2, \tag{4}$$

   où $\|\cdot\|_2$ dénote la norme euclidienne. Calculer la matrice hessienne de cette fonction à minimiser. Ce problème d'optimisation est-il convexe ? Pouvez vous retrouver la formule de l'estimateur au sens des moindres carrés en résolvant ce problème d'optimisation ?

   Le problème que l'on souhaite aborder après ces rappels est celui de trouver les coefficients $a_0, \ldots, a_p$ dans le cas où $p \geq n$ voire bien plus grand que n. Ce cas arrive en fait très souvent en pratique et est donc impossible à traiter par la minimisation de (4) car la matrice hessienne est dans ce cas dégénérée, c'est-à-dire que son noyau est non réduit à $\{0\}$.
3. Pourquoi cette dernière remarque est-elle vraie ? Combien y'a-t-il de solution dans ce cas au problème (4) ?

Exercice 2 (Régression Ridge)**.** Pour parer au problème soulevé dans l'exercice précédent, deux approches sont possibles. La première s'appelle la ridge regression et consiste à chercher a comme le vecteur qui minimise :

$$\|Y - Xa\|_2^2 + \lambda \|a\|_2^2. \tag{5}$$

1. Ce problème est-il convexe ? admet-il une solution unique ?
2. Implanter cette solution à l'aide de la fonction `quapro` pour les données "detroit" qui donnent le nombre d'homicides en fonction de divers paramètres dans la ville de Detroit de 1961 à 1970. Faites varier le paramètre λ et affichez sur la même figure l'évolution jointe des composantes de a en fonction de λ.

Exercice 3 (Méthode du LASSO)**.** Une autre approche pour le problème précédent consiste à utiliser le LASSO, une technique dont le succès est dû à l'étude mathématique et à la promotion ultérieure qu'en a fait Robert Tibshirani de l'université de Stanford en Californie. Cette méthode consiste à remplacer la pénalisation quadratique par une pénalisation utilisant la norme l_1 du vecteur a, et revient donc à résoudre le problème :

$$\min_{a\in\mathbb{R}^{p+1}} \|Y - Xa\|_2^2 + \lambda\|a\|_1. \tag{6}$$

1. Montrer de manière intuitive que ce problème peut se mettre sous la forme :

$$\min_{a,\, u\in\mathbb{R}^{p+1}} \|Y - Xa\|_2^2 + \lambda\sum_{k=1}^{p} u_k \text{ sous les contraintes } -u_k \le a_k \le u_k \quad k = 1,\dots,p. \tag{7}$$

2. Implémenter cette solution à l'aide de la fonction quapro pour les données "detroit" qui donnent le nombre d'homicides en fonction de divers paramètres dans la ville de Detroit de 1961 à 1970. Faites varier le paramètre λ et affichez sur la même figure l'évolution jointe des composantes de a en fonction de λ.

A remarquer : à partir d'une valeur assez grande de λ, certaines composantes de a deviennent nulles. Ainsi, la méthode effectue un choix automatique des variables importante pour expliquer le phénomène observé. On résoud alors ce qu'on appelle la question du choix de modèle d'une manière globale. L'étude théorique de cette méthode est actuellement en pleine effervescence (voir par exemple l'article http ://www.acm.caltech.edu/ emmanuel/papers/LassoPredict.pdf parmi plein d'autres). Cette technique est aussi extrêment utlisée pour la détection des groupes de gènes important pour une maladie spécifique étudiée mais dans ce cas, le modèle introduit n'est plus une régression linéaire mais une régression logistique, ce qui donne un problème d'optimisation un peu plus compliqué ! Un article intéressant à ce sujet est : http ://www-stat.stanford.edu/ hastie/Papers/JRSSB.69.4%20(2007)%20659-677%20Park.pdf, parmi une tonne d'autres bien sûr.

Exercice 4 (Choix de portefeuille de Markowitz, portefeuille tangent et short selling)**.** Un marché est constitué de trois actions à risque : Axa, BNP et Carrefour et deux autres encore dont certainement les plus redoutables et d'un plan épargne action PEA sans risque de taux de rendement 5%. On note la matrice de covariance des actifs risqués C et le vecteur de leurs return espérés μ. Si la premiere variable est le plan d'épargne, la deuxième Axa, la troisième BNP, la quatrième Carrefour etc, alors la matrice de covariance est donnée par :

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2.3 & .93 & .62 & .74 & -.23 \\ 0 & .93 & 1.40 & .22 & .56 & .26 \\ 0 & .62 & .22 & 1.8 & .78 & .27 \\ 0 & .74 & .56 & .78 & 3.4 & -.56 \\ 0 & -.23 & .26 & .27 & -.56 & 2.6 \end{bmatrix}$$

et le vecteur des espérances des accroissements est donné par :

$$\mu = \begin{bmatrix} .03 & .151 & .125 & .147 & .902 & .1768 \end{bmatrix} \tag{8}$$

Au jour t, on a le prix suivant des actifs :

$$s(t) = \begin{bmatrix} 327 & .412 & .17 & .78 & .34 & .56 \end{bmatrix} \tag{9}$$

Notez que comme le produit plan d'épargne est sans risque, sa variance est nulle et sa covariance avec les autres produits également. On se donne un budget de $B = 1$ qui peut correspondre a n'importe quel budget une fois remis à l'échelle. On dit qu'il n'y a pas de short selling si on ne peut que acheter des produits et non pas en vendre alors qu'on ne les possède pas. Aussi étonnant que cela puisse paraître, cette possibilité est autorisée afin de permettre une opportunité plus riche aux speculateurs de toutes sortes, et surtout les gros.

1. Que signifie le fait que deux produits aient une covariance négative ? Donner un exemple de produits qui ont certainement une covariance positive.
2. Donner le portefeuille de Markowitz sans short selling possible ayant un rendement espéré égale à la moyenne des rendements de chaque actif plus 10 %. La condition de short selling est en fait equivalente au fait qu'on ne peut prendre que des quantités positives de chaque produit, une contrainte que l'on a posée explicitement en cours dans l'étude du problème de Markowitz.
3. Pour chaque valeur r de rendement, c'est-à-dire de profit espéré, on obtient un portefeuille différent. Tracez la courbe efficace pour cet ensemble d'actifs, c'est à dire le risque du portefeuille en fonction de r.
4. Reprendre la question 1. dans le cas où le short selling est autorisé.

Exercice 5 (SVM linéaires pour la classifictation)**.** On a vu en cours des Support Vector Machines (SVM) dans le cas linéaire. L'idée consiste à trouver un hyperplan affine qui sépare l'espace en deux demi-espaces de telle manière que :

(i) Toutes les données appartenant à la même classe sont du **même côté** de l'hyperplan.

(ii) L'hyperplan est positionné de **manière optimale**, c'est-à-dire que la marge (la distance au plan du point le plus proche du plan) est la plus grande possible.

La solution est donnée par l'hyperplan d'équation $a^t x + b = 0$ où a et b sont solution du problème :

$$\min_{a \in \mathbb{R}^n,\, b \in \mathbb{R}} \|a\|^2 \text{ sous les contraintes } y_i(a^t x_i + b) \geq 1, \;\; i = 1, \ldots, n. \tag{10}$$

1. Ce problème est-t-il linéaire ? quadratique ?
2. On va générer des données artificiellement. On tire pour cela 10 vecteurs gaussien bidimensionnels d'esperance le vecteur $(0, 0)^t$ et de variance l'identité et 10 vecteurs gaussien bidimensionnels également d'espérance $(10, 5)^t$ et de variance 2 fois l'identité. La fonction pour tirer ces points au hasard est grand avec l'option mn comme "multivariate normal". Représentez les points tirés dans le plan à l'aide de la fonction plot2d. Les deux groupes de points sont-ils bien séparables ? si non, recommencez l'opération jusqu'à ce que ça le soit ! Cela ne sera pas trop long. Enfin, j'espère ...
3. Résoudre le problème de séparation pour ces données artificielles.
4. Tracer la droite de séparation optimale que vous avez trouvée.

Exercice 6 (SVM avec débordements)**.** Le gros plobème avec la recherche d'un hyperplan séparateur est que souvent, les données ne sont tout simplement pas séparables ! il n'y a alors pas de solution au problème (10). Une façon raisonnable de généraliser le problème est la suivante : au lieu d'imposer une marge supérieure ou égale à 1, on peut autoriser qu'elle soit supérieure à $1 - \xi_i$ pour chaque observation $i = 1, \ldots, n$ avec ξ_i le plus petit possible tout en restant supérieur ou égal à zéro. Si les données sont séparables, on pourra donc prendre tous les $\xi_i = 0$, $i = 1, \ldots, n$. Cela peut se mettre sous la forme du problème :

$$\min_{a \in \mathbb{R}^n,\, b \in \mathbb{R}} \|a\|_2^2 + \lambda \|\xi\|_2^2 \text{ sous les contraintes } y_i(a^t x_i + b) \geq 1 - \xi_i, \;\; i = 1, \ldots, n \text{ et } \xi \geq 0. \quad (11)$$

1. Expliquer pourquoi effectivement ce probème est équivalent au problème (10) dans le cas de données séparables.
2. Implémenter cette méthode pour les données de cancer du sein où les colonnes correspondent aux variables 1. Sample code number (id number), 2. Clump Thickness (1 - 10), 3. Uniformity of Cell Size (1 - 10) 4. Uniformity of Cell Shape (1 - 10) 5. Marginal Adhesion (1 - 10), 6. Single Epithelial Cell Size (1 - 10), 7. Bare Nuclei (1 - 10), 8. Bland Chromatin (1 - 10), 9. Normal Nucleoli (1 - 10), 10. Mitoses (1 - 10), 11. Class (2 for benign, 4 for malignant). Les deux classes à identifier sont les tumeurs benignes et les malignes (dernière variable).
3. Enlever quelques données (une dizaine) de la base puis construire l'hyperplan sur les données restantes. Combien sont mal classée ?
4. Une méthode systématique détude consiste en le fait d'enlever une donnée, de calculer l'hyperplan et de voir si cette donnée est oui ou non bien classée et de recommencer en tirant à chaque fois une donnée différente. Cela s'appelle la validation croisée. Implémenter cette méthode. Quel pourcentage de mauvais classement obtenez vous ?

Exercice 7 (Pierre, papier, ciseaux)**.** On a vu en cours que la stratégie optimale pour ce jeu correspondait à jouer "pierre", "papier" ou "ciseaux" avec la même probabilité : 1/3.

1. Tirez une stratégie au hasard pour le joueur 1 (votre adversaire) à l'aide de la fonction rand. On suppose qu'il se bornera à cette stratégie pour toutes les parties à venir.
2. Comment pourriez-vous essayer de deviner la stratégie de votre adversaire après une dizaine de parties ?
3. Jouez 100 parties en adaptant votre stratégie au fur et à mesure. Quel est votre gain ?