Travaux pratiques n°1
Initiation à Python
Méthode des moindres carrés

Objectif : blabla

1 Introduction à Python

```
//comment
function c=fct1(n)
if (n<0) | (int(n)<>n)  then
     c=0;
else
     c=1;
     for i=1:n
          c=c*i;
     end
end
endfunction

a=5+6
sqrt(12)
ans -1
(1+%i)^3
a=[1,2,3], b=[1 2 3]
cos(a)
```

transposée
inverse d'une matrice
Affichage graphique d'une courbe
Chargement & manipulation images

2 Ajout d'un bruit blanc gaussien additif

1D et 2D.

3 Méthode des moindres carrés

La méthode des moindres carrés est utilisée lorsque l'on souhaite ajuster une fonction sur un ensemble de données. Plus précisément, on considère une famille de fonctions $\mathcal{F}$ dont les éléments f_a dépendent d'un paramètre $a \in \mathbb{R}^d$. Un exemple très simple est celui des fonctions linéaires, paramétrées par un paramètre réel $a \in \mathbb{R}$:

$$\mathcal{F} = \left\{ f_a : \mathbb{R} \to \mathbb{R}, x \mapsto a\,x \mid a \in \mathbb{R} \right\}$$

On dispose d'un ensemble de données, dont on suppose qu'il est généré par une fonction de la famille $\mathcal{F}$, c'est-à-dire qu'on dispose d'un ensemble de points $\{(x_i, y_i)\}_{1\leq i\leq n}$ tel que

$$\forall\, i \in [\![\, 1\,;\, n \,]\!], \qquad y_i \approx f_{a_0}(x_i)$$

où $x_i \in \mathbb{R}^p$ et $y_i \in \mathbb{R}^q$ pour tout $i \in [\![\, 1\,;\, n \,]\!]$. L'absence de l'égalité exacte peut être due à l'existence d'un bruit d'acquisition (arrondi numérique, erreur ou imprécision) ou simplement à l'approximation de la modélisation. Dans de nombreuses applications, on cherche à trouver la *meilleure* fonction de $\mathcal{F}$ qui modélise la génération des données. Autrement dit, étant donné l'ensemble des données $u = \{(x_i, y_i)\}_{1\leq i\leq n}$, quelle est la fonction $f_a \in \mathcal{F}$ pour laquelle le nuage de points $v_{f_a} = \{(x_i, f_a(x_i))\}_{1\leq i\leq n}$ se rapproche le plus des données initiales u ? Ces deux ensembles de points appartenant un espace euclidien, on peut [1] mesurer la distance entre ces deux ensembles à l'aide d'une distance de Frobenius :

$$\|v_{f_a} - u\|_F^2 = \sum_{i=1}^{n} \|x_i - x_i\|_2^2 + \|f_a(x_i) - y_i\|_2^2 = \sum_{i=1}^{n} \|f_a(x_i) - y_i\|_2^2$$

Le problème d'ajustement peut donc s'écrire comme le problème d'optimisation suivant :

$$\min_{f_a \in \mathcal{F}} \sum_{i=1}^{n} \|f_a(x_i) - y_i\|_2^2$$

ce qui revient à minimiser l'erreur au carré moyenne entre les données y_i et le modèle théorique $f_a(x_i)$; d'où le nom de méthode des *moindres carrés*. Puisque la famille $\mathcal{F}$ est paramétrée par le vecteur $a \in \mathbb{R}^d$, la forme finale du problème d'ajustement devient

$$\min_{a \in \mathbb{R}^d} \sum_{i=1}^{n} \|f_a(x_i) - y_i\|_2^2$$

On se place à présent dans le cas où $\mathcal{F}$ est l'ensemble des formes linéaires :

$$\mathcal{F} = \Big\{ f_a : \mathbb{R}^p \to \mathbb{R}, x \mapsto \langle a, x \rangle \mid a \in \mathbb{R}^p \Big\}$$

On dispose d'un ensemble de données $u = (X, y) \in \mathcal{M}_{n,p}(\mathbb{R}) \times \mathbb{R}^n$ avec $X = (x_i)_{1\leq i\leq n}$ et $y = (y_i)_{1\leq i\leq n}$. On suppose que $n > p$. On cherche à résoudre le problème

$$\min_{a \in \mathbb{R}^p} J(a) = \sum_{i=1}^{n} \big(\langle a, x_i \rangle - y_i\big)^2 = \|X\, a - y\|_2^2 \qquad (\mathcal{P}_{\mathrm{mm}})$$

Exercice 1 – Théorie

(a) Justifier que la fonction J est différentiable et montrer que

$$\forall\, a \in \mathbb{R}^p, \qquad \nabla J(a) = 2\, X^\top (X\, a - y)$$

(b) Montrer que J est une fonction convexe.

(c) En déduire que l'ensemble des solutions de $(\mathcal{P}_{\mathrm{mm}})$ est l'ensemble des $a \in \mathbb{R}^p$ tels que

$$X^\top X\, a = X^\top\, y$$

(d) On suppose que X est de rang n. Montrer que J est infinie à l'infini. En déduire que J admet au moins une solution.

(e) Justifier que $X^\top X$ est inversible. En déduire que l'unique solution de $(\mathcal{P}_{\mathrm{mm}})$ est le vecteur

$$a = (X^\top X)^{-1}(X^\top\, y)$$

1. Plusieurs choix sont possibles ; celui fait dans la méthode des moindres carrés correspond à l'hypothèse d'un bruit blanc gaussien additif.

> Pour toute matrice $M \in \mathcal{M}_{p,q}(\mathbb{R})$ de rang q, la matrice
> $$M^\dagger = (M^\top M)^{-1} M^\top$$
> est appelée *pseudo-inverse* de M.

On se place dans le cas de la régression linéaire en dimension 1. Prenons pour cela l'exemple loi d'OHM, qui stipule qu'il existe une relation linéaire entre l'intensité I du courant électrique qui traverse un conducteur et la tension U à ses bornes :

$$U = R\,I$$

Le rapport R (constant) entre la tension et l'intensité est appelé *résistance*. Afin d'estimer la valeur de la résistance d'un conducteur donné, on peut faire passer un courant d'une intensité connue I_0 à travers le conducteur, puis mesurer la tension U_0 associée ; enfin, faire le rapport et en déduire une valeur de résistance. Si l'on réitère l'expérience avec une autre valeur d'intensité I_1 connue, on *devrait* trouver la même valeur de résistance.

En pratique, ce n'est pas le cas (et ce, même si $I_0 = I_1$!). De nombreux facteurs expliquent une telle différence : le voltmètre utilisé n'est pas suffisamment précis, l'affichage tronque la valeur de la tension, la loi d'OHM n'est qu'une loi théorique qui *approche* le comportement réel du conducteur... Ainsi, si on fait une série de n mesures de tension U_i associées à n valeurs d'intensité connues I_i, on a seulement

$$\forall\, i \in [\![\, 1\,;\, n \,]\!]\,, \qquad U_i \approx R\,I_i$$

Estimer la valeur de la résistance revient donc à réaliser un ajustement linéaire sur les données $\{(I_i, U_i)\}_{1 \leq i \leq n}$, pour lequel on peut appliquer la méthode des moindres carrés. Le problème d'optimisation associé est le suivant

$$\min_{R \in \mathbb{R}} J(R) = \sum_{i=1}^{n} \big(R\,I_i - U_i\big)^2$$

Exercice 2 – Régression linéaire

- **(a)** Choisir une valeur de $R_{\text{théorique}} \in \mathbb{R}$. Générer $n = 10$ valeurs d'intensité I_i, puis générer n valeurs de tension U_i. Pour simuler l'imprécision des U_i, on ajoutera à chaque valeur théorique $R_{\text{théorique}}\,I_i$ un bruit blanc gaussien de variance $\sigma =$???.
- **(b)** Comment s'interprète la condition de rang maximal sur X dans le cadre de ce problème ?
- **(c)** Montrer que la valeur de résistance R optimale selon la méthode des moindres carrés est donnée par
$$R_{\text{optimal}} = \left(\sum_{i=1}^{n} I_i^2\right)^{-1} \left(\sum_{i=1}^{n} I_i\,U_i\right)$$
- **(d)** Afficher sur le même graphique le nuage de points $\{(I_i, U_i)_{1 \leq i \leq n}\}$ et les droites $U = R_{\text{théorique}}\,I$ et $U = R_{\text{optimal}}\,I$.
- **(e)** Réaliser la même expérience en faisant varier le nombre de mesures n et le niveau de bruit σ. Qu'en concluez-vous ?

Si, au lieu d'estimer la valeur de résistance R, on choisit d'estimer la conductance $G = 1/R$, cela revient à considérer le modèle suivant

$$I = G\,U = \frac{U}{R}$$

On peut donc estimer G (et, de manière équivalente, R) en *imposant* des valeurs U_i (connues) de tension aux bornes du conducteur et en *mesurant* l'intensité I_i qui traverse le conducteur ; ces valeurs vérifient

$$\forall\, i \in [\![\, 1\,;\, n \,]\!]\,, \qquad I_i \approx G\, U_i = \frac{U_i}{R}$$

Ce sont cette fois les mesures I_i qui peuvent être entachées d'erreur ou d'imprécision. La méthode des moindres carrés appliquée à ce modèle revient à résoudre

$$\min_{G \in \mathbb{R}} \sum_{i=1}^{n} (G\, U_i - I_i)^2$$

Autrement dit, on inverse les rôles de U et de I.

Exercice 3 – Estimation de la conductance

- **(a)** À partir des mêmes données que l'exercice précédent, mettre en œuvre la méthode des moindres carrées pour calculer la conductance optimale G_{optimal}.
- **(b)** Afficher sur le même graphique le nuage de points $\{(I_i, U_i)_{1 \leq i \leq n}\}$ et les droites $U = R_{\text{théorique}}\, I$, $U = R_{\text{optimal}}\, I$ et $U = I / G_{\text{optimal}}$.
- **(c)** Qu'observez-vous ?

Lorsque l'on estime la résistance, la droite que l'on obtient minimise en moyenne la distance “verticale” aux points du nuage ; lorsque l'on estime la conductance, la droite minimise en moyenne la distance “horizontale” aux points du nuage. Il est naturel de chercher à trouver la droite qui minimise la distance euclidienne aux points du nuage. Celle-ci est définie pour la droite $\mathcal{D}$ d'équation $R\, I - U = 0$ par

$$\operatorname{dist}(I_i, U_i; \mathcal{D}) = \frac{|R\, I_i - U_i|}{R + 1} = \frac{|I_i - G\, U_i|}{G + 1}$$

Les problèmes d'optimisation s'écrivent donc

$$\min_{R \in \mathbb{R}} \sum_{i=1}^{n} \frac{(R\, I_i - U_i)^2}{(R+1)^2} \qquad \text{et} \qquad \min_{G \in \mathbb{R}} \sum_{i=1}^{n} \frac{(I_i - G\, U_i)^2}{(G+1)^2}$$

Du point de vue de la modélisation, cela revient à considérer que les valeurs d'intensité **et** de tension ne sont pas fiables.

Exercice 4 – Minimisation de la distance des points à la droite

- **(a)** Justifier que, si l'on part des mêmes données, estimer la résistance ou la conductance conduira au même résultat. Autrement dit, $R_{\text{optimal}} = 1/G_{\text{optimal}}$.
- **(b)** Montrer que

$$R_{\text{optimal}} = \left(\sum_{i=1}^{n} I_i\, (I_i + U_i) \right)^{-1} \left(\sum_{i=1}^{n} U_i\, (I_i + U_i) \right)$$

On termine ce premier TP avec un exemple de régression polynomiale. On suppose cette fois que l'on dispose de n couples de données $(x_i, y_i) \in \mathbb{R}^2$ satisfaisant

$$\forall\, i \in [\![\, 1\,;\, n \,]\!]\,, \qquad y_i \approx P(x_i)$$

avec P un polynôme réel de degré au plus p. Tout tel polynôme est défini de manière unique par un vecteur $a = (a_i)_{0\leq k\leq p}$ en écrivant

$$P(X) = \sum_{k=0}^{p} a_k\, X^k$$

L'ajustement polynomial des données par méthode des moindres carrés consiste donc à résoudre le problème d'optimisation

$$\min_{a\in\mathbb{R}^{p+1}} \sum_{i=1}^{n} \left(\sum_{k=0}^{p} a_k\, x_i^k - y_i \right)^2$$

Exercice 5 – Théorie

(a) Montrer que le problème ci-dessus peut se réécrire

$$\min_{a\in\mathbb{R}^{p+1}} \|X\,a - y\|_2^2$$

avec $X = (x_i^k)_{\substack{1\leq i\leq n\\ 0\leq k\leq p}}$ et $y = (y_i)_{1\leq i\leq n}$.

(b) En déduire que les solutions du problème sont les vecteurs $a \in \mathbb{R}^{p+1}$ solutions du système linéaire

$$X^\top X\, a = y$$

Sous quelles conditions ce système admet-il une solution ?

> La matrice X définie dans l'exercice précédent est connue sous le nom de matrice de Vandermonde.

Exercice 6 – Régression polynomiale

(a) Choisir un vecteur $a \in \mathbb{R}^6$. Générer $n = 10$ valeurs x_i, puis générer n valeurs $y_i \approx P(x_i)$ avec P le polynôme associé. Pour simuler l'imprécision des y_i, on ajoutera à chaque valeur théorique $P(x_i)$ un bruit blanc gaussien de variance $\sigma =???$.

(b) Réaliser l'ajustement polynomial pour des degrés entre $p = 0$ et $p = 6$. Afficher sur le même graphique le nuage de points $\{(x_i, y_i)_{1\leq i\leq n}\}$, le polynôme P théorique et les 7 ajustements polynomiaux.