République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

Faculté de Technologie

Département d'électronique

Module: Réseaux de neurones artificiels

S2-M1 Automatique et systèmes

Réseaux de neurones atificiels

TP N°1

Régression linéaire simple et multiple

Préparé par:

Nabil Benoudjit, Mohamed Bahaz & Abdelghani Tafsast

But du TP

Construction et évaluation d'un modèle linéaire par le biais de la régression linéaire simple et de la régression linéaire multiple.

I- Partie théorique

1- Régression linéaire simple

La régression s'adresse à un type de problème où les 2 variables quantitatives continues **x** et **y** ont un rôle asymétrique : la variable **y** dépend de la variable **x**. La liaison entre la variable **y** dépendante (**dite expliquée**) et la variable **x** indépendante (**dite explicative**) peut être modélisée par une fonction de type **y = a*x + b**, représentée graphiquement par une droite. Les coefficients du modèle linéaire sont **a** et **b** calculés comme suit :

$$a = \frac{\sum_{i=1}^{n} \frac{x_i y_i}{n} - \bar{x}\bar{y}}{\sum_{i=1}^{n} \frac{x_i^2}{n} - \bar{x}^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad (1)$$

$$b = \bar{y} - a * \bar{x} \qquad (2)$$

Sous la forme matricielle on a **Y = X*A**, où le vecteur **A** contient les coefficients **a** et **b** du modèle linéaire.

Le vecteur **A** est calculé comme suit : $A = (X^T X)^{-1} X^T Y$ (3)

2- Régression linéaire multiple

La régression linéaire multiple est une méthode d'analyse de données quantitatives. Elle a pour but de mettre en évidence la liaison pouvant exister entre une variable dite **expliquée**, que l'on notera **Y** et plusieurs autres variables dites **explicatives** que l'on notera $X_1, X_2, \ldots, X_k$.

Les k variables X_i, i = 1, ... , k sont contrôlées c'est-à-dire qu'elles sont connues **sans erreur**. Nous nous intéressons aux modèles dits **linéaires**, c'est-à-dire aux modèles du type :

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \ldots + a_k X_k$$

Où, le vecteur ***A*** = [**b, a_1, ... , a_k**] représente les **coefficients du modèle.**

II- Partie simulation

Activité 1:

Nous disposons d'un ensemble de paires de données (voir tableau 1) et nous voulons déterminer la relation entre la variable **expliquée** y et la **variable explicative** x. En d'autres termes trouver la fonction f qui nous permet d'écrire :

$$y = f(x).$$

i	1	2	3	4	5	6
$\mathbf{x_i}$	0	10	20	30	40	50
$\mathbf{y_i}$	4	22	44	60	75	104

Tableau (1)

1- Afficher le nuage de points ($\mathbf{x_i}$, $\mathbf{y_i}$).
2- Calculer le centre de gravité du nuage de points $\bar{x}$ et $\bar{y}$ et afficher le sur la même figure.
3- Quelle remarque pouvez-vous déduire ?
4- Estimer les paramètres **a** et **b** du modèle en utilisant les équations (1) et (2).
5- Calculer le coefficient de détermination (R^2).
6- Calculer l'erreur quadratique moyenne (MSE).
7- Estimer les paramètres **a** et **b** du modèle, cette fois-ci en utilisant l'équation (3).
8- Donner vos conclusions sur cet exemple.

Pour le même exemple, nous avons pour chaque **variable explicative 05 variables expliquées**. Nous proposons de déterminer la relation entre la variable expliquée $\bar{y}$ et la variable explicative x. En d'autre terme trouver la fonction **f** qui nous permet d'écrire :

$$\bar{y} = f(x).$$

i	1	2	3	4	5	6
$\mathbf{x_i}$	0	10	20	30	40	50
$\mathbf{y_{i1}}$	4	22	44	60	75	104
$\mathbf{y_{i2}}$	3	20	46	63	81	109
$\mathbf{y_{i3}}$	4	21	45	60	79	107
$\mathbf{y_{i4}}$	5	22	44	63	78	101
$\mathbf{y_{i5}}$	4	21	44	63	77	105

Tableau (2)

Refaire la même chose (1 à 8) pour les données du tableau 2.

Activité 2:

Nous cherchons à établir la relation du poids du cerveau (Y) en fonction du poids du corps (X) dans une série d'espèces vivantes (homme + animaux).

N°	Espèce	Corps (Kg)	Cerveau (g)	N°	Espèce	Corps (Kg)	Cerveau (g)
1	Castor	**1,35**	**8,1**	15	Eléphant Afrique	**6654**	**5712**
2	Vache	**465**	**423**	16	Triceratops	**9400**	**70**
3	Loup gris	**36,3**	**119,5**	17	Singe Rhésus	**6,8**	**179**
4	Chèvre	**27,66**	**115**	18	Kangourou	**35**	**56**
5	Cochon d'inde	**1,04**	**5,5**	19	Hamster	**0,12**	**1**
6	Diplodocus	**11700**	**50**	20	Souris	**0,023**	**0,4**
7	Eléphant d'Asie	**2547**	**4603**	21	Lapin	**2,5**	**12,1**
8	Ane	**187,1**	**419**	22	Mouton	**55,5**	**175**
9	Cheval	**521**	**655**	23	Jaguar	**100**	**157**
10	Macaque	**10**	**115**	24	Chimpanzé	**52,16**	**440**
11	Chat	**3,3**	**25,6**	25	Brachiosaure	**87000**	**154,5**
12	Girafe	**529**	**680**	26	Rat	**0,28**	**1,9**
13	Gorille	**207**	**406**	27	Taupe	**0,122**	**3**
14	Homme	**62**	**1320**	28	Cochon	**192**	**180**

Tableau (3)

1- Afficher le nuage de points ($\mathbf{x}_i$, $\mathbf{y}_i$). Quelle remarque pouvez-vous déduire ?
2- Utiliser le logarithme sur les $\mathbf{x}_i$ et $\mathbf{y}_i$. Afficher le nouveau nuage de points (log($\mathbf{x}_i$), log($\mathbf{y}_i$)).
3- Quelle remarque pouvez-vous déduire après cette transformation ?
4- Calculer le centre de gravité du nouveau nuage de points $\bar{x}$ *et* $\bar{y}$ et afficher le sur la même figure.
5- Estimer les paramètres **a** et **b** du modèle en utilisant l'équation (3).
6- Calculer le coefficient de détermination (R^2).
7- Calculer l'erreur quadratique moyenne (MSE).
8- Enlever les échantillons (mesures) **aberrants 6, 16 et 25** de la base de données et afficher le nouveau nuage.
9- Calculer le centre de gravité du nouveau nuage de points $\bar{x}$ *et* $\bar{y}$ et afficher le sur la même figure.
10- Estimer les nouveaux paramètres **a** et **b** du modèle en utilisant l'équation (3).
11- Calculer le nouveau coefficient de détermination (R^2).
12- Calculer la nouvelle erreur quadratique moyenne (MSE).
13- Donner vos conclusions sur cet exemple.

<u>Activité 3</u>:

Nous voulons réaliser une régression linéaire multiple pour expliquer la hauteur de neige ($\mathbf{y}_i$) en fonction de l'altitude ($\mathbf{x}_{i1}$), de la rugosité ($\mathbf{x}_{i2}$), de la pente ($\mathbf{x}_{i3}$), de l'orientation ($\mathbf{x}_{i4}$), de la latitude ($\mathbf{x}_{i5}$) et de la longitude ($\mathbf{x}_{i6}$).

N°	Hauteur neige	Altitude	Rugosité	Pente	Orientation	Latitude	Longitude
1	**95**	**2768**	**252**	**22**	**324**	**8760219**	**438465.0625**
2	**150**	**4108**	**333**	**29**	**308**	**8760195**	**438474.0625**
3	**4**	**4045**	**62**	**5**	**249**	**8760168**	**438480.0625**
4	**0**	**4572**	**85**	**8**	**14**	**8760135**	**438489.0625**
5	**0**	**4614**	**115**	**10**	**63**	**8760105**	**438495.0625**
6	**80**	**4321**	**176**	**16**	**130**	**8760072**	**438498.0625**
7	**95**	**3886**	**72**	**6**	**199**	**8760039**	**438504.0625**
8	**20**	**4206**	**57**	**5**	**32**	**8760012**	**438507.0625**

9	**90**	**4192**	**266**	**23**	**197**	**8759985**	**438513.0625**
10	**10**	**4051**	**69**	**6**	**113**	**8759955**	**438519.0625**
11	**10**	**3746**	**62**	**5**	**149**	**8759922**	**438519.0625**
12	**50**	**3789**	**42**	**3**	**218**	**8759895**	**438525.0625**
13	**45**	**3771**	**44**	**4**	**53**	**8759865**	**438531.0625**
14	**60**	**3796**	**48**	**4**	**101**	**8759838**	**438534.0625**
15	**55**	**3885**	**77**	**7**	**332**	**8759811**	**438537.0625**
16	**3**	**4295**	**113**	**10**	**18**	**8759787**	**438540.0625**
17	**33**	**4467**	**147**	**13**	**50**	**8759760**	**438546.0625**
18	**0**	**4764**	**12**	**1**	**276**	**8759730**	**438552.0625**
19	**35**	**4313**	**38**	**3**	**350**	**8759703**	**438552.0625**
20	**45**	**4387**	**40**	**3**	**46**	**8759673**	**438558.0625**

Tableau (4)

1- Estimer les paramètres du modèle (le vecteur ***A***) en utilisant l'équation (3).
2- Calculer le coefficient de détermination (R^2).
3- Calculer l'erreur quadratique moyenne (MSE).
4- Donner vos conclusions sur cet exemple.

Bon courage