Master 1 Informatique – ENSICAEN 2A

Module « Apprentissage »

TD 1: Régression

A. Lechervy

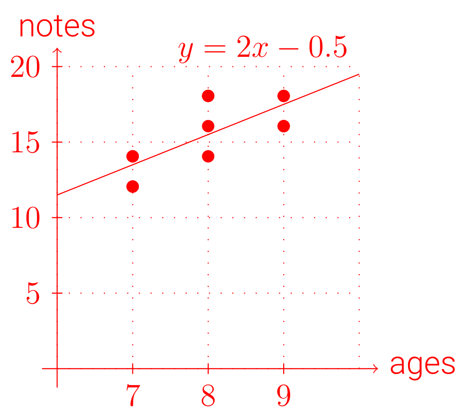
---

1 Régression linéaire

Un Test de lecture a été réalisé sur 12 enfants de 7 à 9 ans dont les notes ont été les suivantes:

i	1	2	3	4	5	6	7	8	9	10	11	12
ages	7	8	9	7	9	8	7	9	8	9	8	7
notes	12	16	16	14	18	18	12	18	14	16	18	14

Question 1.1. *Tracez les notes en fonction de l'age.*



Question 1.2. *Calculez les moyennes et la matrice de covariance des données.*

Moyennes:

$$\begin{aligned}
\overline{x} &= \frac{4 \times 7 + 4 \times 8 + 4 \times 9}{12} && (1)\\
&= \frac{96}{12} && (2)\\
&= 8 && (3)
\end{aligned}$$

$$\begin{aligned}
\overline{y} &= \frac{2 \times 12 + 3 \times 14 + 3 \times 16 + 4 \times 18}{12} && (4)\\
&= \frac{184}{12} && (5)\\
&= 15.5 && (6)
\end{aligned}$$

i	1	2	3	4	5	6	7	8	9	10	11	12
x	7	8	9	7	9	8	7	9	8	9	8	7
y	12	16	16	14	18	18	12	18	14	16	18	14
$x\breve{}\overline{x}$	-1	0	1	-1	1	0	-1	1	0	1	0	-1
$y\breve{}\overline{y}$	-3.5	0.5	0.5	-1.5	2.5	2.5	-3.5	2.5	-1.5	0.5	2.5	-1.5
$(x\breve{}\overline{x})^2$	1	0	1	1	1	0	1	1	0	1	0	1
$(y\breve{}\overline{y})^2$	12.25	0.25	0.25	2.25	6.25	6.25	12.25	6.25	2.25	0.25	6.25	2.25
$(x\breve{}\overline{x})(y\breve{}\overline{y})$	3.5	0	0.5	1.5	2.5	0	3.5	2.5	0	0.5	0	1.5

$$cov = \begin{bmatrix} 2/3 & 4/3 \\ 4/3 & 19/4 \end{bmatrix} = \begin{bmatrix} 0.6666666667 & 1.3333333333 \\ 1.3333333333 & 4.75 \end{bmatrix}.$$

Question 1.3. *Déduisez de l'équation précédente la droite passant le plus proche des données, puis tracez là.*

$$\begin{aligned} a &= \frac{cov(x,y)}{cov(x,x)}, && (7)\\ &= \frac{4/3}{2/3}, && (8)\\ &= 2. && (9) \end{aligned}$$

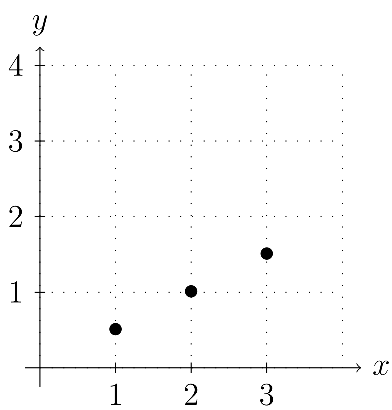
$$\begin{aligned} b &= \overline{y} - a\overline{x}, && (10)\\ &= 15.5 - 2 \cdot 8, && (11)\\ &= -0.5 && (12) \end{aligned}$$

Question 1.4. *A votre avis, quelle devrait être la note d'un enfant de 10 ans ?*

$$y = 2 \cdot 10 - 0.5 = 19.5$$

2 Descente de gradient

Nous voulons faire une régression par descente de gradient sur les données suivantes:

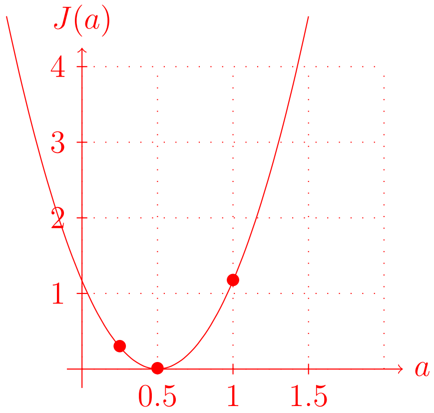


Question 2.1. *Calculez la valeur de la fonction de coût du critère par moindre carré dans le cas d'une régression de type $y = ax$ pour les droites suivantes:*

- $y = x$,
- $y = 0.5x$,
- $y = 0.25x$.

$$
\begin{aligned}
J(a=1) &= \frac{1}{3}\left((0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right) = \frac{7}{6} \simeq 1.17 && (13)\\
J(a=0.5) &= \frac{1}{3}\left((0.5-1\cdot 0.5)^2 + (1-2\cdot 0.5)^2 + (1.5-3\cdot 0.5)^2\right) = 0 && (14)\\
J(a=0.25) &= \frac{1}{3}\left((0.5-1\cdot 0.25)^2 + (1-2\cdot 0.25)^2 + (1.5-3\cdot 0.25)^2\right) = \frac{7}{24} \simeq 0.3 && (15)\\
& && (16)
\end{aligned}
$$

Question 2.2. *Tracez à partir des points précédents la fonction de coût $J(a)$ du problème de régression.*



Question 2.3. *Calculez les 3 premières itérations d'une descente de gradient en partant de la droite $y = 1 \cdot x$ pour un taux d'apprentissage de $\alpha = 0.1$. Tracez la descente sur la figure précédente.*

$$J(a) = \frac{1}{n}\sum_i (y_i - ax_i)^2 = \frac{1}{3}((0.5-a)^2 + (1-2a)^2 + (1.5-3a)^2).$$

$$
\begin{aligned}
\frac{\partial J(a)}{\partial a} &= \frac{1}{n}\sum_i 2ax_i^2 - 2x_iy_i, && (17)\\
&= \frac{2}{3}((1^2a - 1\cdot 0.5) + (2^2a - 2\cdot 1) + (3^2a - 3\cdot 1.5)), && (18)\\
&= \frac{2}{3}(14a - 7), && (19)\\
&= \frac{28}{3}a - \frac{14}{3}, && (20)\\
&\simeq 9.333a - 4.667. && (21)\\
& && (22)
\end{aligned}
$$

$$
\begin{aligned}
a &:= a - \alpha\left(\frac{28}{3}a - \frac{14}{3}\right), && (23)\\
&:= \left(1 - \alpha\frac{28}{3}\right)a + \alpha\frac{14}{3}. && (24)
\end{aligned}
$$

t	0	1	2	3
a	1	0.533	0.502	0.500

Question 2.4. *Calculez la solution analytique du problème de régression. Trouvez vous la même solution que par descente de gradient ?*

$$
\begin{aligned}
\frac{\partial J(a)}{\partial a} &= 0, && (25)\\
\Longrightarrow \frac{28}{3}a - \frac{14}{3} &= 0, && (26)\\
\Longrightarrow a = \frac{14}{28} &= 0.5. && (27)\\
& && (28)
\end{aligned}
$$

ou

$$
\begin{aligned}
a &= \frac{\overline{xy}}{\overline{x^2}}, && (29)\\
&= \frac{1\cdot 0.5 + 2\cdot 1 + 3\cdot 1.5}{1^2 + 2^2 + 3^2}, && (30)\\
&= \frac{1\cdot 0.5 + 2\cdot 1 + 3\cdot 1.5}{1^2 + 2^2 + 3^2}, && (31)\\
&= \frac{7}{14}, && (32)\\
&= 0.5. && (33)
\end{aligned}
$$

Question 2.5. *Nous allons effectué des descentes de gradients avec d'autres valeurs de α:*

- $\alpha = 0.3$,
- $\alpha = 0.01$.

Que constatez vous ? Calculez et tracer la valeur de la fonction de coût au cours des itérations et expliquer son comportement.

$\alpha = 0.3$

t	0	1	2	3
a	1	-0.4	2.12	-2.416
J(a)	1.17	3.78	12.25	39.68

$\alpha = 0.01$

t	0	1	2	3
a	1	0.953	0.911	0.872
J(a)	1.17	0.959	0.788	0.648

3 Régression polynomiale

Lors du lancement d'une balle nous avons mesurer l'altitude de la balle en fonction du temps. On obtient les valeurs suivantes:

t[s]	0	3	4
h[m]	1	13	9

Question 3.1. *On suppose que l'on peux approximer cette fonction par un polynôme de degré 2. Calculez les coefficients de ce polynôme.*

Rappel:

$$X^{-1} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}^{-1}, \tag{34}$$

$$= \frac{1}{\det(X)} \begin{bmatrix} ei - fg & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{bmatrix}. \tag{35}$$

Avec:

$$\det(X) = aei + bfg + cdh - ceg - fha - ibd.$$

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \text{ et } Y = \begin{bmatrix} 1 \\ 13 \\ 9 \end{bmatrix}$$

$$\begin{aligned} A &= (X^\top X)^{-1} X^\top Y, & (36)\\ &= \begin{bmatrix} 1 \\ 10 \\ -2 \end{bmatrix} & (37)\\ &\implies -2x^2 + 10x + 1 \end{aligned}$$

Question 3.2. *Déduisez de la question précédente l'instant $\hat{t}$ où la balle atteint l'altitude maximale.*

$$\frac{\partial - 2x^2 + 10x + 1}{\partial x} = -4x + 10 = 0 \implies x = 2.5$$

4 Démonstration

Question 4.1. *Rappelez la formule du problème de régression d'arête.*

$$\theta^\star_{\text{ridge}} = \arg\min_w \sum_{i=1}^m \left(y_i - \theta_0 - \sum_{j=1}^d \theta_j x_i^{(j)} \right)^2 + \lambda \sum_{j=1}^d \theta_j^2.$$

Question 4.2. *Donnez la solution analytique de la question précédente et faites la démonstration de votre résultat.*

$$\begin{aligned} \frac{\partial J(\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta} (y - X\theta)^\top (y - X\theta), & (38)\\ &= \frac{\partial}{\partial \theta} \left[y^\top y - \theta^\top X^\top y - y^\top X\theta + \theta^\top X^\top X\theta + \lambda\theta^\top\theta \right], & (39)\\ &= 0 - X^\top y - (y^\top X)^\top + 2X^\top X\theta + \lambda\theta, & (40)\\ &= -2(X^\top y - \left[X^\top X + \lambda \text{Id}\right] \theta), & (41)\\ &\implies X^\top y - \left[X^\top X + \lambda \text{Id}\right] \theta = 0, & (42)\\ &\implies \theta = \left[X^\top X + \lambda \text{Id}\right] X^\top y. & (43) \end{aligned}$$