Centre Universitaire de Mila	Examen de rattrapage en « Apprentissage Automatique »
Institut : ST, Département : MI	Durée de l'examen : 1h30 / Mila le : 10-06-2017
Master 1 – STIC	Documents non autorisés

Exercice 1 (8 points).

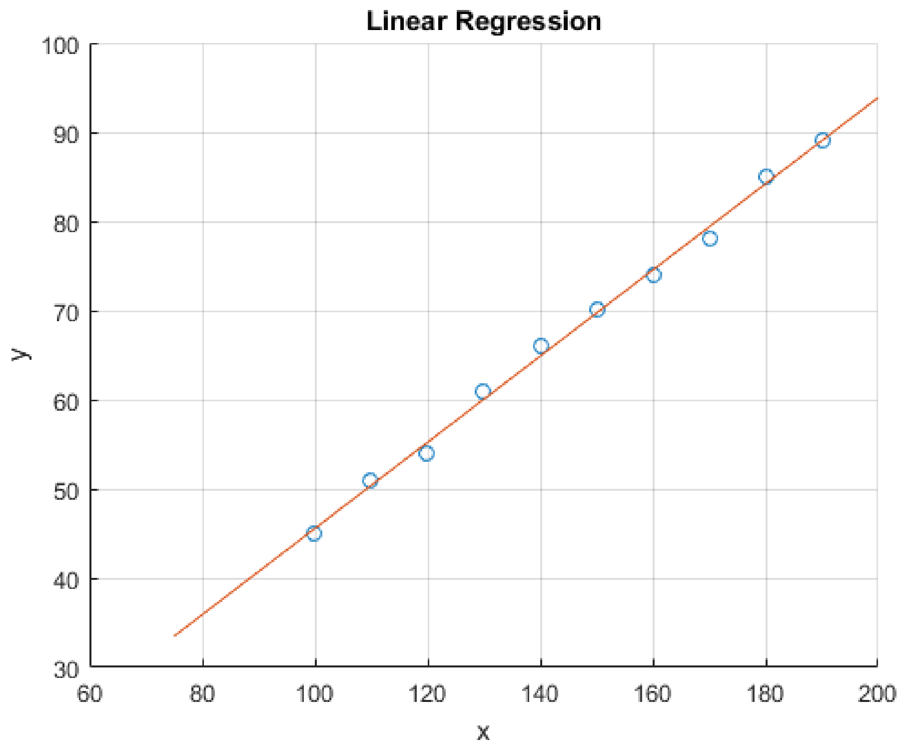
L'analyse de la température de fonctionnement d'un procédé chimique sur le rendement du produit a donné les valeurs suivantes pour la température et le rendement correspondant :

Température °C	Rendement %	Température °C	Rendement %
100	45	150	70
110	51	160	74
120	54	170	78
130	61	180	85
140	66	190	89

1) Donner une représentation graphique de ces données.
2) Trouver la fonction de régression linéaire par la méthode des moindres carrés, qui permet d'associer à la température la valeur de rendement correspondante.
3) Utilisant cette fonction de régression, prédire la valeur de rendement pour la température 80°C.
4) Déterminer (en utilisant la droite de régression) quand la valeur de rendement sera supérieure à 100.

Solution :

1) Représentation graphique de ces données.



2) Modèle de régression linéaire :

On a

$$x = [100; 110; 120; 130; 140; 150; 160; 170; 180; 190];$$

et l'étiquette y

$$y = [45; 51; 54; 61; 66; 70; 74; 78; 85; 89];$$

Application du modèle de régression linéaire :

Dans le modèle de la régression linéaire, l'ensemble des paramètres est calculé par la formule suivante :

$$\widetilde{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{y}$$

Application numérique sur les données de l'exercice :

$$\boldsymbol{X} = \begin{bmatrix} 100 & 1 \\ 110 & 1 \\ 120 & 1 \\ 130 & 1 \\ 140 & 1 \\ 150 & 1 \\ 160 & 1 \\ 170 & 1 \\ 180 & 1 \\ 190 & 1 \end{bmatrix} \quad \boldsymbol{y} = \begin{bmatrix} 45 \\ 51 \\ 54 \\ 61 \\ 66 \\ 70 \\ 74 \\ 78 \\ 85 \\ 89 \end{bmatrix}$$

$$\boldsymbol{X^TX} = \begin{bmatrix} 100 & 110 & 120 & 130 & 140 & 150 & 160 & 170 & 180 & 190 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} 100 & 1 \\ 110 & 1 \\ 120 & 1 \\ 130 & 1 \\ 140 & 1 \\ 150 & 1 \\ 160 & 1 \\ 170 & 1 \\ 180 & 1 \\ 190 & 1 \end{bmatrix} = \begin{bmatrix} 218500 & 1450 \\ 1450 & 10 \end{bmatrix}$$

$$\boldsymbol{X^Ty} = \begin{bmatrix} 100 & 110 & 120 & 130 & 140 & 150 & 160 & 170 & 180 & 190 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} 45 \\ 51 \\ 54 \\ 61 \\ 66 \\ 70 \\ 74 \\ 78 \\ 85 \\ 89 \end{bmatrix} = \begin{bmatrix} 101570 \\ 673 \end{bmatrix}$$

On a $\boldsymbol{det}(\boldsymbol{X^TX}) = \mathbf{82500}$**,** donc $(\boldsymbol{X^TX})^{-1} = \begin{bmatrix} 10/82500 & -1450/82500 \\ -1450/82500 & 218500/82500 \end{bmatrix} =$

$$\begin{bmatrix} 0.0001 & -0.0176 \\ -0.0176 & 2.6485 \end{bmatrix}$$

D'où $\widetilde{\mathbf{w}} = (\boldsymbol{X^TX})^{-1}\boldsymbol{X^Ty} = \begin{bmatrix} 0.4830 \\ -2.7394 \end{bmatrix}$

$$\boldsymbol{y} = -\mathbf{2.7394} + \mathbf{0.4830}\boldsymbol{x}$$

3) Prédire la valeur de rendement pour la température 80°C :

$$\boldsymbol{y_{80°C}} = -\mathbf{2.7394} + \mathbf{0.4830} * \mathbf{80} = \mathbf{35.9006}$$

4) La valeur de rendement sera supérieure à 100 si on a :

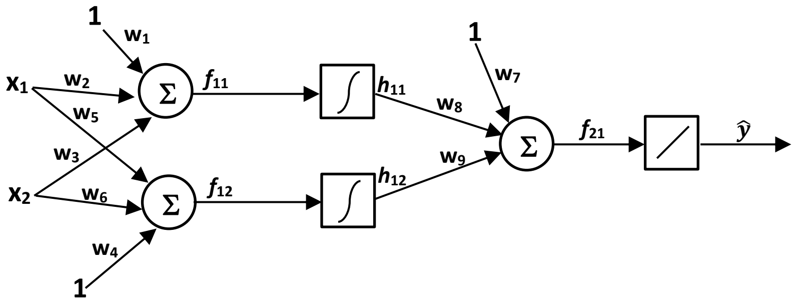
$$\boldsymbol{y} > \mathbf{100} \Rightarrow -\mathbf{2.7394} + \mathbf{0.4830}\boldsymbol{x} > \mathbf{100}$$

$$\Rightarrow \boldsymbol{x} > \frac{\mathbf{100 + 2.7394}}{\mathbf{0.4830}}$$

$$\Rightarrow \boldsymbol{x} \geq \mathbf{212°C}$$

Exercice 2 (8 points).

Soit le réseau de neurones multicouches décrit par le graphe suivant :



1) Donner les formules mathématiques qui déterminent les sorties intermédiaires f_{11}, f_{12}, h_{11}, h_{12}, f_{21} ainsi que la sortie finale $\hat{y}$.
2) Soit la fonction d'erreur définie par :

$$\mathrm{E(w)} = \left(y - \hat{\boldsymbol{y}}_{(x,w)}\right)^2$$

a- Donner la formule de la dérivée partielle $\frac{\partial E(w)}{\partial w_j}$.

b- En appliquant l'algorithme de propagation en arrière (backpropagation), trouver les expressions des mises à jour des paramètres Δw_j pour j = 1, …, 9.

Solution :

1) Propagation en avant (forwardpropagartion algorithm)

$$f_{11} = w_1 + w_2 x_1 + w_3 x_2$$

$$f_{12} = w_4 + w_5 x_1 + w_6 x_2$$

$$h_{11} = sigm(f_{11}) = \frac{1}{1 + e^{-f_{11}}}$$

$$h_{12} = sigm(f_{12}) = \frac{1}{1 + e^{-f_{12}}}$$

$$\boldsymbol{\hat{y} = f_{21} = w_7 + w_8 h_{11} + w_9 h_{12}}$$

2) Propagation en arrière (backpropagartion algorithm)

a- La fonction d'erreur est donnée par

$$\mathrm{E(w)} = \left(y - \hat{\boldsymbol{y}}_{(x,w)}\right)^2$$

Donc, on aura

$$\frac{\partial E(w)}{\partial w_j} = -2\left(y - \hat{\boldsymbol{y}}_{(x,w)}\right)\frac{\partial \hat{\boldsymbol{y}}_{(x,w)}}{\partial w_j}$$

b- D'après la propagation en avant, on a :

$$\hat{y} = f_{21} = w_7 + w_8 h_{11} + w_9 h_{12}$$

Donc, les dérivées $\frac{\partial \hat{\boldsymbol{y}}_{(x,w)}}{\partial w_j}$ sont données par :

$$\frac{\boldsymbol{\partial \hat{y}}}{\boldsymbol{\partial w_7}} = 1$$

$$\frac{\boldsymbol{\partial \hat{y}}}{\boldsymbol{\partial w_8}} = h_{11}$$

$$\frac{\boldsymbol{\partial \hat{y}}}{\boldsymbol{\partial w_9}} = h_{12}$$

$$\frac{\boldsymbol{\partial \hat{y}}}{\boldsymbol{\partial w_1}} = \frac{\partial \hat{y}}{\partial h_{11}} \frac{\partial h_{11}}{\partial f_{11}} \frac{\partial f_{11}}{\partial w_1} = w_8 h_{11}(1 - h_{11})$$

$$\frac{\boldsymbol{\partial \hat{y}}}{\boldsymbol{\partial w_2}} = \frac{\partial \hat{y}}{\partial h_{11}} \frac{\partial h_{11}}{\partial f_{11}} \frac{\partial f_{11}}{\partial w_2} = w_8 h_{11}(1 - h_{11})x_1$$

$$\frac{\boldsymbol{\partial \hat{y}}}{\boldsymbol{\partial w_3}} = \frac{\partial \hat{y}}{\partial h_{11}} \frac{\partial h_{11}}{\partial f_{11}} \frac{\partial f_{11}}{\partial w_2} = w_8 h_{11}(1 - h_{11})x_2$$

$$\frac{\boldsymbol{\partial \hat{y}}}{\boldsymbol{\partial w_4}} = \frac{\partial \hat{y}}{\partial h_{12}} \frac{\partial h_{12}}{\partial f_{12}} \frac{\partial f_{12}}{\partial w_3} = w_9 h_{12}(1 - h_{12})$$

$$\frac{\boldsymbol{\partial \hat{y}}}{\boldsymbol{\partial w_5}} = \frac{\partial \hat{y}}{\partial h_{12}} \frac{\partial h_{12}}{\partial f_{12}} \frac{\partial f_{12}}{\partial w_4} = w_9 h_{12}(1 - h_{12})x_1$$

$$\frac{\boldsymbol{\partial \hat{y}}}{\boldsymbol{\partial w_6}} = \frac{\partial \hat{y}}{\partial h_{12}} \frac{\partial h_{12}}{\partial f_{12}} \frac{\partial f_{12}}{\partial w_4} = w_9 h_{12}(1 - h_{12})x_2$$

En fin, la mise à jour de chaque paramètre est donnée par la formule :

$$\Delta \boldsymbol{w_j} = \boldsymbol{\alpha}(\boldsymbol{y} - \boldsymbol{\hat{y}}_{(x,w)}) \frac{\boldsymbol{\partial \hat{y}}_{(x,w)}}{\boldsymbol{\partial w_j}}$$

Exercice 3 (4 points).

Soit **A** et **B** deux variables booléennes.

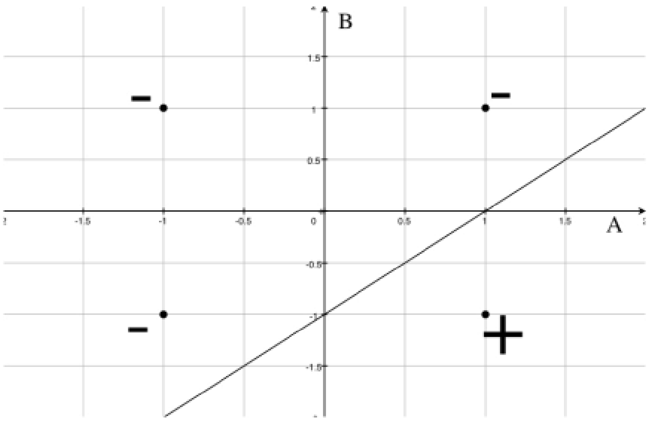
1) Concevoir un réseau de neurones à deux entrées permettant d'implémenter la fonction booléenne $\mathbf{A} \wedge \neg\mathbf{B}$.
2) Concevoir un réseau de neurones à deux couches implémentant la fonction booléenne **A XOR B.**

Solution :

1) Le perceptron demandé a **3** entrées : **A**, **B** et la constante $\mathbf{1}$. Les valeurs de **A** et **B** sont $\mathbf{1}$ (vrai) ou $-\mathbf{1}$ (faux). Le tableau suivant décrit la sortie $\boldsymbol{y}$ du perceptron :

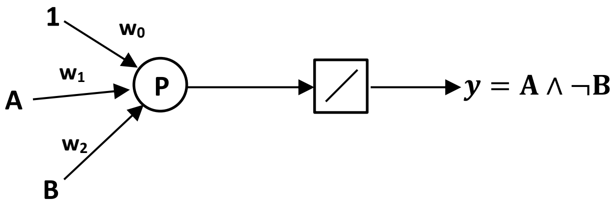
A	**B**	$\mathbf{y = A \wedge \neg B}$
-1	**-1**	**-1**
-1	**1**	**-1**
1	**-1**	**1**
1	**1**	**-1**

Une des surfaces de décision correctes (n'importe quelle droite séparant le point positif des points négatifs serait correcte) est affichée dans la figure suivante :



La droite traversant l'axe $\mathbf{A}$ en $\mathbf{1}$ et l'axe $\mathbf{B}$ en $\mathbf{-1}$. Donc l'équation de la droite est

$$\frac{\boldsymbol{A-0}}{\mathbf{1-0}} = \frac{\boldsymbol{B-(-1)}}{\mathbf{0-(-1)}} \Rightarrow \boldsymbol{A = B + 1} \Rightarrow \boldsymbol{1 - A + B = 0}$$



Les valeurs possibles pour les poids **w_0**, **w_1** et **w_2** sont **1** et **-1**. En utilisant ces valeurs, la sortie du perceptron pour **A = 1, B = -1** est négative. Par conséquent, nous pouvons conclure que **w_0 = -1, w_1 = 1, w_2 = -1.**

2) $\mathbf{A\ XOR\ B}$ ne peut pas être calculé par un perceptron unique, nous devons donc construire un réseau à deux couches de perceptrons. La structure du réseau peut être dérivée par :

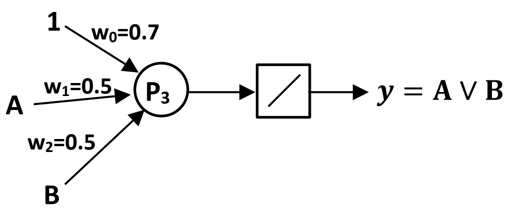
- Exprimer $\mathbf{A\ XOR\ B}$ en fonction des composantes logiques :

$$\mathbf{A\ XOR\ B} = (\mathbf{A} \wedge \neg\mathbf{B}) \vee (\neg\mathbf{A} \wedge \mathbf{B})$$

- Définir les perceptrons $\mathbf{P_1}$ et $\mathbf{P_2}$ pour $(\mathbf{A} \wedge \neg\mathbf{B})$ et $(\neg\mathbf{A} \wedge \mathbf{B})$
- Combiner les sorties de $\mathbf{P_1}$ et $\mathbf{P_2}$ dans un perceptron $\mathbf{P_3}$ qui implémente $f(\mathbf{P_1}) \vee f(\mathbf{P_2})$

Les perceptrons $\mathbf{P_1}$ et $\mathbf{P_2}$ sont définis de manière similaire à $\mathbf{P}$. $\mathbf{P_3}$ défini la fonction $\vee$ (le ou logique) comme suit :

$\mathbf{A}$	$\mathbf{B}$	$\mathbf{y = A \vee B}$
-1	**-1**	**-1**
-1	**1**	**1**
1	**-1**	**1**
1	**1**	**1**



En fin, le réseau demandé est donné dans la figure suivante :