Rappels de base (vecteurs, matrices, distances). Notations. Codage de caractéristiques.

Remarque : les exercices de cette feuille sont à faire par écrit, sans ordinateur.

Exercice 1 *Rappels : vecteurs, distances,...*
On considère une base d'apprentissage $\mathbf{X}$ contenant n exemples décrits par d attributs (ou variables), et soit $\mathbf{x}_1$ et $\mathbf{x}_2$ deux exemples de $\mathbf{X}$. Dans cet exercice, on considère que les d attributs sont tous numériques (ie. à valeurs dans $\mathbb{R}$). Avec les notations que nous utiliserons ce semestre, on note $\mathbf{x}_1 = (x_{1,1}, x_{1,2}, \ldots, x_{1,d})$ et $\mathbf{x}_2 = (x_{2,1}, x_{2,2}, \ldots, x_{2,d})$.

Question 1. Donner l'expression analytique complète de $\mathbf{X}$ sous la forme d'une matrice.

Question 2. Donner l'expression analytique du produit scalaire $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle$ (parfois noté $K(\mathbf{x}_1, \mathbf{x}_2)$) entre $\mathbf{x}_1$ et $\mathbf{x}_2$ selon leurs composantes. Représenter ensuite $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle$ sous la forme d'un produit de matrices.

Question 3. Donner l'expression analytique de la distance euclidienne $d_E(\mathbf{x}_1, \mathbf{x}_2)$ entre $\mathbf{x}_1$ et $\mathbf{x}_2$.

Question 4. Soit $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^+$ une mesure. Rappeler les propriétés que doit vérifier f pour être une distance.

Question 5. Donner l'expression analytique de la norme euclidienne $||\mathbf{x}_1||$ (aussi appelée norme 2 et notée $||\mathbf{x}_1||_2$) de $\mathbf{x}_1$.

Question 6. En science des données, la librairie Python `numpy` est utilisée pour les calculs scientifiques. Elle permet, entre autres, de représenter des vecteurs et des matrices et de leur appliquer des opérations matricielles. Par exemple, les instructions suivantes permettent de créer 2 vecteurs `X1` et `X2`, et une matrice `M` :

```
# importation de la librairie (à faire au préalable)
import numpy as np
X1 = np.array([1,2,0,1,5])
X2 = np.array([1,0,1,3,3])
M = np.array([[1, 2, 0, 1, 5], [4, 2, 11, 3, 8]])
```

Les opérations suivantes sont alors possibles[1] :

```
Produit terme à terme : X1*X2 =  [ 1   0   0   3  15]
Produit terme à terme : M*M =
[[  1    4    0    1   25]
 [ 16    4  121    9   64]]
Transposée: X1.T =  [1 2 0 1 5]
Transposée: M.T =
[[ 1   4]
 [ 2   2]
 [ 0  11]
 [ 1   3]
 [ 5   8]]
Produit matriciel : np.dot(X1,X2) =  19
Produit matriciel : np.dot(M,X1.T) =  [31  51]
Produit matriciel : X1@X2 =  19
Produit matriciel : M@X1.T =  [31  51]
```

Analyser et expliquer les résultats obtenus en réalisant "à la main" les opérations présentées.

Question 7. Donner la valeur du produit de la matrice M avec elle-même. Donner ensuite l'instruction Python permettant de réaliser ce calcul.

1. L'étude de numpy sera faite de façon plus approfondie en TME.

Question 8. Donner le code python permettant de représenter en numpy le vecteur $\mathbf{x} = (4\ \ 2\ \ 0)$ et la matrice

$$\mathbf{M} = \begin{pmatrix} 0 & 4 & 1 \\ 11 & 0 & 3 \\ 1 & 1 & 3 \\ 4 & 1 & 3 \end{pmatrix}$$

puis donner les instructions pour calculer le produit scalaire de $\mathbf{x}$ avec chacune des lignes de $\mathbf{M}$. Faire les calculs à la main pour vérifier ensuite en TME.

Question 9. Donner les instructions pour calculer $||\mathbf{x}||$, puis celle permettant de calculer la distance de $\mathbf{x}$ à chaque vecteur de $\mathbf{M}$ (pris ligne par ligne).

Exercice 2 *Droites et vecteurs (1)*

Dans cet exercice, on se place dans un repère orthonormé en 2 dimensions $\mathbb{R} \times \mathbb{R}$ comme indiqué dans la figure 1.

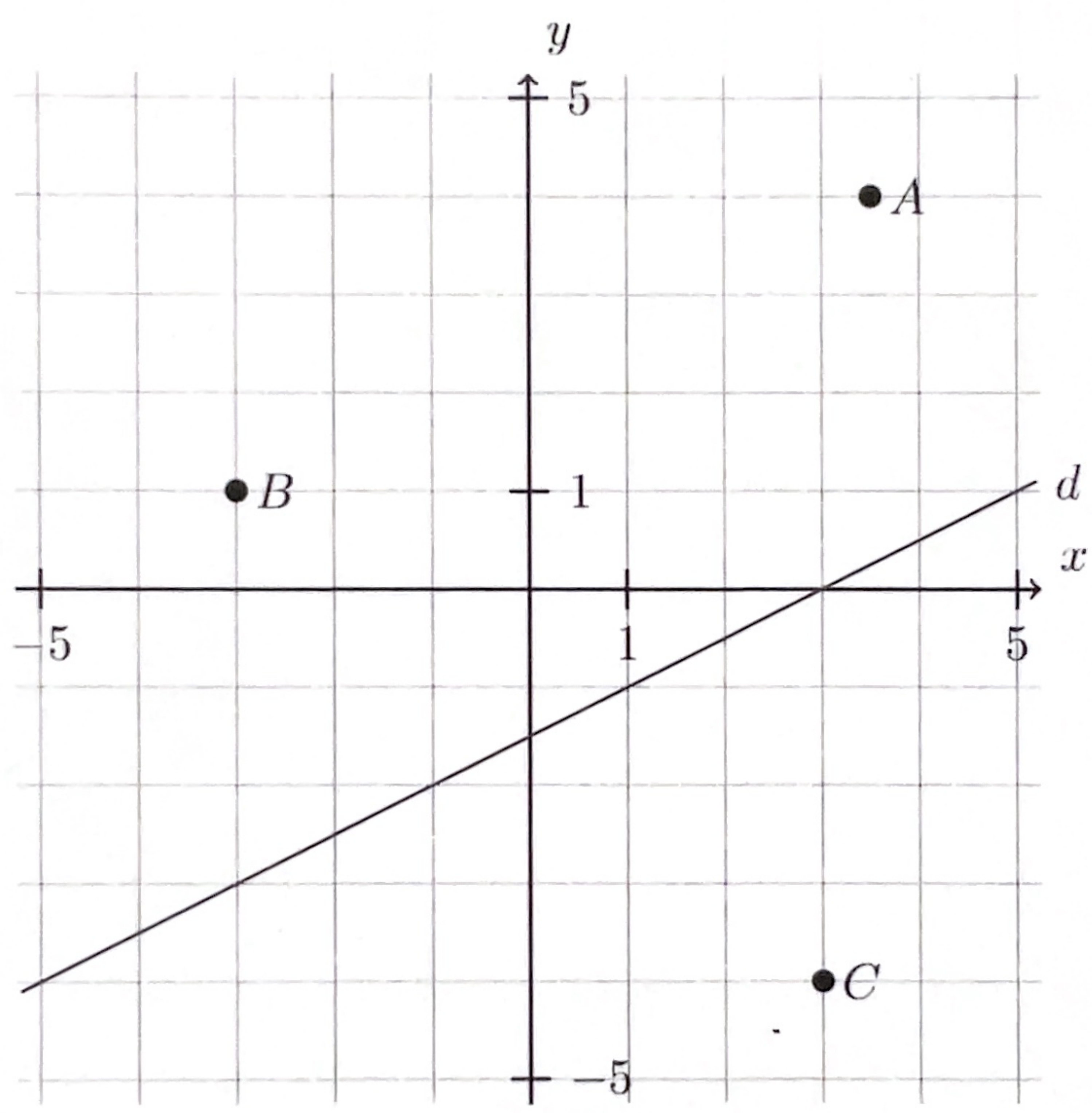

FIGURE 1 – Droite et points

Question 1. Déterminer l'équation de la droite d.

Question 2. Déterminer l'équation de la droite d_A, parallèle à la droite d et passant par le point A.

Question 3. Donner un vecteur directeur de la droite d.

Question 4. Déterminer l'équation de la droite d_B, perpendiculaire à la droite d et passant par le point B.

Exercice 3 *Droites et vecteurs (2)*

Dans cet exercice, on se place dans un repère orthonormé en 2 dimensions $\mathbb{R} \times \mathbb{R}$. On note $\mathbf{x} = (x_1, x_2)$ un point de $\mathbb{R} \times \mathbb{R}$.

Soit d_w la droite d'équation $w_1 x_1 + w_2 x_2 + c = 0$ avec w_1, w_2 et c à valeurs dans $\mathbb{R}$ tels que $(w_1, w_2) \neq (0, 0)$.

Question 1. Donner l'expression de v_d un vecteur directeur de la droite d_w, puis en déduire l'ensemble de tous les vecteurs directeurs de d_w, ie. l'ensemble des vecteurs colinéaires à v_d.

Question 2. Montrer que le vecteur $\mathbf{w} = (w_1, w_2)$ est normal à la droite d_w.

Question 3. On se place dans le cas où $c = 0$. Représenter graphiquement une droite d'équation $w_1 x_1 + w_2 x_2 = 0$ ainsi que le vecteur $\mathbf{w}$ correspondant (on prendra le point $O = (0, 0)$ comme origine).

Question 4. Toujours avec $c = 0$, exprimer l'équation de d_w sous la forme d'un produit scalaire. En déduire que si 2 points $\mathbf{x}_1$ et $\mathbf{x}_2$ sont tels que $\langle \mathbf{w}, \mathbf{x}_1 \rangle$ et $\langle \mathbf{w}, \mathbf{x}_2 \rangle$ ont le même signe, alors ils se trouvent du même côté de la droite d_w (ie. dans le même sous-espace délimité par d_w).

Astuce : utiliser les angles.

Question 5. Généraliser le résultat précédent pour le cas général où $c \neq 0$ (on fait l'hypothèse que w_2 est non nul. Vérifier en prenant les valeurs numériques de l'exercice 2 (droite, et points).