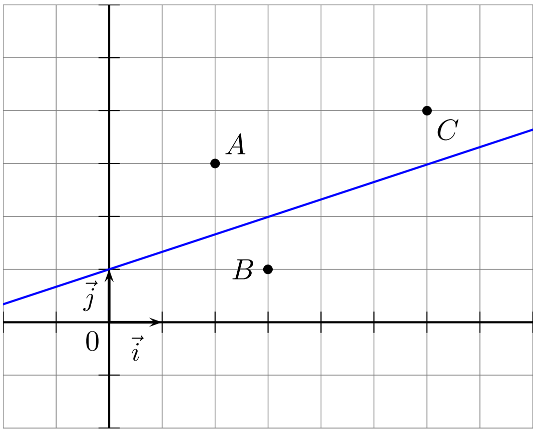
IA – Droite de régression

1 Objectif de l'exercice et quelques définitions

Le but de cet exercice est de définir la droite la plus proche d'un ensemble de points. Si on considère deux points, la droite la plus proche de ces deux points est simplement la droite qui passe par ces deux points. Mais si on considère 3 points ou plus qui ne sont pas alignés, comme dans l'exemple ci-dessous, il n'y a pas, bien sûr, de droite qui passe par tout les points.



On peut alors chercher la droite qui soit la plus proche des trois points, c'est à dire telle que la distance de l'ensemble des points à la droite soit minimal. Pour cela, il faut commencer par définir ce qu'on entend par : **distance d'un ensemble de points à une droite**.

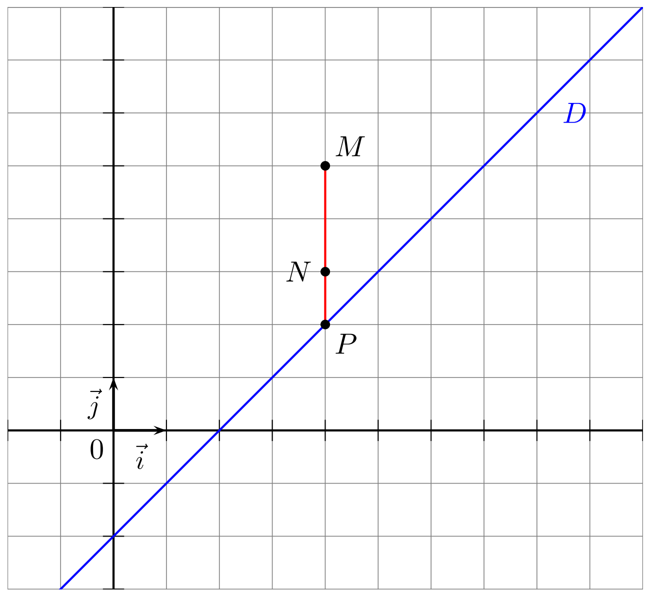
On commence par définir la **distance d'un point à une droite** :

Définition : Soit D la droite d'équation $y = ax + b$ et M un point de coordonnées $(x; y)$, on définit la distance du point M à la droite D, que l'on note $d(M; D)$, par la formule :

$$d(D, M) = [y - (ax + b)]^2 \qquad (1)$$

La distance est donc le carré de la différence entre l'ordonnée du point M et l'ordonnée du point de la droite (d) qui a le même abscisse que M.

Sur l'exemple ci-dessous, la distance du point $M(4; 5)$ à la droite D d'équation $y = x - 2$ est la longueur au carré du segment en rouge, donc 9. Comme on peut s'y attendre, la distance du point N à la droite D est inférieur (égal à 1) car le point N est plus proche de D que le point M. Et si un point appartient à la droite D alors sa distance à la droite D est nulle (vous pouvez le vérifier par le calcul avec le point $P(4; 2)$).



La distance d'un ensemble E constitué de plusieurs points à une droite D est alors définie comme la somme des distances des points de l'ensemble E à la droite D.

Définition : Si on note E l'ensemble des points A, B, C, ... c'est à dire $E = \{A, B, C, \ldots\}$, alors la distance de la droite D à l'ensemble E que l'on note $d(D, E)$ est définit par :

$$d(D, E) = d(D, A) + d(D, B) + d(D, C) + \ldots \tag{2}$$

On définit alors la droite de régression par :

Définition : Soit E un ensemble de points du plan, on appelle **droite de régression** de l'ensemble E, la droite telle que la distance de l'ensemble E à cette droite soit minimal.

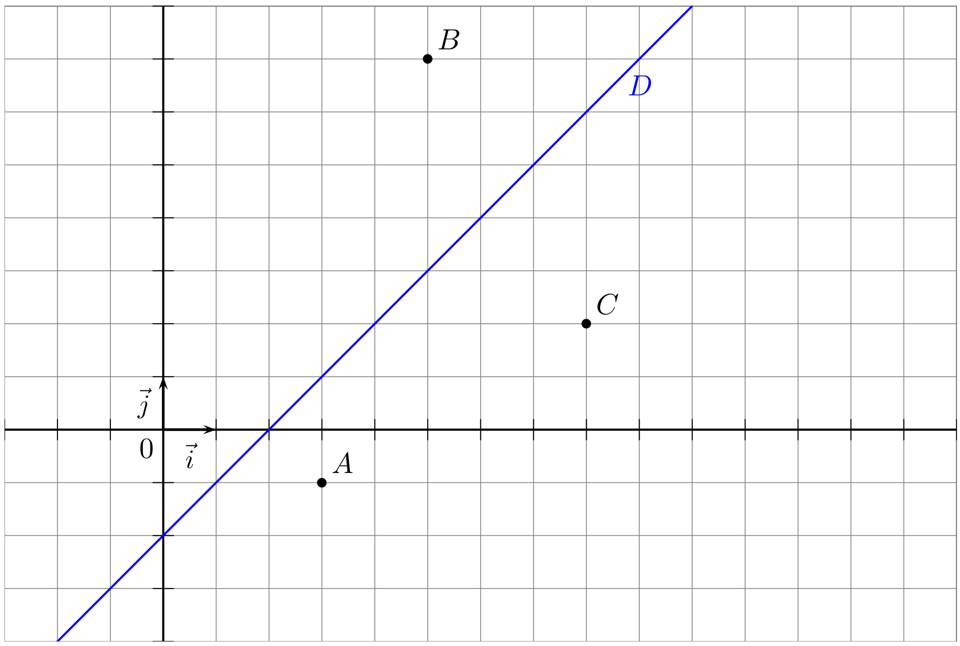
Remarques

1. Un premier travail serait de démontrer qu'une telle droite existe et, est unique. Nous admettrons ce résultat.

2. Il est tout à fait possible de définir d'une autre manière la distance d'un ensemble de points à une droite. On peut, par exemple, considérer la valeur absolue ou une puissance supérieure à 2 à la place du carré ou ne pas faire la somme des distances mais considérer la plus grande des distances. Chaque choix définie une nouvelle distance qui ne donnera peut-être pas la même droite de régression.

2 Exercice d'application

On considère les trois points du plan : $A(3;-1)$, $B(5;7)$ et $C(8;2)$ et la droite D d'équation $y = x - 2$ représentée dans le repère ci-dessous :



1. Déterminer graphiquement puis par le calcul $d(A;D)$; $d(B;D)$ puis $d(C;D)$.
2. En déduire $d(E;D)$ la distance de l'ensemble E à la droite D.
3. Déterminer soit graphiquement, soit par le calcul, $d(E;(AC))$.
4. Tracer une droite qui vous semble plus proche de l'ensemble E, déterminer alors la distance de l'ensemble E à cette nouvelle droite. Avez-vous trouver une droite plus proche de E que la droite D ou la droite (AC) ?
5. A la calculatrice, creer un programme qui prend en entrée $(a;b)$ le coefficient directeur et l'ordonnée à l'origine d'une droite et qui renvoie la distance de cette dernière à l'ensemble de points E. (Ce programme permettra de tester rapidement plusieurs droites)