

Math Data

Former à l'IA et favoriser l'émergence
d'une **culture citoyenne de l'IA** -
Enseigner les **mathématiques du lycée**
via des **challenges de données**



Anne Keller

Professeure de mathématiques

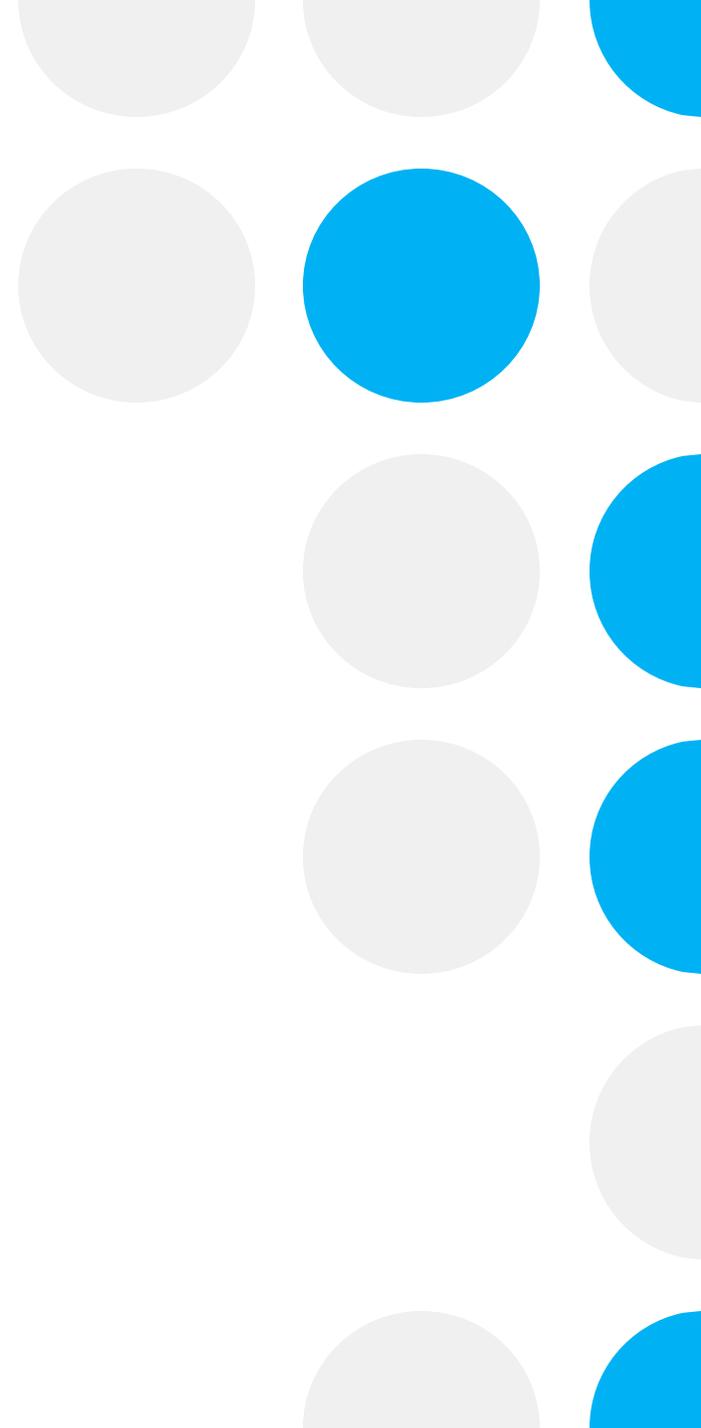
Formatrice académique

*INSPE de l'académie de Lille, EAFC de
l'académie de Lille*

A circular portrait of Louis Pasteur, a French chemist and microbiologist, is centered on the right side of the page. The portrait is overlaid with several semi-transparent blue circles of varying sizes. The text 'Lycée Louis Pasteur Lille' is written in a bold, dark blue font across the center of the portrait.

**Lycée Louis
Pasteur
Lille**

- Qu'est-ce que MathAData ?
 - Motivation et objectifs poursuivis
 - Qu'est-ce qu'un challenge de données ?
 - Une démarche de co-développement
 - Gros plan sur un module d'enseignement de
MathAData
 - À l'échelle d'un établissement
-



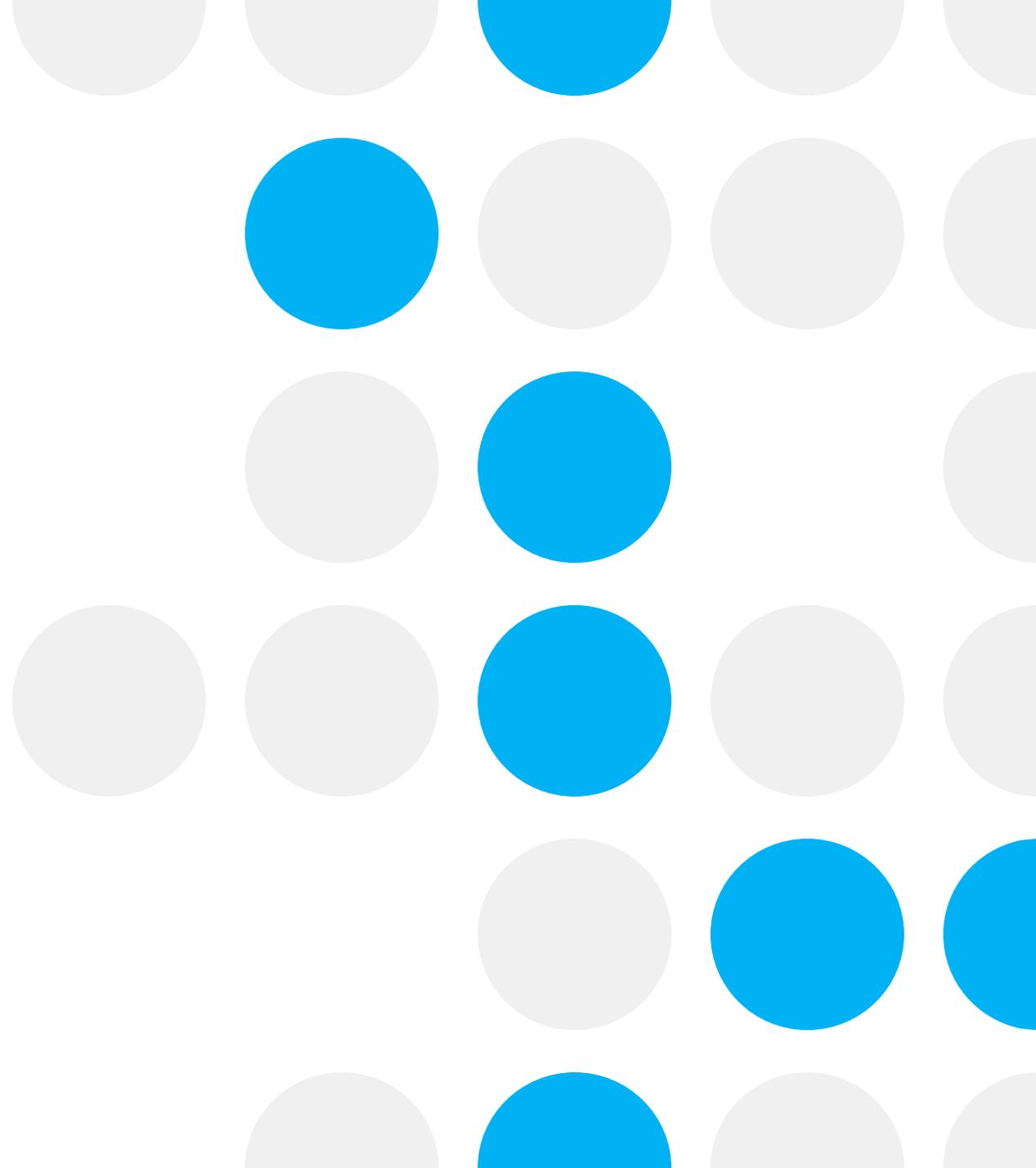


Stéphane MALLAT

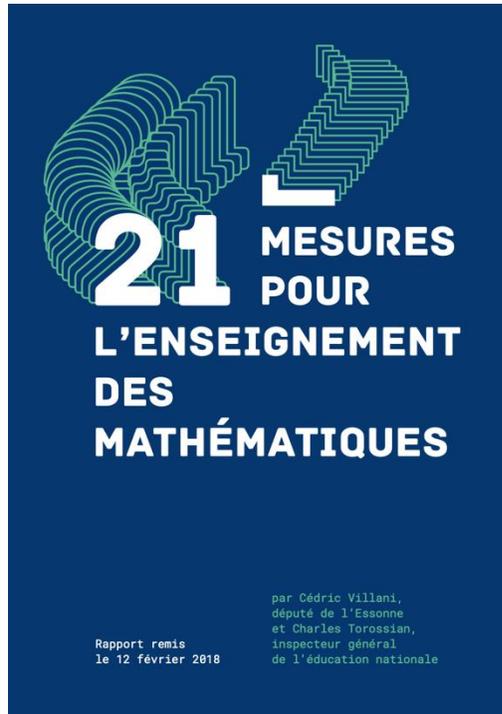
Chaire de sciences des données

Collège de France

challengedata.ens.fr



Motivation et objectifs



ChALLENGE Data

By MathA



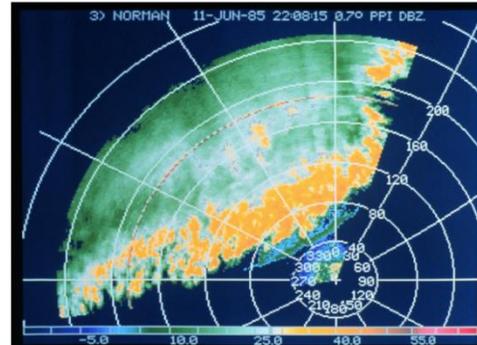
Anticipez l'affluence au sein des gares SNCF-Transilien !

- Challenge compétitif
- Sciences économiques
- Industriel
- Régression



Football : Qui va gagner ?

- Challenge compétitif
- Sport
- Classification



- Challenge compétitif
- Physique
- Environnement
- Régression
- Séries temporelles
- Plus de 1Go
- Niveau avancé



Prédiction de prix de l'électricité

- Challenge compétitif
- Sciences économiques
- Industriel
- Classification
- Régression

Un cadre unique pour toutes ces applications

Donnée

d



Algorithme

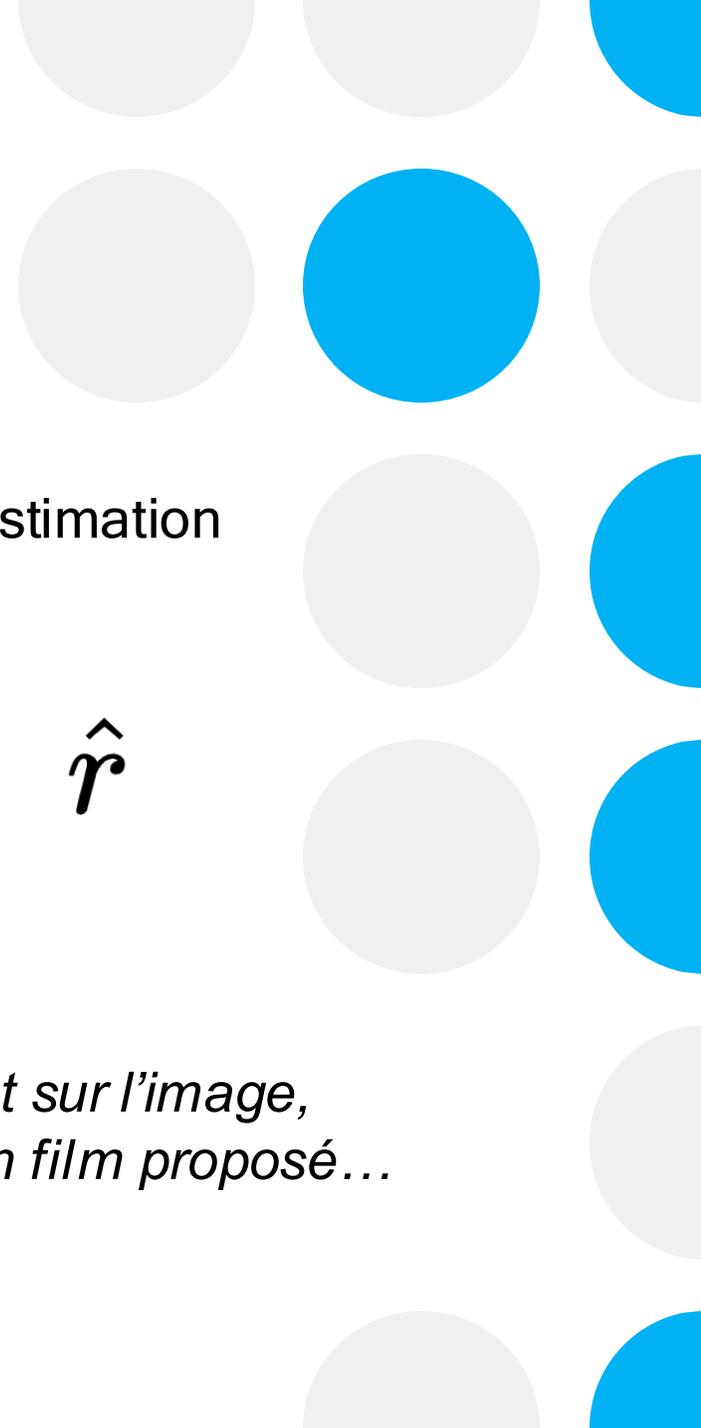


Estimation

\hat{r}

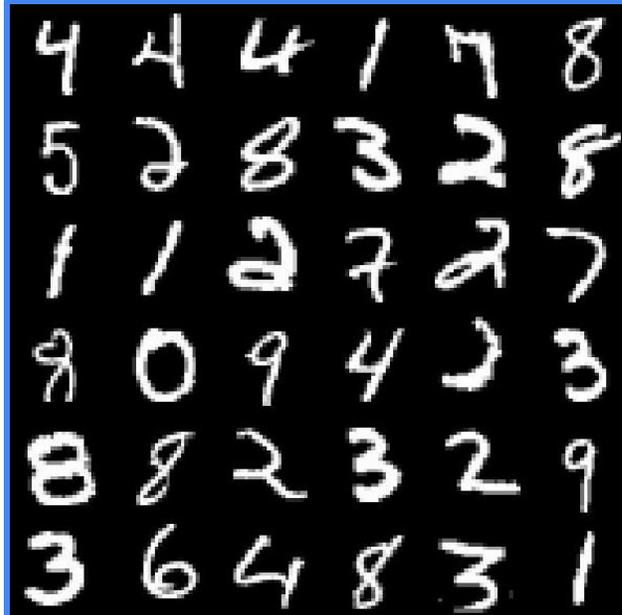
*Images, historiques
de vos films...*

*Objet sur l'image,
prochain film proposé...*



Un module d'enseignement MathAData, c'est...

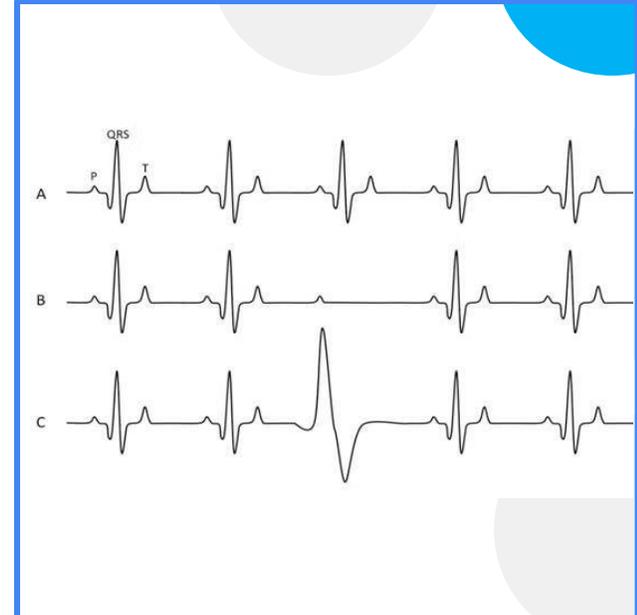
Un problème concret sur des thématiques variées



Reconnaissance de chiffres sur une image



Reconnaissance d'animaux à partir d'un son de sonar



Détection de troubles cardiaques sur un ECG

Concrètement, un module MathAData c'est...

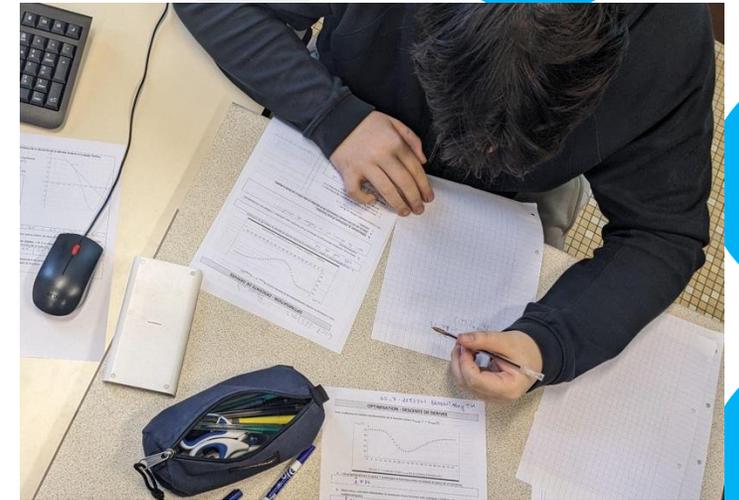
Des séquences organisées en 3 temps



Cours en classe
avec discussion pour
susciter questionnement
chez les élèves



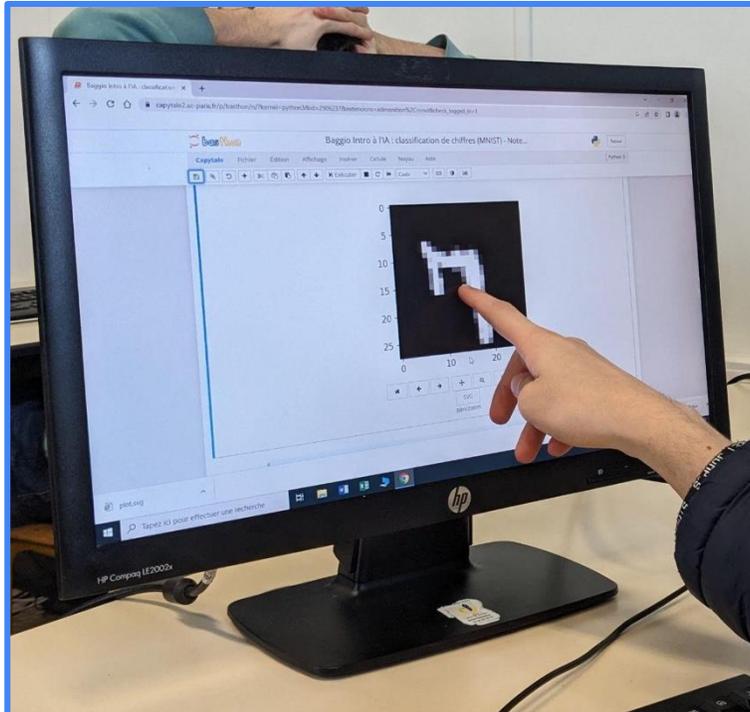
Expérimentation des élèves sur
Notebooks



Exercices sur feuilles
pour pratiquer,
s'entraîner et
automatiser

Un challenge d'IA MathAData c'est...

Une phase de recherche maniplatoire par les élèves



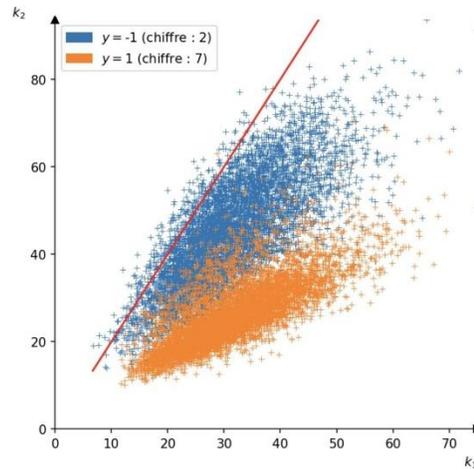
Manipulation des données et des objets mathématiques grâce à des **Notebooks Python**

On rappelle que l'équation est $x_2 = mx_1 + p$

Entrée[4]:

```
m = 2  
p = 0
```

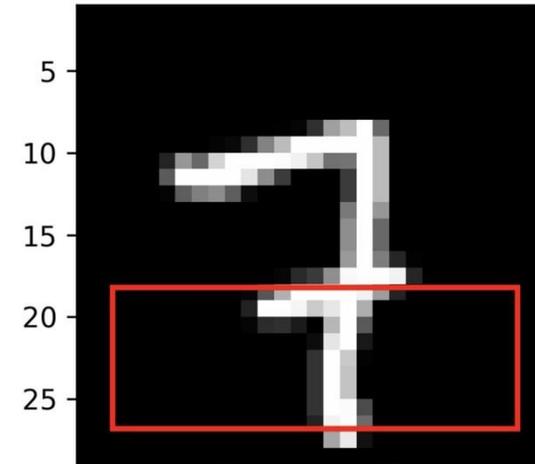
```
tracer_separatrice(m, p, c_train)
```



Les élèves testent leurs idées : droit à l'erreur et recherche expérimentale

```
1 # Complétez avec des valeurs entre 1 et 28  
2 numero_ligne_debut = ...  
3 numero_ligne_fin = ...  
4 numero_colonne_debut = ...  
5 numero_colonne_fin = ...
```

Figure 1



Pas de prérequis Python nécessaire grâce à des **Notebooks à compléter**

Un module d'enseignement MathAData, c'est...

Des ressources éditables et adaptables pour chaque étape de l'enseignement

Modéliser un challenge

Apprendre de nouveaux outils mathématiques

Manipuler, expérimenter sur ordinateur

S'entraîner, calculer, démontrer

Bloc 1, modélisation : critère de discrimination = aire de recouvrement

Critère de discrimination = recouvrement des histogrammes

Le challenge de données

Exemple pour la reconnaissance chien/chat

Bloc 1, mathématique : calcul de l'estimateur

Objectif : comprendre et montrer que la position du point par rapport à la droite est donné par le signe d'un produit scalaire

Notions de géométrie analytique :

- Équation de droite
- Vecteur normal
- Produit scalaire pour projection

$$D : ax_1 + bx_2 + c = 0$$
$$\hat{y} = \begin{cases} 1 & \text{si } a\bar{x}_1 + b\bar{x}_2 + c \geq 0 \\ -1 & \text{si } a\bar{x}_1 + b\bar{x}_2 + c < 0 \end{cases}$$

Des slides : challenge, modélisation mathématique

~45 min

Intro à l'IA : classification de chiffres (MNIST) - Notebook s...

II.3. Classificateur linéaire

II.4 Minimisation de la fonction erreur

Figure 1

Figure 1

Erreur d'entraînement en fonction du paramètre seuil, MNIST 2 & 7

Des Notebooks pour manipuler les objets mathématiques + expérimenter leurs idées

~1h

Statistique et erreur

Erreur d'entraînement et histogrammes de la caractéristique

Séparation Linéaire

Séparation linéaire à deux classes dans le plan

Statistique et erreur

Erreur d'entraînement et histogrammes de la caractéristique

Exercices

- Soit B l'événement « x est mal classé ». Soit C l'événement « la classe y de x est égale à -1 ».
- Exprimer $P(B)$ en fonction de $P(A)$ et $P(C)$ en fonction de $P(C)$.
- Montrer que l'erreur moyenne d'entraînement est la probabilité de mal classer une image $\epsilon_{\text{moy}} = P(B) = 1 - P(A)$.
- On décide que minimiser l'erreur ϵ_{moy} revient à maximiser $P(A)$.

Des fiches d'exercices personnalisables
S'entraîner : calculs et démonstrations

Le programme MathAData : une démarche de co-développement

Entre l'inspection académique, les professeurs expérimentateurs et l'équipe MathAData

Février
2023

Lancement du programme

- Faire face à la désaffection grandissante pour les maths
- Liens entre le lycée, le supérieur et le monde professionnel
- Introduire de la manipulation et de l'expérimentation dans l'enseignement des maths au lycée

AGIR POUR
L'ÉDUCATION
— UN ENJEU SCIENTIFIQUE —
POUR LA SOCIÉTÉ

Octobre
2023

Co-développement avec l'académie de Lille

- Tests en classes
- Co-développement de ressources grâce aux labos math
- Préparation de supports de formation


ACADÉMIE
DE LILLE
*Liberté
Égalité
Fraternité*



Miguel Toquet, IA-IPR
Mathématiques, Académie de
Lille

Avril 2024

Co-développement entre Lille, Paris et MathAData

- Tests en classes à partir des ressources
- Intégration des retours de terrain
- Co-développement de nouvelles ressources

Octobre
2024

Premier passage à l'échelle dans les académies de Lille, Paris et Créteil

- **Lille : ≈ 50 profs dans 14 établissements / ≈ 2000 élèves concernés**



L'accompagnement proposé

→ pour permettre de tester des séances en classe

Formation en établissement (1/2 journée)

- Détail des modules
- Prise en main des Notebooks Python et des fiches éditables

Échanges en visio - préparation séances

- Individuelles ou en groupe (labo math)
- But : définir déroulé de quelques séances
- Personnalisation des supports

Retour d'expérience

Choix d'un thème + 1 niveau

- Ex : *probas conditionnelles en 1ère Spé*
- Fixer une date indicative cible d'une 1ère séance

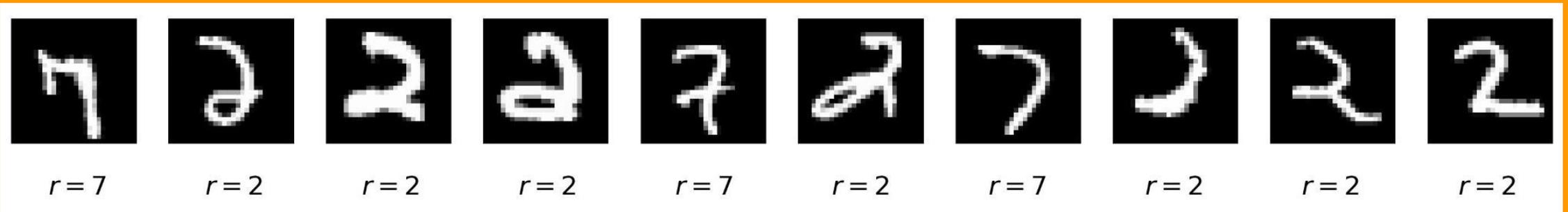
1ères séances en classe

- Avec notre équipe si souhaité
- Idéalement 2 séances : Intro puis 1 thème

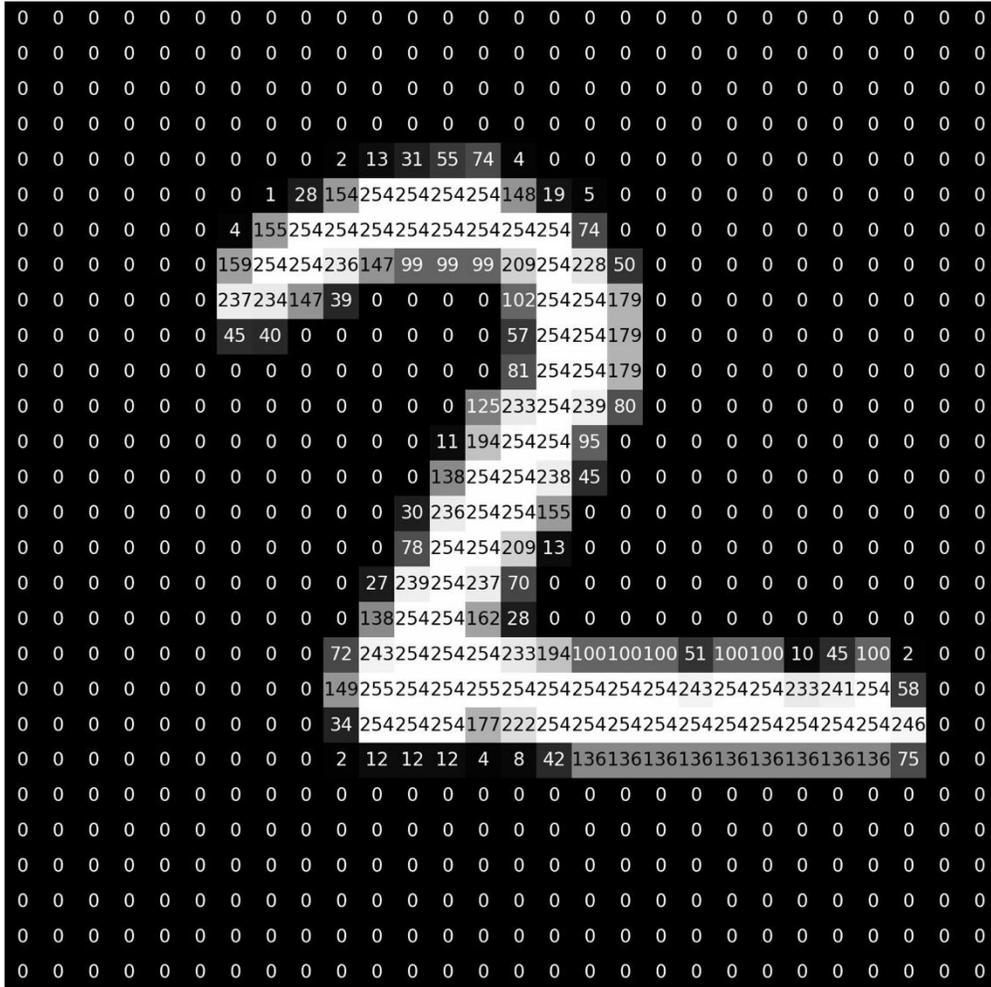
Mise en commun des ressources sur la plateforme MathAData et la page académique

Intelligence artificielle et classification d'images

But : comprendre comment l'intelligence artificielle identifie et classe des images

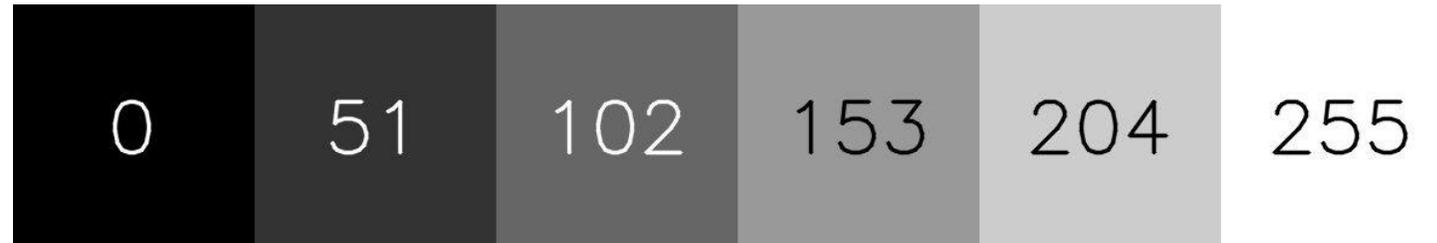


Qu'est-ce qu'une image pour l'ordinateur ?

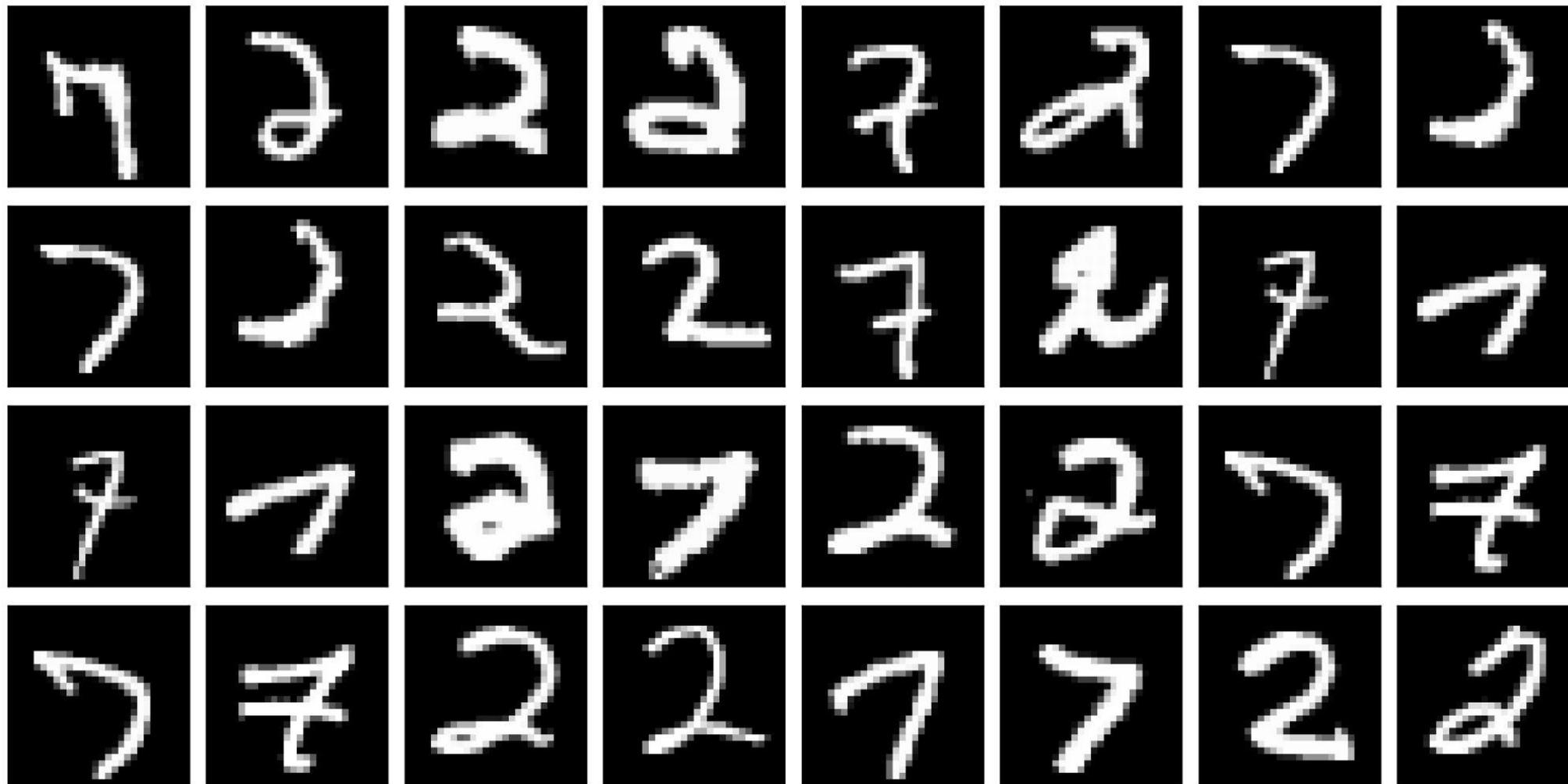


Images numériques : tableau de 28 x 28 pixels

- Chaque nombre : intensité lumineuse de chaque pixel
- 0 → pixel noir
- 255 → pixel blanc
- Valeur intermédiaire → gris plus ou moins clair



La base de données d'entraînement



Base de données d'entraînement : 12 000 images de chiffres de 2 et de 7.

Classification : comparaison à un seuil

Observations sur dix images :

- images de 7 : caractéristiques entre 20 et 30
- images de 2 : caractéristiques > 34

→ **Idée : comparer à un seuil x**

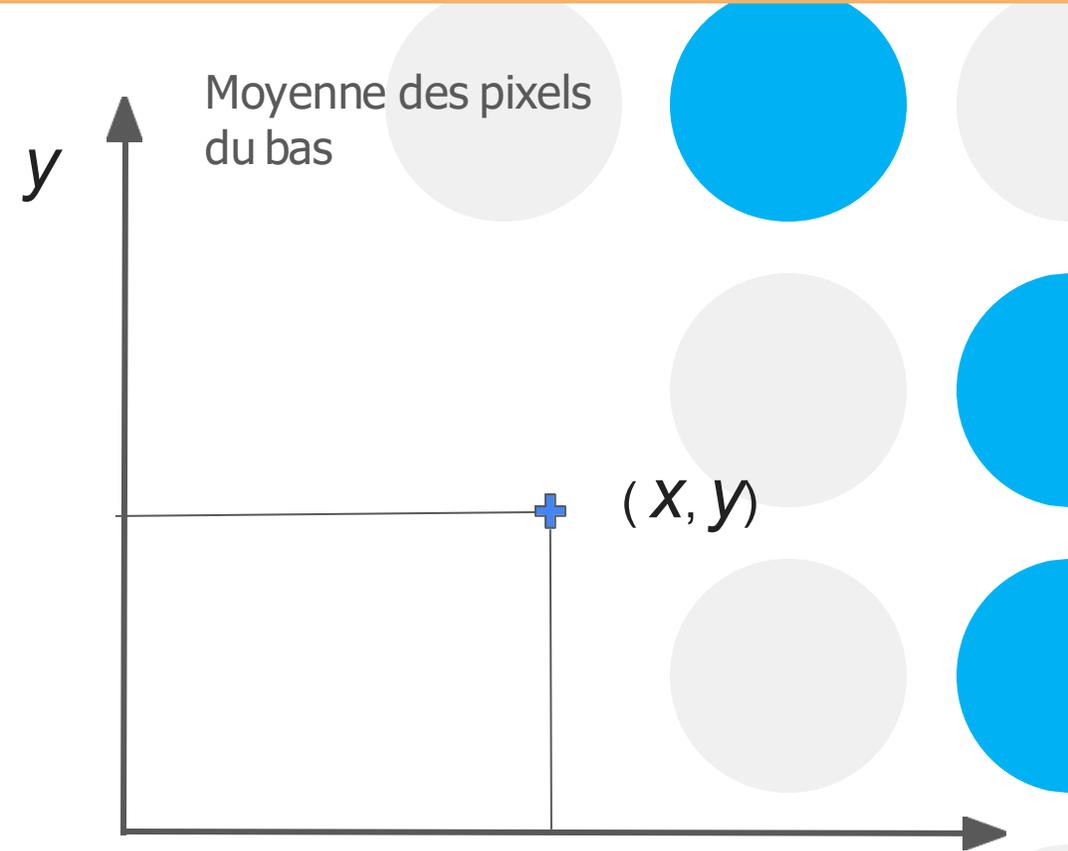
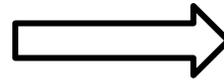
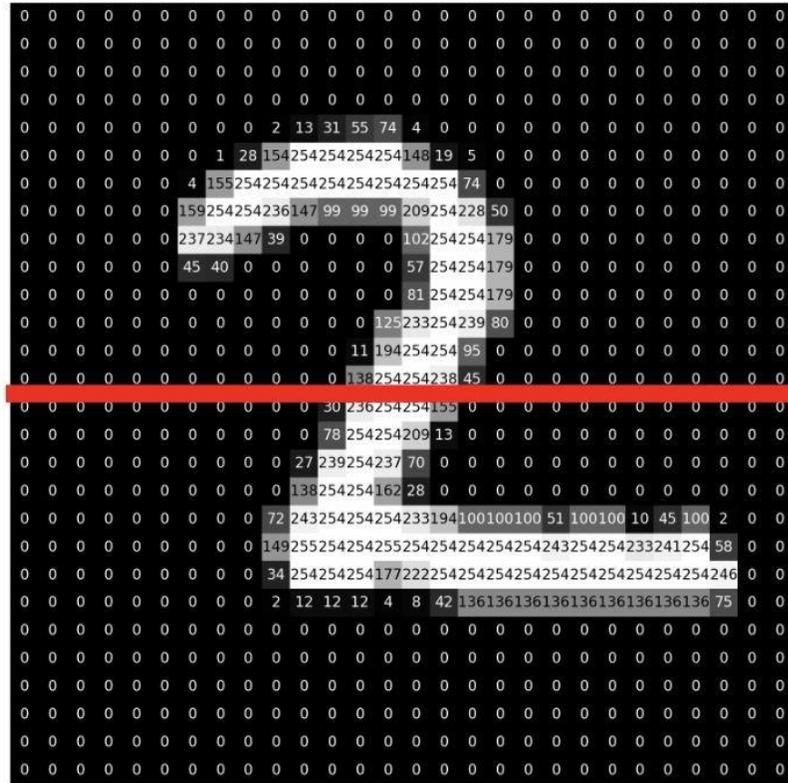
$$\hat{r} = \begin{cases} 7 & \text{si } k < x \\ 2 & \text{si } k \geq x \end{cases}$$

→ **Quelle valeur pour x ?**

Exemple : la moyenne

Chiffre	Caractéristique k
7	27
2	37
2	63
2	73
7	30
2	51
7	21
2	37
2	34
2	38
7	25
2	58

Caractéristiques : combiner intelligences humaine et artificielle



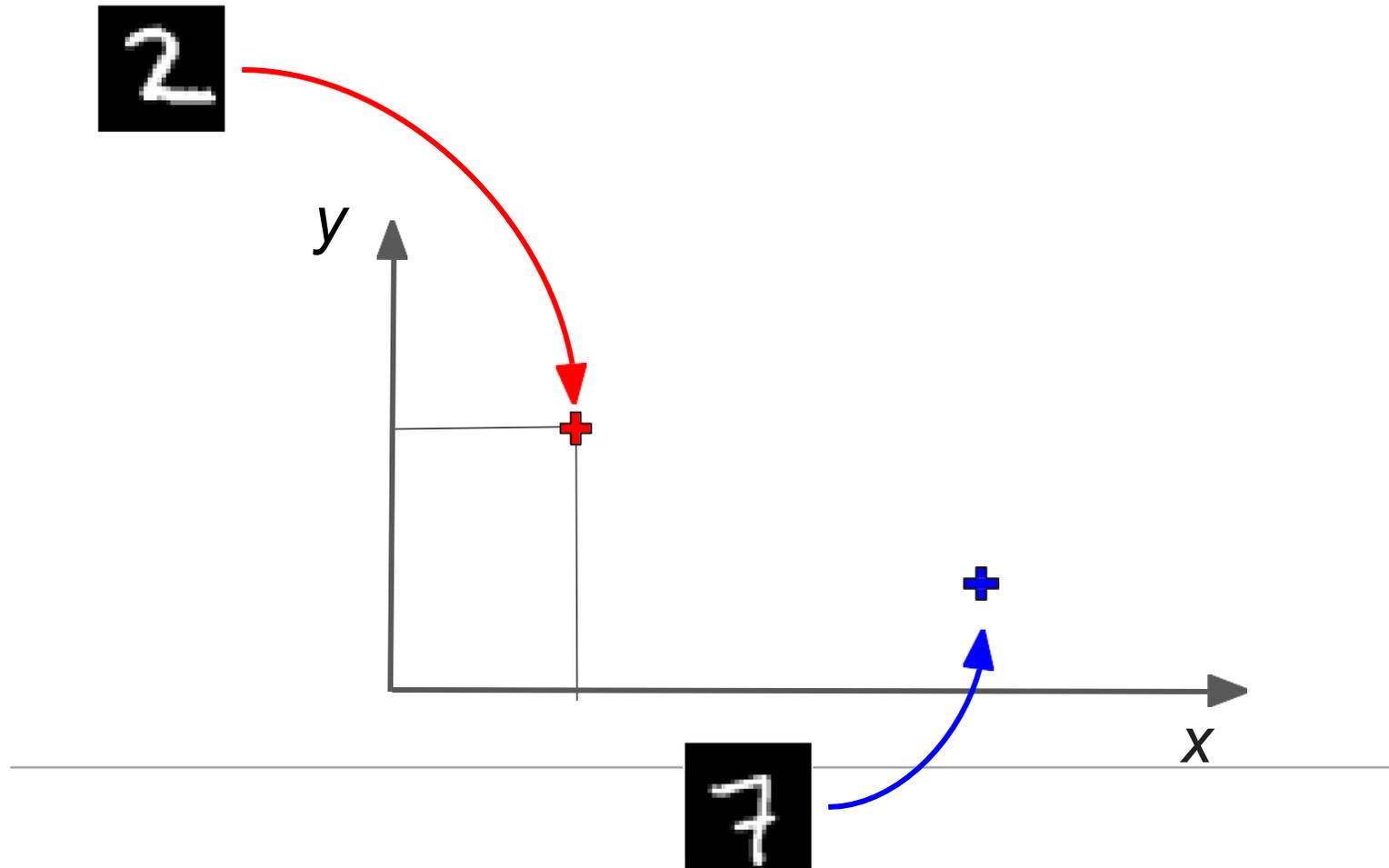
Moyenne des valeurs des pixels moitié haute et moitié basse

- Nombre facile à calculer pour un ordinateur
- Moyenne grande si beaucoup de pixels blancs, petite sinon

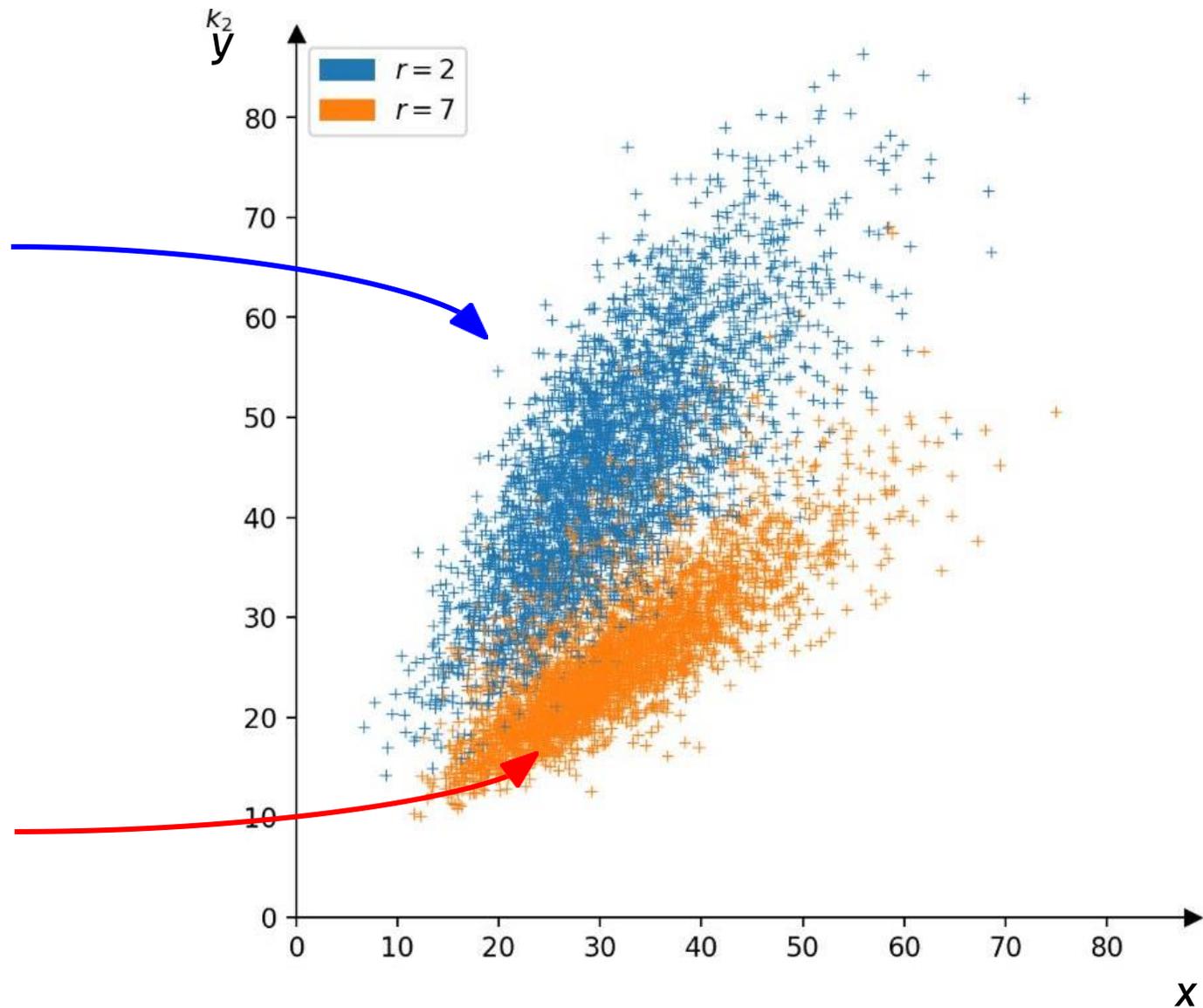
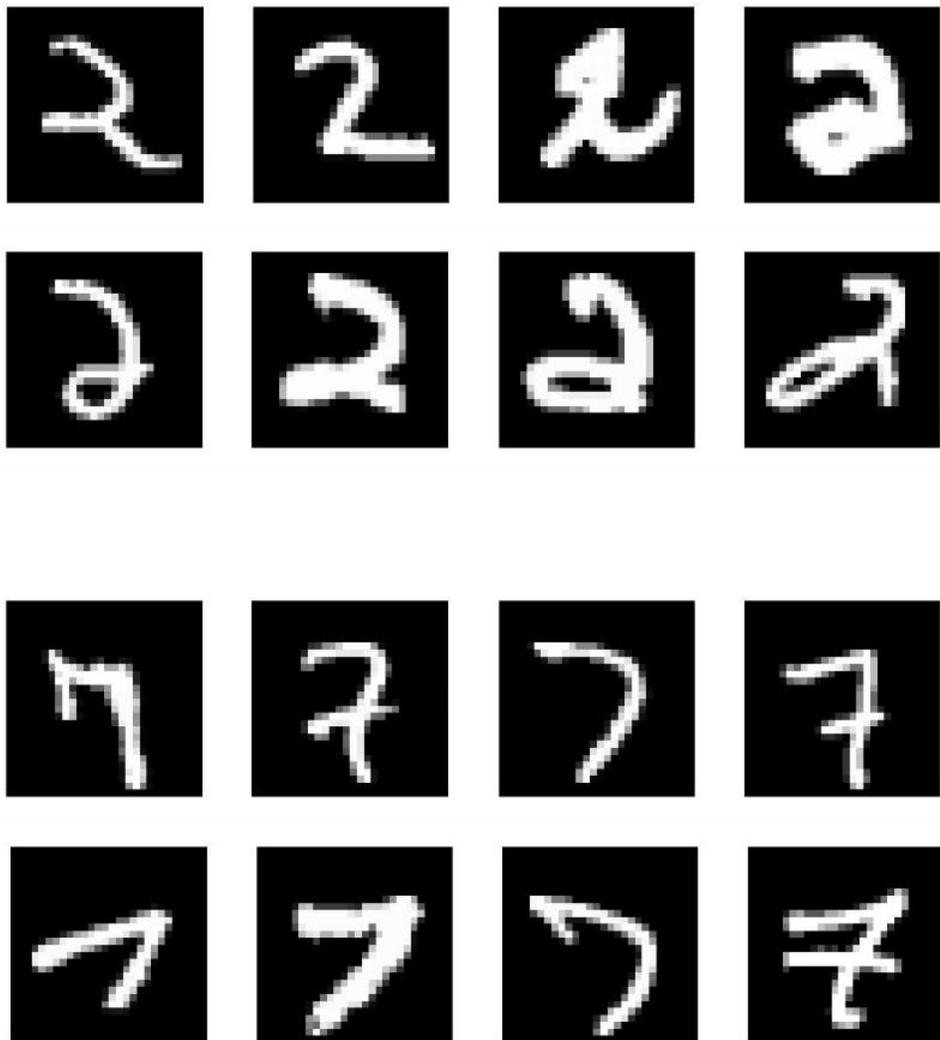
Représenter les images dans un plan pour les différencier

Idée centrale :

- représenter chaque image par **2 nombres caractéristiques** (x, y) , qui définissent **un point dans un plan**
- on veut que dans ce plan les images de 2 soient **séparées** des images de 7 .

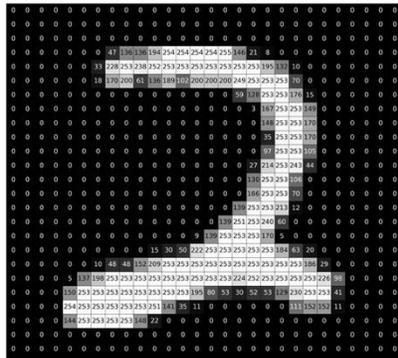


Moyenne en haut et en bas sur les 12 000 images d'entraînement

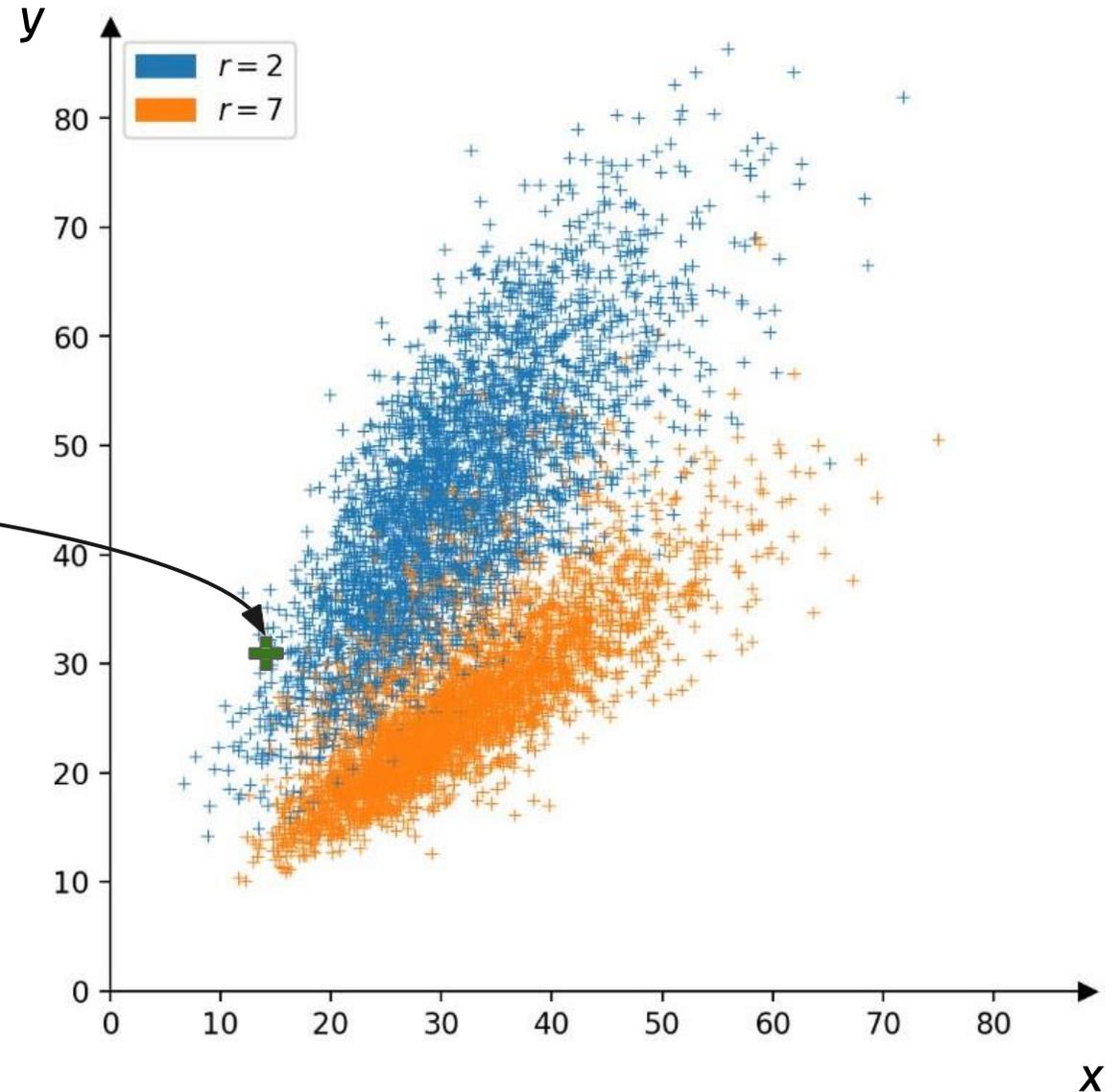


Classification dans le plan

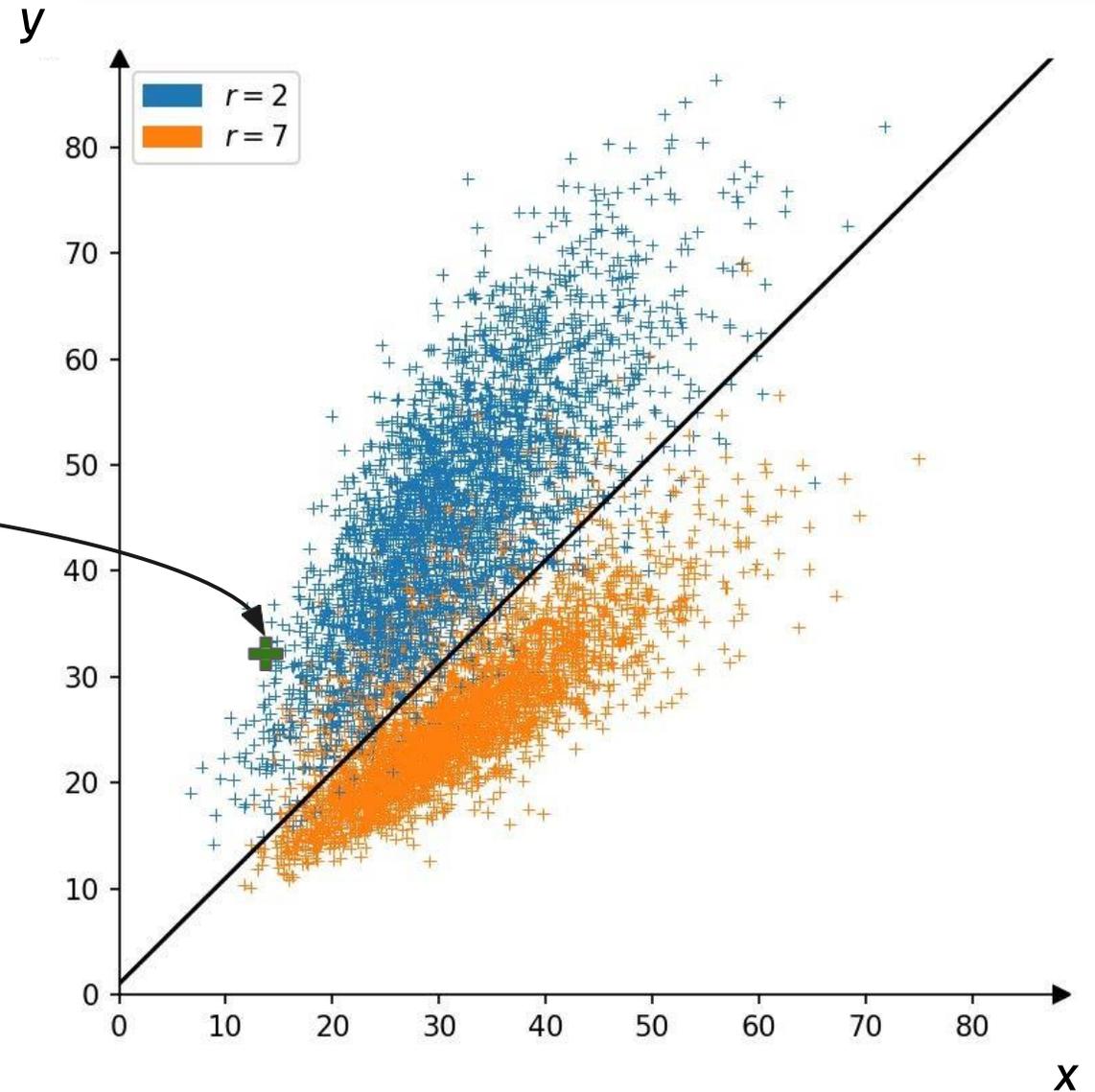
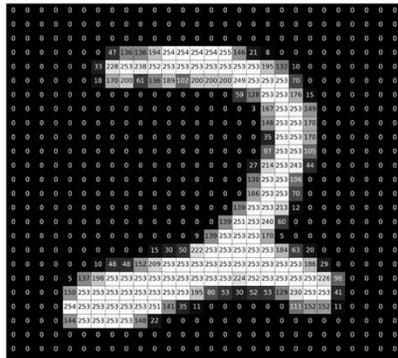
Comment classer notre image à partir de cette représentation ?



Quelle frontière simple à définir pour séparer les nuages de points et donc classer les images ?



Classification dans le plan

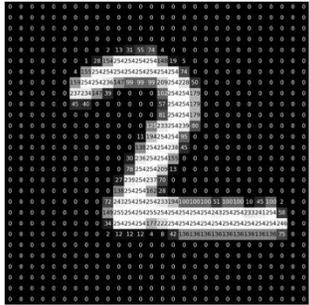


Idée : une droite

Résumé - Classificateur complet

plan

Image d



Calcul des caractéristiques

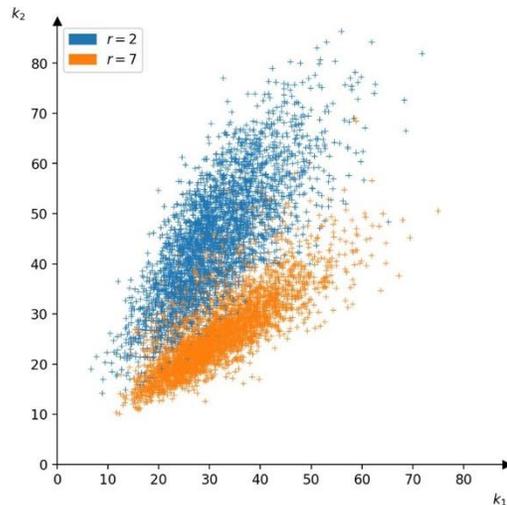
(X, y)

Classification :
Comparaison à
une droite
 $y = mx + p$

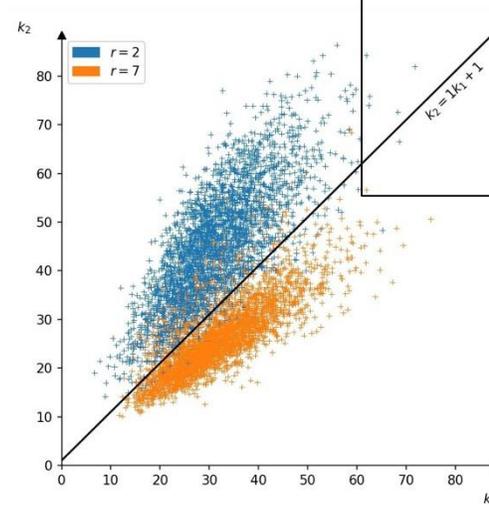
Estimation

\hat{r}

28 x 28 = 784
nombres



2 nombres



Réponse : 2 ou 7

Calcul
erreur

Apprentissage =
trouver la droite qui
minimise l'erreur

Math  Data

**À l'échelle
d'un
établissement**

**Partage
d'expérimentation
locale**
