# Projet d’examen de l’UE 2MTH2129

M1 promotion 2025-2026

Durée deux semaines

## Objectifs du projet

- **Réduire la dimensionnalité des données** : Utiliser l'ACP pour identifier les principales composantes qui expliquent la majorité de la variance dans l'ensemble de données, réduisant ainsi le nombre de variables à traiter pour la régression.

- **Construire un modèle de régression prédictif** : Appliquer la régression linéaire multivariée sur les composantes principales issues de l'ACP pour prédire la variable dépendante.
- **Optimiser l'entraînement via GPU** : Exploiter la puissance de calcul d'un GPU pour accélérer le processus d'entraînement, surtout sur des ensembles de données massifs où l'ACP et la régression peuvent être coûteuses en temps de calcul.

## Étapes du projet

**1. Préparation des données**

Séparer les variables explicatives et la variable cible.

1. Effectuer une mise à l'échelle des variables si nécessaire (par exemple, avec une standardisation), car l'ACP y est sensible.

2. **Analyse en Composantes Principales (ACP)**

**a.** Appliquer l'ACP sur l'ensemble des variables explicatives

b. Sélectionner le nombre de composantes principales qui retiennent une par suffisante de variance totale (par exemple 95% du total).

c. Transformer les données d'origine en un nouvel ensemble de données de dimensionnalié

réduite, composé des composantes principales.

NP : Faire tous les graphiques et interprétations possible de l’ACP

**3. Régression linéaire multivariée sur les composantes**

a. Utiliser les composantes principales comme nouvelles variables explicatives pour prédire la variable dépendante.

b. Entraîner le modèle de régression linéaire sur le jeu de données transformé.

c. Utiliser des bibliothèques comme TensorFlow ou PyTorch pour tirer parti de l'accélération GPU lors de l'entraînement.

4. **Évaluation et interprétation**

a. Évaluer la performance du modèle à l'aide de métriques appropriées (par exemple, le R^2, l'erreur quadratique moyenne).

b. Analyser les coefficients de régression sur les composantes principales pour comprendre leur impact sur la variable cible.

**Exemple de cas d'usage :** régression linéaire multiple appliquée aux données de tomates

Après une analyse à composantes principales, prédire des caractéristiques (rendement, qualité, teneur en nutriments) en utilisant plusieurs variables explicatives (engrais, lumière, eau, etc.), comme le montre une étude sur la prédiction de la productivité des parcelles de tomates avec un MAPE de 5%, ou des exemples illustrant comment l'engrais influence le rendement. Ces modèles permettent d'optimiser l'agriculture en comprenant les relations complexes entre les intrants et la production de tomates, souvent avec des données capturées par des systèmes d'imagerie ou des capteurs. En pièce jointe, vous avez deux fichiers csv (donnees_agro_part1.csv, donnees_agro_part2.csv) qu’il faut concaténer en un seul fichier. Les variables indépendantes (X) sont : Dose d'engrais, heures d'ensoleillement, volume d'eau, température, humidité, types de nutriments, et la variable dépendante (Y) est : Rendement (kg), taille des fruits, teneur en solides solubles (Brix), niveau de défauts.