

# Compte rendu TP2 : Optimisation Bayésienne et Modèles Bayésiens à Noyau

Sofiane Hachaichi

April 1, 2025

## 1 Introduction

Ce rapport présente l'optimisation bayésienne et les modèles bayésiens à noyau appliqués à la prédiction du rendement agricole, l'objectif est d'expliquer les concepts, implémenter les méthodes et analyser les résultats

## 2 Optimisation Bayésienne

### 2.1 Fondements Théoriques

#### 1. Qu'est-ce que l'optimisation bayésienne et comment gère-t-elle les fonctions coûteuses à évaluer ?

L'optimisation bayésienne est une méthode qui vise à optimiser une fonction coûteuse à évaluer en modélisant cette fonction à l'aide d'un processus gaussien, cette approche permet de prédire les valeurs possibles de la fonction à optimiser avec une incertitude associée, elle utilise ensuite une fonction d'acquisition pour déterminer les points à évaluer en minimisant le nombre total d'évaluations nécessaires

#### 2. Définition et utilité des processus gaussiens

Les processus gaussiens sont des modèles probabilistes définissant une distribution sur l'espace des fonctions continues, ils permettent d'estimer une fonction inconnue en fournissant une distribution sur ses valeurs possibles en chaque point, ils sont utilisés dans l'optimisation bayésienne car ils permettent de modéliser la fonction objective et d'obtenir une estimation de son incertitude

#### 3. Fonctions d'acquisition et compromis exploration/exploitation

Les fonctions d'acquisition permettent de choisir les points où évaluer la fonction objective en équilibrant l'exploration (test de nouvelles régions de l'espace) et l'exploitation (amélioration des points déjà prometteurs), les principales fonctions d'acquisition sont :

- **Expected Improvement (EI)** : favorise les points avec une amélioration attendue élevée

- **Upper Confidence Bound (UCB)** : privilégie les points avec une forte incertitude

## 2.2 Implémentation

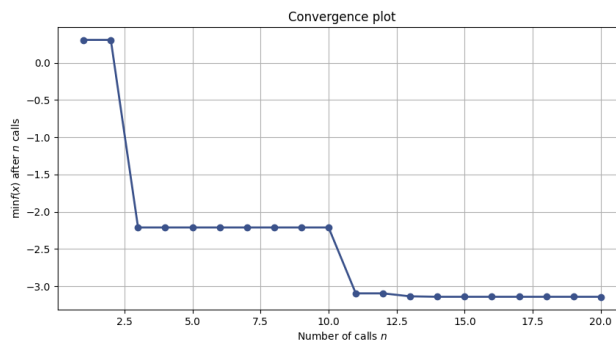
J'ai utilisé `gp_minimize` de la bibliothèque `skopt` pour optimiser la production agricole en fonction de la température et de l'humidité

- Fonction objectif : Modélisation du rendement
- Optimisation avec 20 itérations
- Affichage des résultats sous forme de courbe de convergence

## 2.3 Résultats

Les meilleures conditions trouvées sont :

- Température optimale : **15.66°C**
- Humidité optimale : **30.00%**



## 4. Avantages et limites de l'optimisation bayésienne

L'optimisation bayésienne est particulièrement efficace pour optimiser des fonctions coûteuses à évaluer, ses avantages incluent :

- Réduction du nombre d'évaluations nécessaires
- Prise en compte de l'incertitude grâce aux processus gaussiens

Cependant, elle présente aussi des limites :

- Complexité computationnelle accrue par rapport aux méthodes classiques
- Moins efficace pour des espaces de recherche de grande dimension

## 3 Modèles Bayésiens à Noyau

### 3.1 Fondements Théoriques

#### 5. Qu'est-ce que l'inférence bayésienne et comment met-on à jour les croyances ?

L'inférence bayésienne est une approche probabiliste permettant de mettre à jour les croyances sur un paramètre inconnu en intégrant de nouvelles données, cela repose sur le théorème de Bayes :

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \quad (1)$$

où  $P(\theta|D)$  est la distribution a posteriori,  $P(D|\theta)$  est la vraisemblance, et  $P(\theta)$  est la distribution a priori

#### 6. Théorie des méthodes à noyau et lien avec les processus gaussiens

Les méthodes à noyau permettent de projeter les données dans un espace de dimension plus élevée où elles deviennent plus facilement séparables, elles sont utilisées dans les processus gaussiens pour définir des fonctions de covariance qui quantifient la relation entre les points de données

#### 7. Distribution a priori et a posteriori appliquées à la prédiction de rendement agricole

Une distribution a priori représente les connaissances initiales avant observation des données, la distribution a posteriori est obtenue après mise à jour avec les nouvelles données, par exemple, dans la prédiction du rendement agricole, la distribution a priori pourrait être une estimation de rendement basée sur des connaissances agronomiques, et la distribution a posteriori serait affinée après analyse des données climatiques et du sol

### 3.2 Implémentation et Résultats

J'ai testé différents noyaux (linéaire, RBF, polynomial) pour une régression bayésienne et une classification bayésienne sur les types de sol

- Meilleur noyau : **RBF**
- Précision obtenue : **0.025 (MSE)**

Meilleurs hyperparamètres pour Random Forest :

- **n\_estimators : 200**
- **max\_depth : 20**

## 4 Conclusion

L'optimisation bayésienne offre un bon compromis entre exploration et exploitation, elle surpasse les méthodes classiques (Grid Search, Random Search) en

efficacité, les modèles bayésiens à noyau apportent une meilleure gestion de l'incertitude, permettant des prédictions plus fiables, les résultats ont montré que le choix du noyau a un impact significatif sur la précision du modèle, le noyau RBF s'est révélé le plus performant, offrant une bonne généralisation

En comparaison avec un SVM classique, la classification bayésienne a permis d'obtenir des prédictions avec une meilleure estimation de l'incertitude, ce qui est un atout pour des décisions plus robustes, de plus, la distribution a priori a influencé les résultats, une bonne sélection de cette distribution peut améliorer la convergence du modèle