

## Rapport TP3 ATDN :

# Optimisation Bayésienne et Modèles Bayésiens à Noyau

## Partie 1 : Optimisation Bayésienne

### Fondements théoriques

#### 1. Le principe de l'optimisation bayésienne et description de ce principe

L'optimisation bayésienne est une technique utilisée pour optimiser des fonctions coûteuses à évaluer.

Elle consiste à optimiser une fonction objective en sélectionnant les meilleurs points à évaluer grâce à une fonction d'acquisition en se basant sur une évaluation de la valeur et de l'incertitude effectuée par un modèle probabiliste.

L'idée principale est de construire un modèle probabiliste (souvent basé sur les processus gaussiens) qui approxime la fonction objective. Ensuite, ce modèle est utilisé pour choisir intelligemment les points suivants à évaluer, en équilibrant exploration (tester des zones inconnues) et exploitation (se concentrer sur les zones prometteuses).

#### 2. Définition des processus gaussiens et leur utilité pour modéliser la fonction objective

Un processus gaussien est un processus stochastique (une collection de variables aléatoires avec un index temporel ou spatial) de telle sorte que chaque collection finie de ces variables aléatoires suit une loi normale multidimensionnelle ; c'est-à-dire que chaque combinaison linéaire est normalement distribuée.

En résumé un processus gaussien est un modèle probabiliste qui permet de faire des prédictions sur une fonction inconnue en supposant que les valeurs de cette fonction suivent une distribution gaussienne (normale).

Ils sont utilisés pour modéliser la fonction objective car :

- Ils offrent une estimation de l'incertitude, ce qui permet de mieux gérer l'exploration/exploitation.
- Ils s'adaptent bien aux fonctions complexes et bruitées.
- Ils sont non paramétriques, donc flexibles pour représenter des formes de fonctions variées.

#### 3. Les principales fonctions d'acquisition (Expected Improvement, Upper

Confidence Bound, etc.)

Les fonctions d'acquisition sont utilisées pour décider où évaluer la fonction objective ensuite. Elles exploitent les prédictions du processus gaussien.

- Leur rôle dans le compromis exploration/exploitation

Expected Improvement (EI) :

- Favorise les points qui ont une forte probabilité d'améliorer le meilleur score actuel.
- Bon compromis entre exploration et exploitation.

Upper Confidence Bound (UCB):

- Sélectionne les points où l'incertitude est élevée, favorisant l'exploration.
- Contrôle l'équilibre exploration/exploitation avec un paramètre (plus il est élevé, plus on explore).

Exploration vs Exploitation

Exploration = tester des zones peu connues pour améliorer le modèle global.

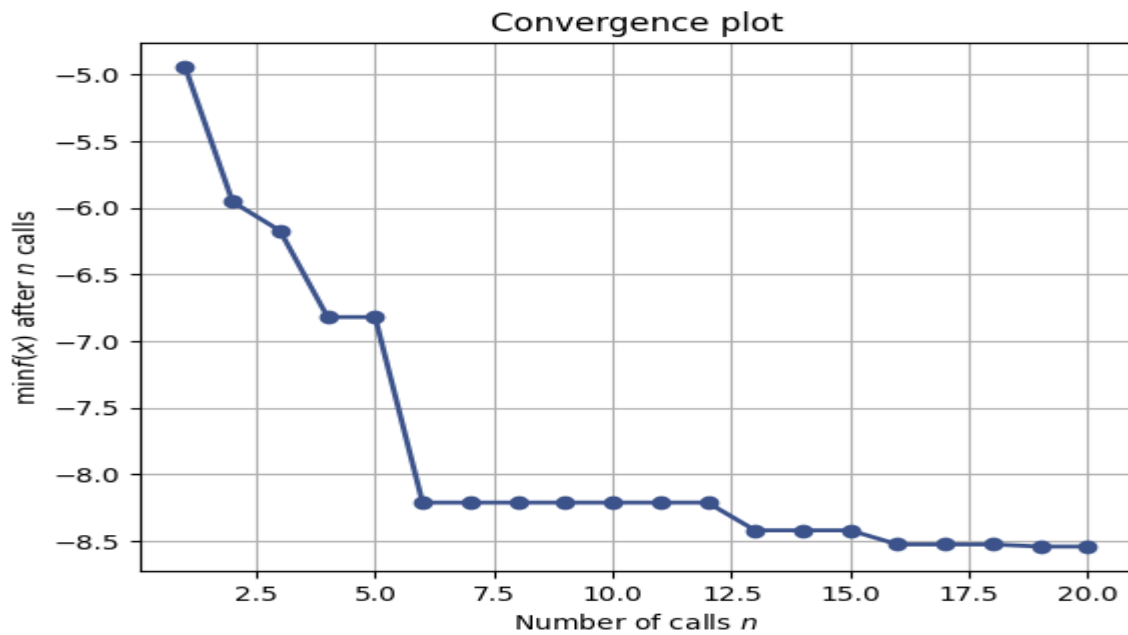
Exploitation = tester les zones déjà connues comme performantes pour affiner les résultats.

L'optimisation bayésienne cherche à équilibrer les deux pour éviter de passer à côté d'un optimum global.

## Implémentation et applications

4. Implémentation de l'optimisation bayésienne pour maximiser la production agricole en fonction de l'humidité et de la température et visualisation des étapes du processus.

```
Meilleures conditions trouvées :  
Humidité = 33.43 %  
Température = 27.76 °C  
Rendement agricole optimal = 8.54 t/ha
```

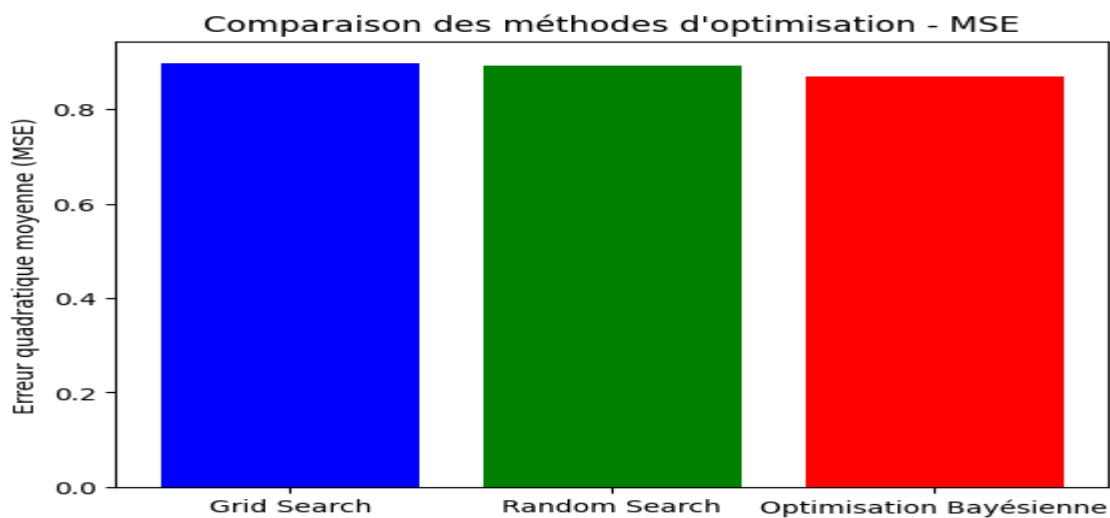


5. Ajustement des hyperparamètres d'un modèle de régression (ex : Random Forest) sur des données agricoles avec l'optimisation bayésienne.

```
Meilleurs hyperparamètres ( avec l'optimisation bayésienne) pour Random Forest :  
n_estimators = 50  
max_depth = 3  
Erreur quadratique moyenne (Optimisation Bayésienne) : 0.8697
```

- Comparaison des résultats avec Grid Search et Random Search.

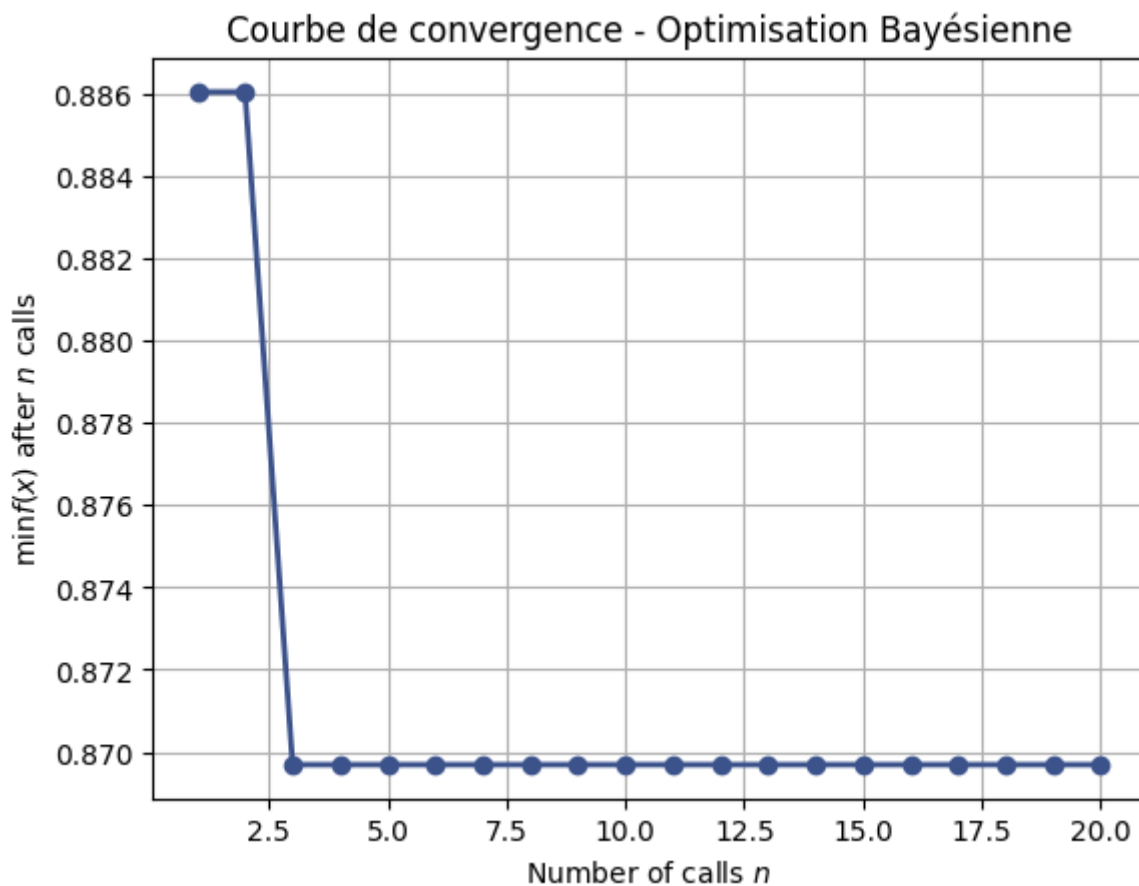
```
Meilleurs hyperparamètres (Grid Search) : {'max_depth': 5, 'n_estimators': 50}  
Erreur quadratique moyenne (Grid Search) : 0.8981  
Meilleurs hyperparamètres (Random Search) : {'max_depth': 4, 'n_estimators': 113}  
Erreur quadratique moyenne (Random Search) : 0.8915
```

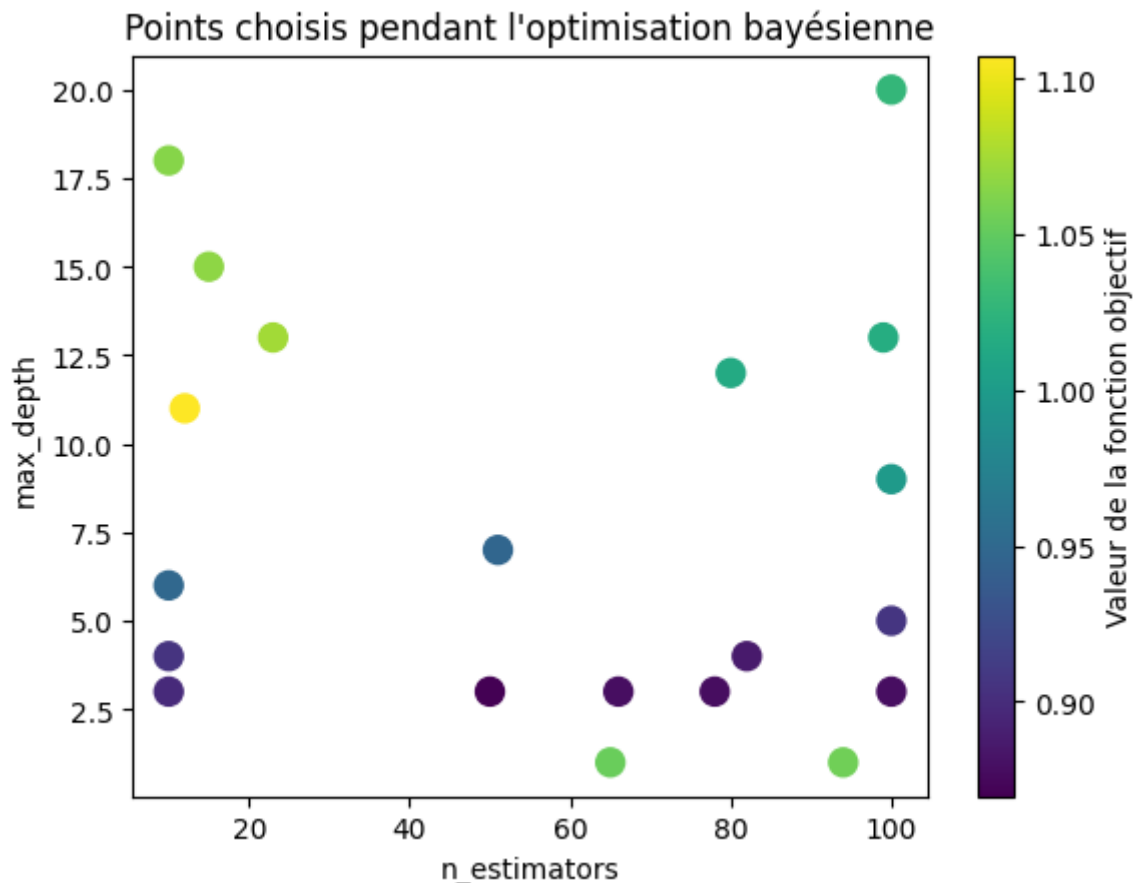


L'optimisation bayésienne donne la plus faible erreur quadratique moyenne, donc l'optimisation bayésienne a trouvé les meilleurs hyperparamètres (permet de mieux explorer l'espace des hyperparamètres) qui permettent de mieux prédire le rendement agricole. Ce qui indique une meilleure performance du modèle pour prédire les données de test.

En revanche les méthodes Grid Search et Random Search ont des erreurs quadratiques un peu plus élevé, ce qui montre que ces modèles ne sont pas aussi bon pour généraliser sur de nouvelles données, ce qui peut être dû à un mauvais choix d'hyperparamètres.

6. Visualisation du processus d'optimisation (courbe de convergence, choix des points).





- Commentaire sur la manière dont le modèle explore l'espace de recherche.

L'optimisation bayésienne commence généralement par explorer l'espace de manière plus large, en se concentrant sur les régions de l'espace qui semblent prometteuses.

À mesure que les itérations avancent, elle affine ses recherches autour des zones où elle a trouvé des solutions intéressantes, ce qui lui permet d'explorer plus finement les zones de l'espace qui ont un fort potentiel.

La courbe de convergence et les points choisis doivent montrer une progression vers les meilleures solutions avec une diminution de la fonction objectif au fil des itérations.

#### 7. Analyse des avantages et limites de l'optimisation bayésienne face aux méthodes classiques

Avantages de l'optimisation bayésienne :

- Efficacité dans l'exploration de l'espace : L'optimisation bayésienne est plus efficace que les méthodes classiques comme Grid Search et Random Search, car elle utilise un modèle probabiliste pour guider l'exploration. Elle se concentre sur les zones de l'espace des hyperparamètres qui sont plus susceptibles d'améliorer les performances du modèle, réduisant ainsi le nombre d'évaluations nécessaires.

- Exploration vs exploitation : Elle combine exploration (chercher des zones inconnues) et exploitation (affiner les zones prometteuses) de manière équilibrée. Cela permet de trouver des solutions optimales tout en réduisant les évaluations inutiles.
- Utilisation d'un modèle prédictif : L'optimisation bayésienne utilise un processus gaussien (ou un autre modèle probabiliste) pour prédire les valeurs de la fonction objective dans des zones non explorées, ce qui permet de faire des choix plus informés pour les itérations suivantes.

Limites de l'optimisation bayésienne :

- Complexité computationnelle : L'optimisation bayésienne peut être coûteuse en termes de calcul, notamment lorsque l'espace des hyperparamètres est très vaste ou lorsque le modèle utilisé pour approximer la fonction objective (comme un processus gaussien) est complexe à entraîner.
- Sensibilité aux hyperparamètres du modèle probabiliste : Les résultats de l'optimisation bayésienne peuvent dépendre des choix du modèle de base (comme le noyau du processus gaussien) et de ses hyperparamètres. Si ces choix sont mauvais, l'optimisation peut ne pas être efficace.
- Non-scaling avec un grand nombre d'hyperparamètres : L'optimisation bayésienne peut être moins efficace si l'espace des hyperparamètres est très grand (par exemple, avec des modèles ayant de nombreux hyperparamètres continus). Cela peut entraîner une exploration inefficace de l'espace.

Comparaison avec les méthodes classiques :

- Grid Search : Exhaustive, mais peut être extrêmement coûteuse en termes de temps et de calculs, surtout si l'espace des hyperparamètres est grand.
- Random Search : Plus rapide que Grid Search, car il explore l'espace de manière aléatoire, mais il ne garantit pas une exploration optimale de l'espace. Il peut passer à côté de zones prometteuses.
- Optimisation Bayésienne : Plus efficace, mais plus coûteuse en termes de calculs et sensible aux paramètres du modèle.

## Partie 2 : Modèles Bayésiens à Noyau

Fondements théoriques

8. Le concept d'inférence bayésienne et la mise à jour des croyances.

Concept d'inférence bayésienne :

L'inférence bayésienne est un cadre statistique qui permet de mettre à jour les croyances (probabilités) à propos d'un modèle ou d'une hypothèse à mesure que de nouvelles données deviennent disponibles. L'idée de base est d'utiliser la règle de Bayes pour ajuster les probabilités a priori (avant observation des données) en probabilités a posteriori (après observation des données).

La façon dont on met à jour les croyances avec de nouvelles données :

**Distribution a priori** : Avant d'observer les données, on commence avec une croyance sur la distribution de notre paramètre ou de notre modèle, exprimée par une distribution a priori.

**Données observées** : Lorsqu'on observe des données, on utilise la vraisemblance de ces données, qui décrit comment les données sont générées en fonction des paramètres.

**Distribution a posteriori** : On combine la distribution a priori et la vraisemblance des données pour obtenir une distribution a posteriori, qui est notre croyance mise à jour.

## 9. La théorie des méthodes à noyau et leur lien avec les processus gaussiens.

Théorie des méthodes à noyau :

Les méthodes à noyau sont une classe de techniques statistiques qui utilisent une fonction de noyau pour transformer les données dans un espace de plus haute dimension (généralement un espace de caractéristiques infinie dimension), tout en effectuant les calculs dans l'espace d'origine sans avoir besoin de connaître explicitement la transformation.

Les noyaux permettent de modéliser des relations complexes entre les variables sans avoir à spécifier explicitement une fonction de transformation.

Lien avec les processus gaussiens :

Dans les modèles bayésiens à noyau, on utilise un noyau pour définir une mesure de similarité entre les observations. Un processus gaussien (GP) est un modèle probabiliste qui génère des distributions sur des fonctions. Lorsqu'on applique un noyau à un GP, on détermine comment les points d'observation sont liés entre eux. Le noyau peut être utilisé pour déterminer la covariance entre les points, ce qui est essentiel pour prédire les valeurs de nouvelles observations en se basant sur les observations précédentes.

Pourquoi utiliser un noyau dans un modèle bayésien ?

L'utilisation d'un noyau dans un modèle bayésien permet de capturer des relations complexes dans les données sans avoir besoin de spécifier explicitement la forme de la fonction sous-jacente. Cela rend le modèle plus flexible et puissant, surtout dans des situations où la fonction réelle est non linéaire ou difficile à modéliser directement.

## 10. Définition distribution a priori et distribution a posteriori.

Distribution a priori :

La distribution a priori représente les croyances ou hypothèses que l'on a sur les paramètres d'un modèle avant d'observer les données. Elle reflète nos connaissances ou nos suppositions sur les paramètres du modèle.

Distribution a posteriori :

La distribution a posteriori est la mise à jour de notre croyance sur les paramètres du modèle après avoir observé les données. Elle est obtenue en combinant la distribution a priori et la vraisemblance des données observées.

Exemple appliqué à la prédiction de rendement agricole :

Distribution a priori : Avant d'observer les données, on pourrait supposer que le rendement agricole est en moyenne autour de 30 t/ha, avec une certaine variabilité (par exemple, entre 20 et 40 t/ha). Cette croyance pourrait être représentée par une distribution normale.

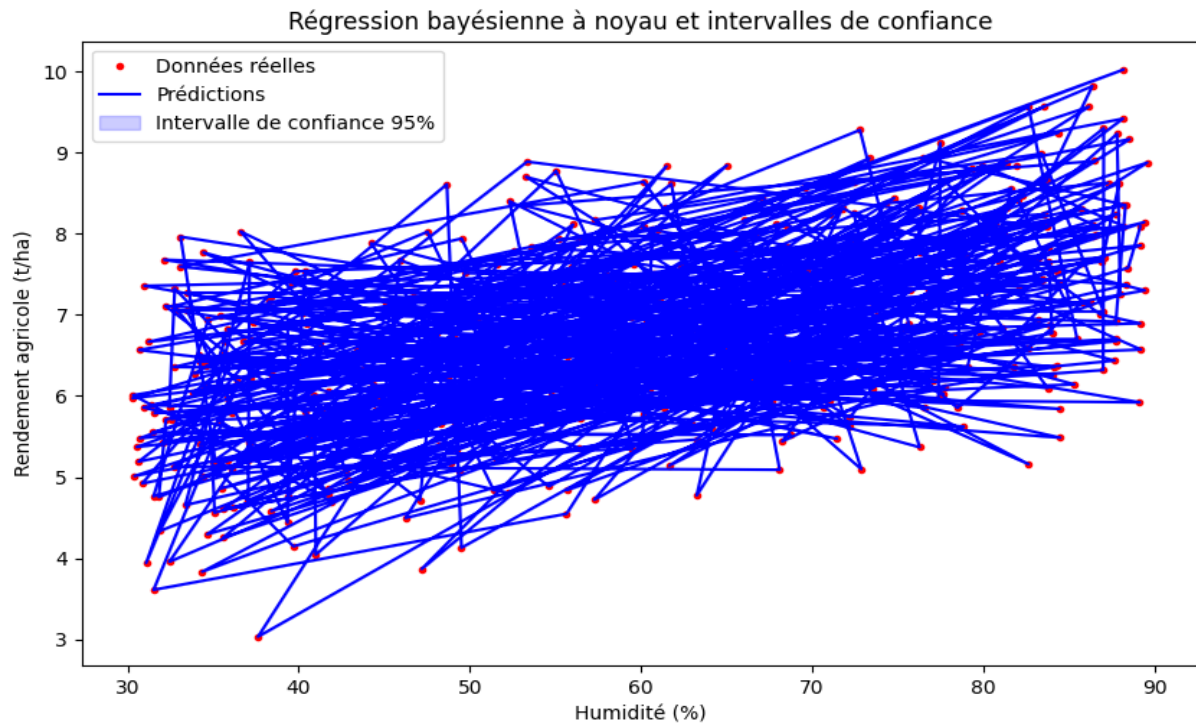
Distribution a posteriori : Après avoir observé des données sur l'humidité et la température, et après avoir effectué l'inférence bayésienne, la distribution a posteriori de notre modèle pourrait nous donner des valeurs mises à jour pour les rendements agricoles, tenant compte de l'effet des nouvelles observations. Par exemple, si l'humidité et la température montrent que le rendement est fortement influencé par ces facteurs, la distribution a posteriori pourrait avoir une forme différente, avec des rendements plus bas ou plus élevés en fonction des conditions observées.

Implémentation et applications

## 11. Implémentation d'une régression bayésienne à noyau sur les données

Visualisation des prédictions et des intervalles de confiance.



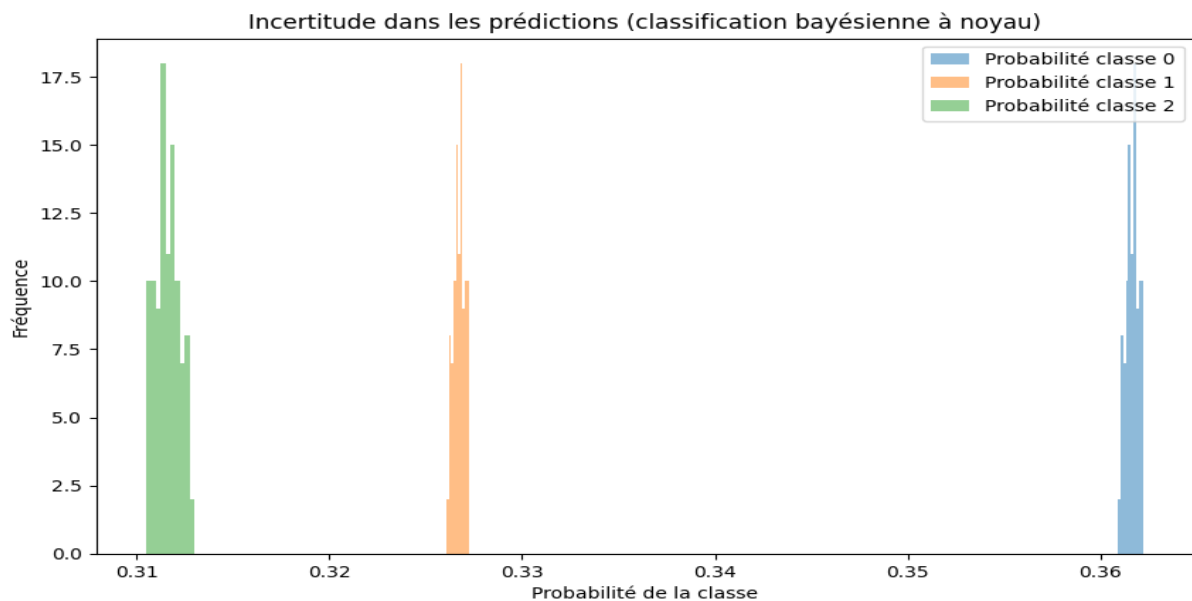


12. Réalisez une classification bayésienne à noyau pour prédire le type de sol (argileux, sableux, limoneux) en fonction des données climatiques.

- Comparaison des résultats avec un SVM classique.

Les performances des deux modèles sont similaires, et aucune méthode n'a montré un meilleur comportement dans ce cas précis. La classification bayésienne à noyau (GPC) et le SVM classique ont du mal pour prédire correctement les classes "Limoneux" et "Sableux".

13. Analyse de l'incertitude dans les prédictions.



- Commentaire des zones où le modèle est moins confiant :

Les probabilités des classes sont proches les unes des autres ( $\approx 0.31 - 0.36$ ), ce qui signifie que le modèle n'attribue pas une forte probabilité à une classe spécifique et hésite dans ses prédictions.

Donc le modèle est moins confiant dans la plage 0.31 - 0.36, car aucune classe ne prend un avantage clair sur les autres.

#### 14. Teste de différents noyaux (linéaire, RBF, polynomial).

- La différence entre les différents noyaux et leur impact sur la précision du modèle

Noyau RBF : est utilisé couramment pour des problèmes où la séparation entre les classes n'est pas linéaire. Ce noyau capture des relations complexes et non linéaires. Donc il est très efficace pour des classes avec des frontières complexes et non linéaires et permet une meilleure généralisation sur des données complexes.

Noyau linéaire : Il modélise une relation linéaire entre les points de données. Il est utile lorsque les classes sont linéairement séparables. Donc il est plus simple et plus efficace si les classes sont séparables linéairement.

Noyau polynomial : Ce noyau est utilisé pour capturer les relations polynomiales entre les caractéristiques, souvent utilisé lorsque les classes ne sont pas linéairement séparables mais peuvent être séparées par un polynôme.

#### 15. L'influence des choix du noyau et de la distribution a priori sur les résultats.

Influence des choix de noyau :

Le choix du noyau influence directement la manière dont le modèle apprend et généralise à partir des données :

Le noyau linéaire est simple et rapide, mais il peut ne pas capturer les relations non linéaires.

Le noyau RBF est flexible et peut modéliser des relations complexes, mais il peut aussi être sensible au choix de ses hyperparamètres (comme la longueur d'échelle).

Le noyau polynomial permet de modéliser des relations plus flexibles que le noyau linéaire, mais peut entraîner un sur-ajustement sur des données bruyantes.

Influence de la distribution a priori :

La distribution a priori dans un modèle bayésien permet de mettre à jour nos croyances initiales avant d'observer les données. Elle affecte fortement les prédictions, surtout dans les cas où les données sont peu informatives. Une distribution a priori plus large ou plus informée peut aider le modèle à être plus robuste, tandis qu'une distribution trop stricte peut limiter la capacité du modèle à bien s'ajuster aux données réelles.

## Conclusion

Dans ce TP nous avons exploré l'optimisation bayésienne et la classification bayésienne à noyau, mettant en évidence leur efficacité dans l'optimisation des hyperparamètres et la gestion de l'incertitude.

L'optimisation bayésienne permet une recherche intelligente des paramètres en limitant les évaluations coûteuses, tandis que l'analyse de l'incertitude dans les prédictions aide à identifier les zones où le modèle est moins confiant.

Le choix du noyau a un impact crucial sur la précision du modèle :

Linéaire : efficace pour des données séparables simplement.

RBF : puissant pour capturer des relations complexes.

Polynomial : bon compromis entre flexibilité et interprétabilité.

En combinant ces approches, nous obtenons une modélisation robuste, capable de s'adapter aux structures des données tout en réduisant les erreurs d'approximation et d'incertitude.