

Compte rendu TP3 ATDN2

Optimisation Bayésienne et Modèles Bayésiens à Noyau

Houda SLIMANI

Partie 1 : Optimisation Bayésienne

1. Principe de l'optimisation bayésienne :

L'optimisation bayésienne est une technique d'optimisation globale particulièrement utile lorsque l'évaluation d'une fonction objective est coûteuse. Contrairement à des méthodes comme Grid Search ou Random Search, l'optimisation bayésienne construit un modèle probabiliste de la fonction à optimiser, en utilisant généralement un processus gaussien. Ce modèle prédit la fonction objective et son incertitude, et guide la recherche des points à évaluer en choisissant ceux qui promettent une amélioration de la fonction objective tout en tenant compte de l'incertitude dans les prédictions.

2. Processus Gaussiens :

Les processus gaussiens sont des modèles probabilistes non paramétriques utilisés pour estimer des fonctions. Ils sont définis par une fonction de covariance, qui mesure la similarité entre les points dans l'espace de recherche. Ces processus sont utilisés dans l'optimisation bayésienne pour modéliser la fonction objective et prédire les résultats à des points non évalués tout en capturant l'incertitude associée à ces prédictions. Cette incertitude permet à l'optimisation de prendre des décisions informées sur les prochains points à évaluer.

3. Fonctions d'acquisition :

Les fonctions d'acquisition, comme l'Expected Improvement (EI), l'Upper Confidence Bound (UCB) et la Probability of Improvement (PI), sont utilisées pour guider l'optimisation bayésienne en équilibrant exploration et exploitation.

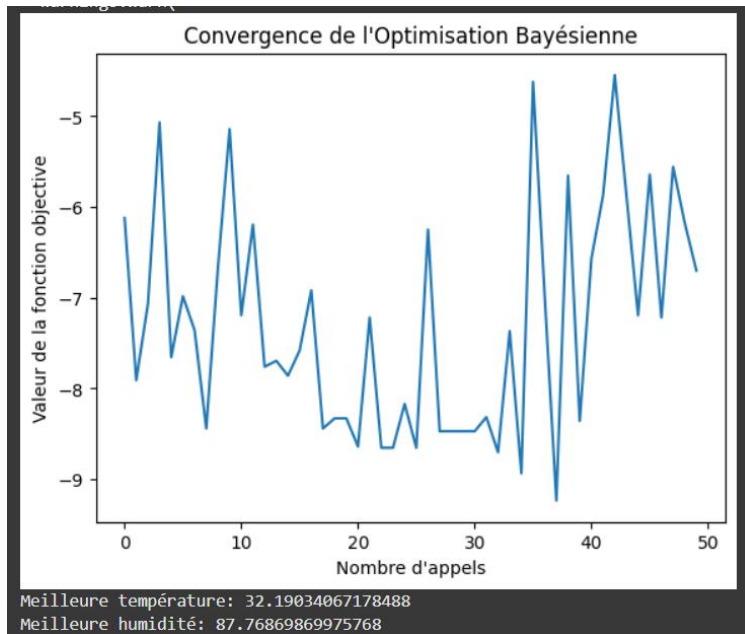
- **Exploration** : tester des zones de l'espace de recherche avec une grande incertitude.
- **Exploitation** : tester des zones où le modèle prédit que la fonction objective pourrait être maximisée.

Ces fonctions sont calculées en fonction de l'incertitude du modèle et de la possibilité d'améliorer la fonction objective.

4. Implémentation de l'optimisation bayésienne pour maximiser la production agricole en fonction de l'humidité et de la température :

5. Utilisation de l'optimisation bayésienne pour ajuster les hyperparamètres d'un modèle de régression .

6. Visualisation de l'optimisation :



```
Meilleurs hyperparamètres : OrderedDict([('max_depth', 30), ('min_samples_split', 20), ('n_estimators', 221)])  
RMSE : 0.9421829570918757
```

7. Analyse des avantages et limites de l'optimisation bayésienne

Si le temps de calcul est limité ou si l'espace de recherche est vaste, Random Search est généralement plus performant. En revanche, si l'objectif est une exploration approfondie, Grid Search peut être plus approprié.

Partie 2 : Modèles Bayésiens à Noyau

8. Inférence bayésienne :

L'inférence bayésienne est un processus dans lequel les croyances ou les connaissances initiales (croyances a priori) sont mises à jour avec de nouvelles données. Cette mise à jour

se fait par la règle de Bayes, qui ajuste les probabilités a priori pour obtenir une probabilité a posteriori plus précise après l'observation de nouvelles données. Dans le contexte de l'optimisation ou de la régression, cette approche permet d'incorporer l'incertitude dans les prédictions et d'affiner les modèles en fonction des nouvelles informations.

9. Méthodes à noyau :

Les méthodes à noyau sont des techniques d'apprentissage automatique qui utilisent une fonction de noyau pour projeter des données dans un espace de dimension plus élevée, où elles deviennent plus faciles à séparer ou à modéliser. Les noyaux comme le linéaire, RBF (Radial Basis Function) ou polynomial sont utilisés pour transformer les données de manière non linéaire. Dans les modèles bayésiens, les noyaux permettent de capturer des relations complexes entre les variables, rendant les modèles plus flexibles.

10. Distribution a priori et a posteriori :

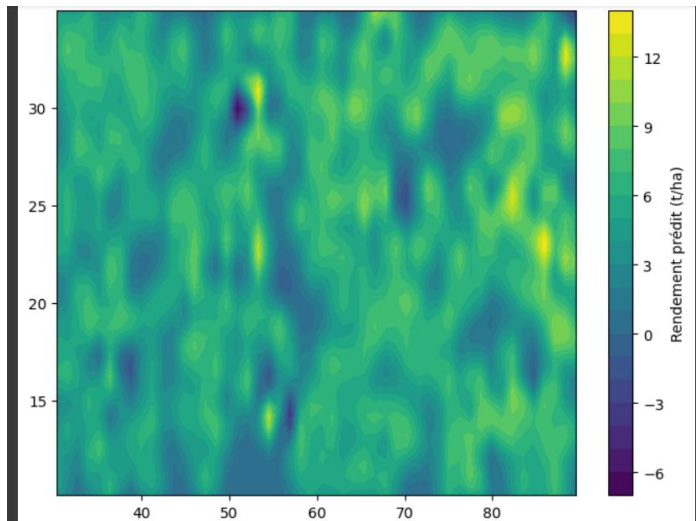
La distribution a priori représente la connaissance initiale ou les croyances sur un paramètre avant l'observation de nouvelles données. Par exemple, dans la prédiction de rendement agricole, on pourrait avoir une croyance a priori sur la température ou l'humidité qui influence la production. La distribution a posteriori, quant à elle, est obtenue en ajustant les croyances a priori avec les nouvelles données observées, ce qui permet de mettre à jour les prédictions de manière plus précise.

11. Régression bayésienne à noyau :

La régression bayésienne à noyau est une approche où un modèle de régression est appris en utilisant un noyau, comme le noyau RBF, qui permet de capturer des relations complexes et non linéaires dans les données. Le modèle est basé sur un processus gaussien qui prédit non seulement la valeur de la fonction objective, mais aussi l'incertitude associée à chaque prédiction. Cela permet de déterminer les intervalles de confiance autour des prédictions, ce qui est essentiel pour comprendre l'incertitude dans les modèles prédictifs.

12. Classification bayésienne à noyau :

La classification bayésienne à noyau est réalisée en utilisant des modèles tels que le SVM avec un noyau RBF. Le noyau RBF est capable de transformer les données non linéaires dans un espace de dimension supérieure où elles deviennent linéairement séparables. En comparaison avec un SVM classique (sans noyau), l'utilisation d'un noyau RBF améliore souvent la performance du modèle sur des ensembles de données complexes où les frontières entre les classes ne sont pas linéaires.



13. Analyse de l'incertitude dans les prédictions :

L'incertitude dans les prédictions peut être analysée en examinant les intervalles de confiance autour des prédictions. Dans le cas de la régression bayésienne à noyau, les intervalles de confiance sont calculés en fonction de l'écart-type des prédictions, ce qui permet de quantifier l'incertitude du modèle. Cela peut aider à identifier les zones où le modèle est moins sûr de ses prédictions et à orienter de futures explorations dans des zones de plus grande incertitude.

```
Précision du modèle bayésien à noyau : 0.34
Précision du SVM classique : 0.31
```

14. Test des noyaux linéaire, RBF et polynomial :

Les noyaux sont des fonctions utilisées pour transformer les données dans un espace de dimension supérieure. Le noyau linéaire est simple et rapide, mais il n'est pas efficace lorsque les données sont non linéaires. Le noyau RBF est plus flexible et est souvent utilisé pour modéliser des relations non linéaires complexes. Le noyau polynomial peut également être utilisé pour capturer des relations non linéaires, mais il peut être plus sensible aux valeurs extrêmes et plus coûteux à calculer que le noyau RBF. Le choix du noyau dépend des caractéristiques des données et du compromis entre complexité et performance.

15. Influence des noyaux et de la distribution à priori :

Le choix du noyau et de la distribution a priori a un impact considérable sur la flexibilité et la performance du modèle. Un noyau bien choisi permet de mieux capturer la structure sous-jacente des données, ce qui peut améliorer les performances du modèle. De même, la distribution a priori influence la manière dont le modèle apprend et ajuste ses

prédictions en fonction des nouvelles données. Les choix incorrects peuvent entraîner des sur-ajustements ou une mauvaise généralisation aux nouvelles données.