



I N S E A





3

8

Bayesian machine learning

Definition

Win

TO

Indépendance:

Les paramètres optimisés $\theta \in \mathbb{R}^p$ en ML décrivent des variables aléatoires

suivant une loi a priori π dont la variance est donnée par \mathcal{O} .

→ ~~no~~ a priori, toutes les valeurs possibles \Rightarrow risque d'overfitting

 \rightarrow a priori très strict, seule une valeur est possible \Rightarrow risque d'underfitting

Discuter de l'effet de C très grand / très petit.

La loi a priori est équivalente à la pénalité de régularisation: elle contrôle la complexité du modèle

Différence de
méthodologie:

Au lieu d'optimiser, on simule suivant la loi a posteriori $\theta|\text{data}$

Résultats: Au lieu d'avoir une seule prédiction on a une *distribution* de prédictions: moyenne, écart-type, HDI ..

Idée principale: Les paramètres optimisés $\theta \in \mathbb{R}^p$ en ML deviennent des variables aléatoires suivant une loi a priori π dont la variance est donnée par C .

Discuter de l'effet de C très grand / très petit.

$C \rightarrow \infty$: aucun a priori, toutes les valeurs sont possibles: \Rightarrow risque d'overfitting

$C \rightarrow 0$: a priori très strict, seule une valeur est possible \Rightarrow risque d'underfitting

La loi a priori est équivalente à la pénalité de régularisation: elle contrôle la complexité du modèle

Différence de méthodologie: Au lieu d'optimiser, on simule suivant la loi a posteriori $\theta|\text{data}$

Résultats: Au lieu d'avoir une seule prédiction on a une **distribution** de prédictions: moyenne, écart-type, HDI ...

Voir TD6

Frequentist machine learning

