INSEA

23

Regularization

Machine learning classique: zero-to-hero

Pour réduire l’overfitting, on peut réduire l’espace d’optimisation en privilégiant des coefficients simples:

$$\min_{\substack{\beta \in \mathbb{R}^{d+1} \\ \|\beta\|_2^2 \leq C}} -\sum_{i=1}^{n} y_i \log(p_i) + (1 - y_i) \log(1 - p_i)$$

Ce problème n’est pas facile à résoudre (contrainte quadratique), on peut montrer que ce problème est équivalent:

$$\min_{\beta \in \mathbb{R}^{d+1}} -\sum_{i=1}^{n} y_i \log(p_i) + (1 - y_i) \log(1 - p_i) + \frac{1}{C} \|\beta\|_2^2$$

$$C \to 0\ ?$$

l'optimisation est sur $\mathbb{R}^{d+1}$ en entier: risque d'overfitting.

le β optimal est le vecteur nul: la fonction de prédiction est constante: underfitting

$$C \to +\infty \, ?$$

Dans les deux cas plus C est petit plus on minimise $\|\beta\|$: “Plus on régularise”.

Pour réduire l'overfitting, on peut réduire l'espace d'optimisation en privilégiant des coefficients simples:

$$\min_{\substack{\beta \in \mathbb{R}^{d+1} \\ \|\beta\|_2^2 \leq C}} - \sum_{i=1}^{n} y_i \log(p_i) + (1 - y_i) \log(1 - p_i)$$

Ce problème n'est pas facile à résoudre (contrainte quadratique), on peut montrer que ce problème est équivalent:

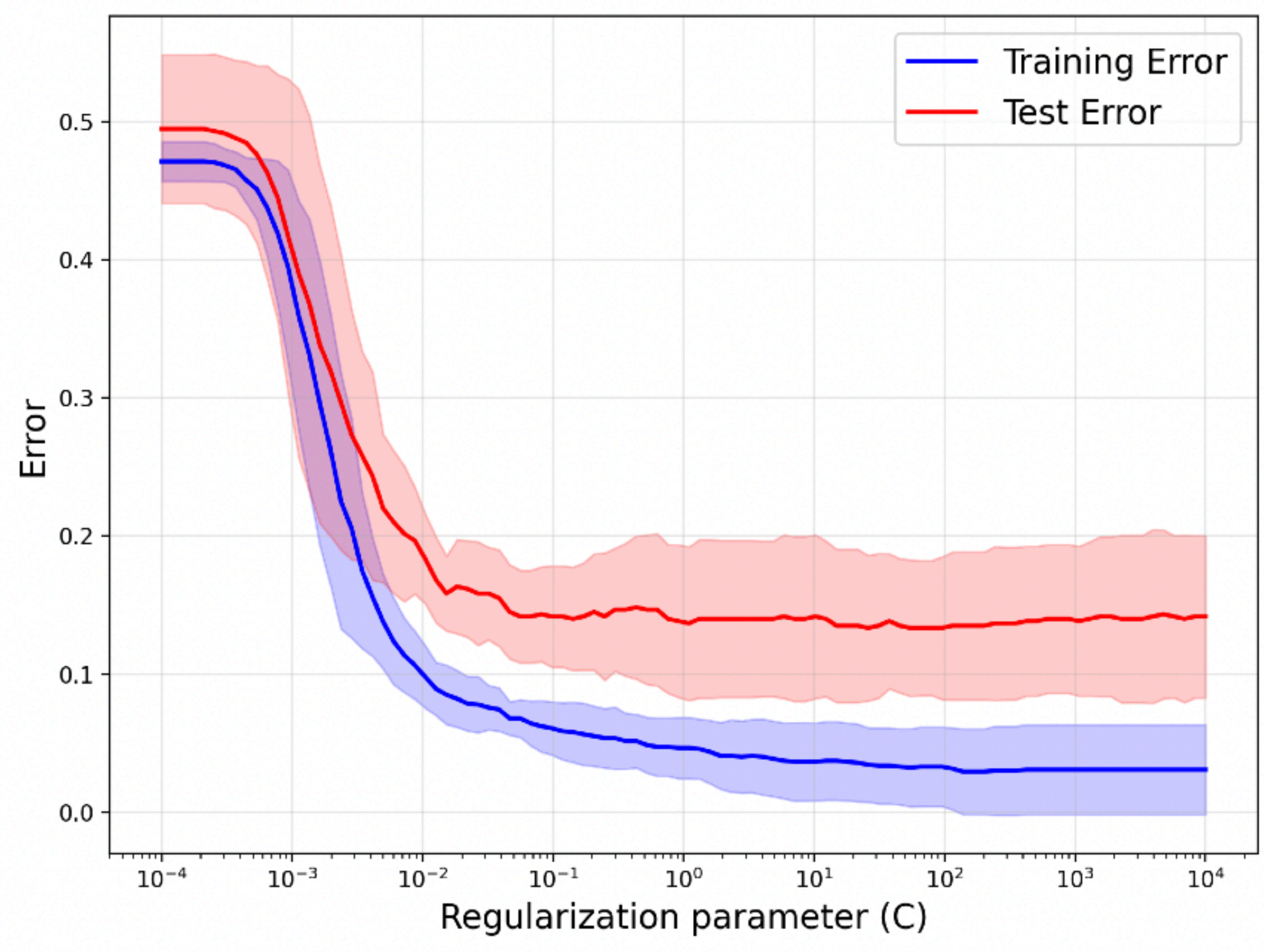
$$\min_{\beta \in \mathbb{R}^{d+1}} - \sum_{i=1}^{n} y_i \log(p_i) + (1 - y_i) \log(1 - p_i) + \frac{1}{C} \|\beta\|_2^2$$

Dans les deux cas plus C est petit plus on minimise $\|\beta\|$: "Plus on régularise".

$C \to 0$? le β optimal est le vecteur nul: la fonction de prédiction est constante: underfitting

$C \to +\infty$? l'optimisation est sur $\mathbb{R}^{d+1}$ en entier: risque d'overfitting.

INSEA


Training Error
Test Error
Error
0.0
0.1
0.2
0.3
0.4
0.5
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
Regularization parameter (C)

INSEA