Tout savoir sur la régression pénalisée

Partie 3

Présenté par **Morgan Gautherot**

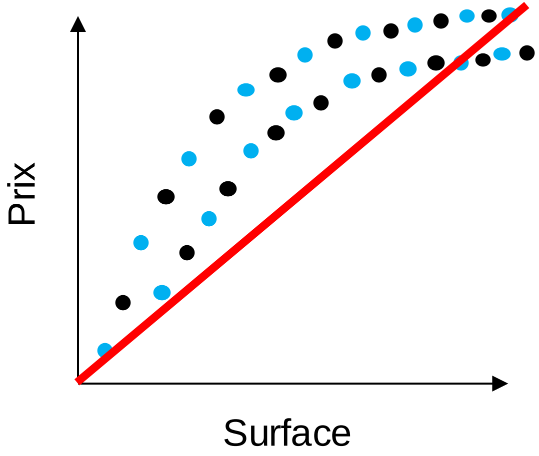
A. Pénalisation de la fonction de coût

Sur-entraînement et sous-entraînement

Sous-entraînement

Erreur sur le jeu d'entraînement : Élevée

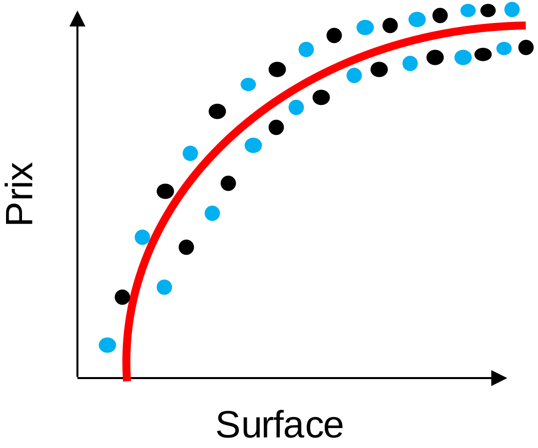
Erreur sur le jeu de test : Élevée



Entraînement correct

Erreur sur le jeu d'entraînement : Faible

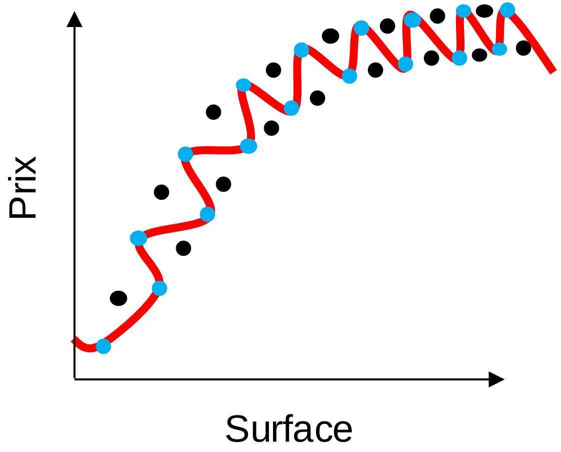
Erreur sur le jeu de test : Faible



Sur-entraînement

Erreur sur le jeu d'entraînement : Nulle

Erreur sur le jeu de test : Moyenne



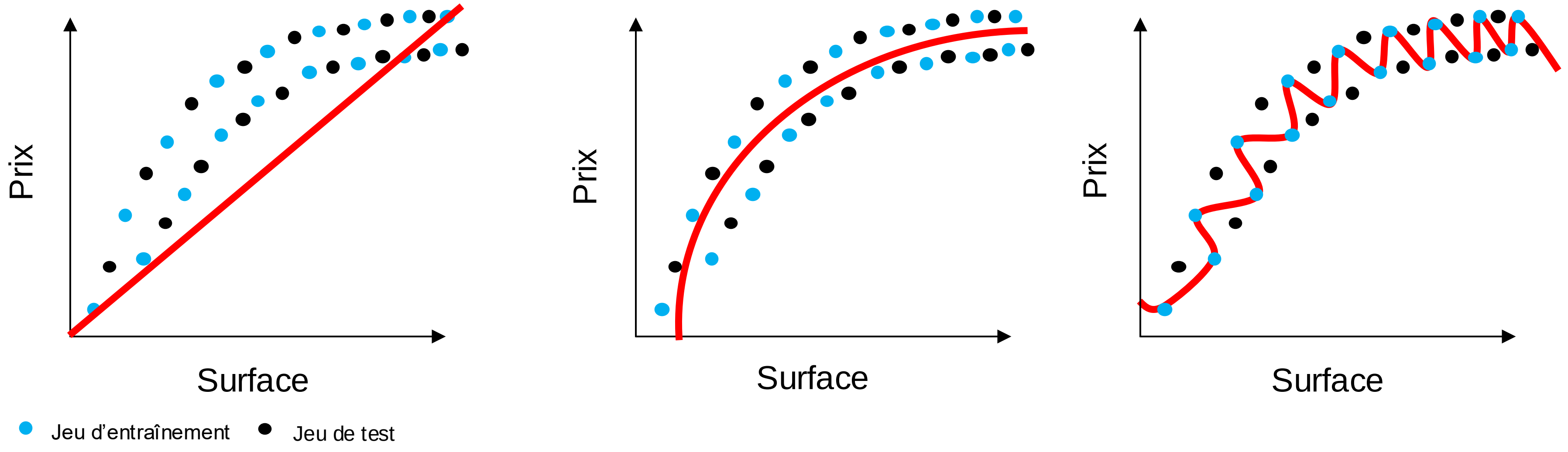
● Jeu d'entraînement ● Jeu de test

Complexité du modèle

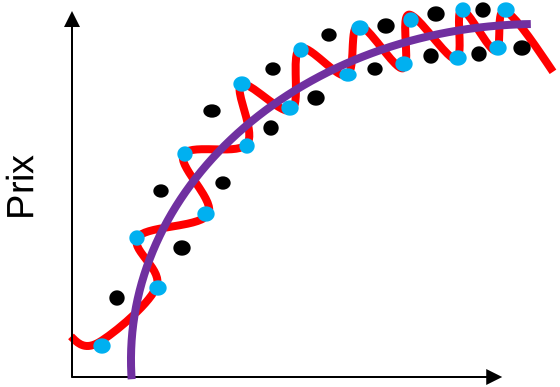
$$\hat{y} = w_0 + w_1.x_1$$

$$\hat{y} = w_0 + w_1.x_1 + w_2.{x_1}^2$$

$$\hat{y} = w_0 + w_1.x_1 + w_2.{x_1}^2 + w_3.{x_1}^3 + w_4.{x_1}^4 + \cdots$$



Pénalisation des paramètres



$$\hat{y} = w_0 + w_1.x_1 + w_2.x_1{}^2 + w_3.x_1{}^3 + w_4.x_1{}^4 + \cdots$$

Pénalisation des paramètres

$$\min_w J(w) = \underbrace{\frac{1}{2m}\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2}_{\text{Minimiser l'erreur de prédiction}} \underbrace{+1000.w_3 + 1000.w_4 + \cdots}_{\text{Minimiser la valeur des paramètres } w_3, w_4, \ldots}$$

B. La régression Lasso

Régression Lasso ou pénalisation L1

- Un modèle plus simple avec moins de paramètres est moins sujet au sur-entraînement.

Paramètre de régularisation

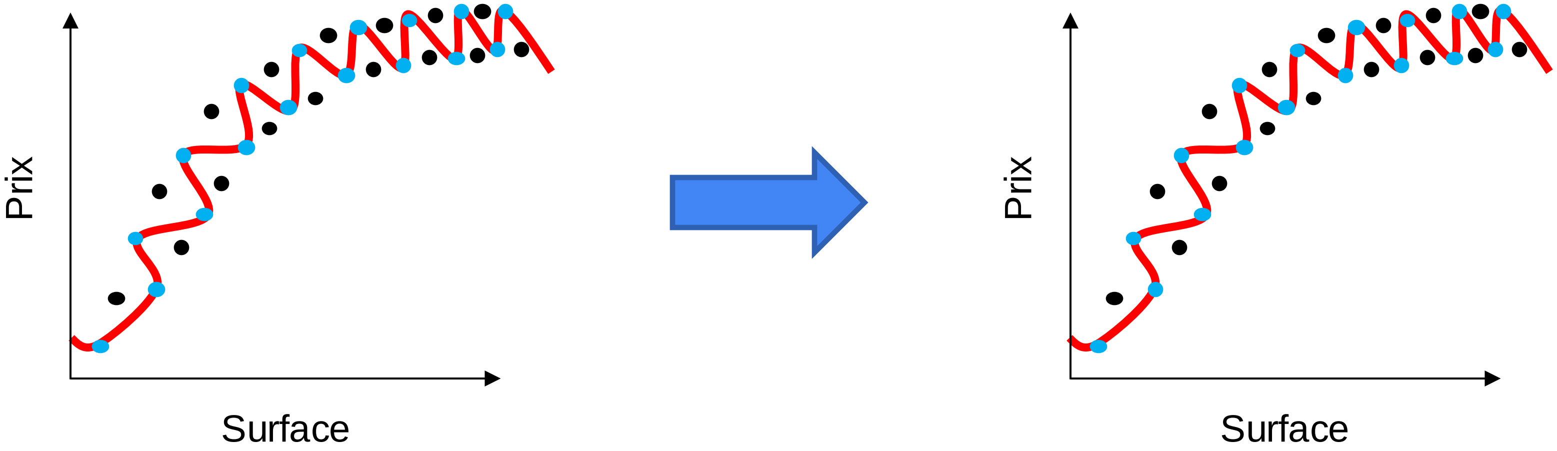
$$J(w) = \frac{1}{2m}\left[\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2 \boxed{+\lambda\sum_{j=1}^{n}|w_J|}\right]$$

Régularisation

Impact du coefficient de régularisation

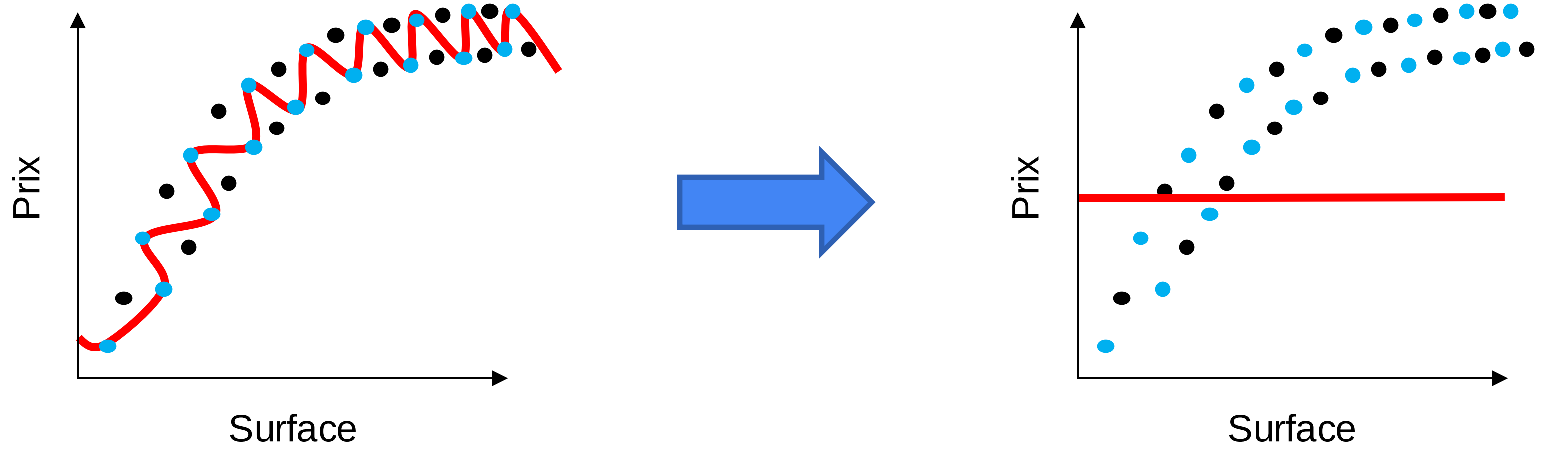
λ trop petit

$$J(w) = \frac{1}{2m}\left[\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2 \cancel{+ \lambda \sum_{j=1}^{n} |w_j|}\right]$$



Impact du coefficient de régularisation

λ trop grand $\quad J(w) = \frac{1}{2m}\left[\cancel{\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2} + \lambda \sum_{j=1}^{n} |w_J|\right]$

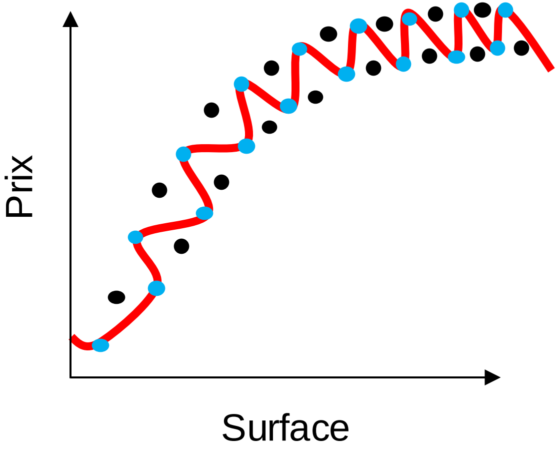


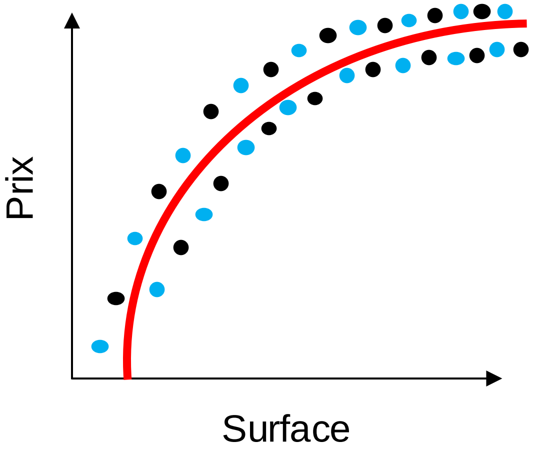
Impact du coefficient de régularisation

[0.01, …, 0.1, …., 0.5]

λ adéquat

$$J(w) = \frac{1}{2m}\left[\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2 + \lambda\sum_{j=1}^{n}|w_J|\right]$$





C. La régression Ridge

Régression Ridge ou pénalisation L2

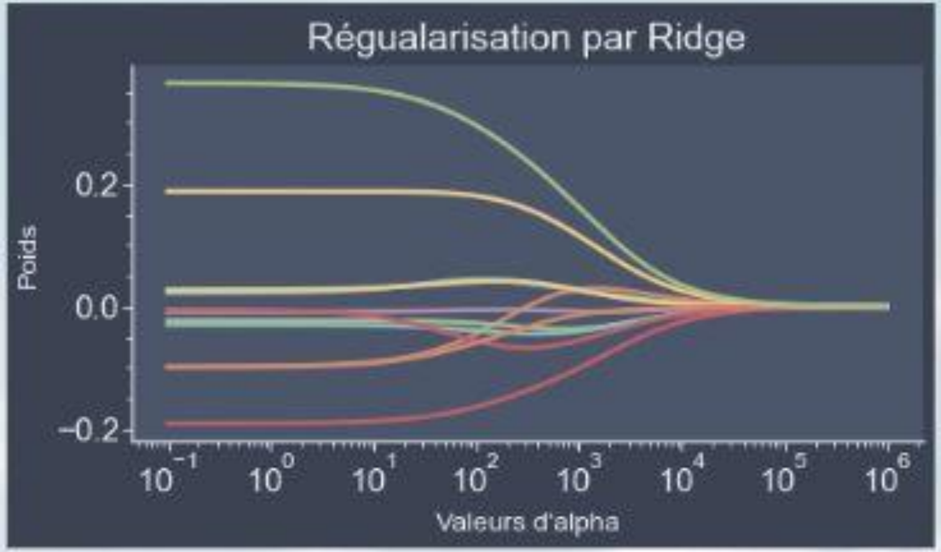
- Un modèle avec des paramètres plus homogène est moins sujet au sur-entraînement.

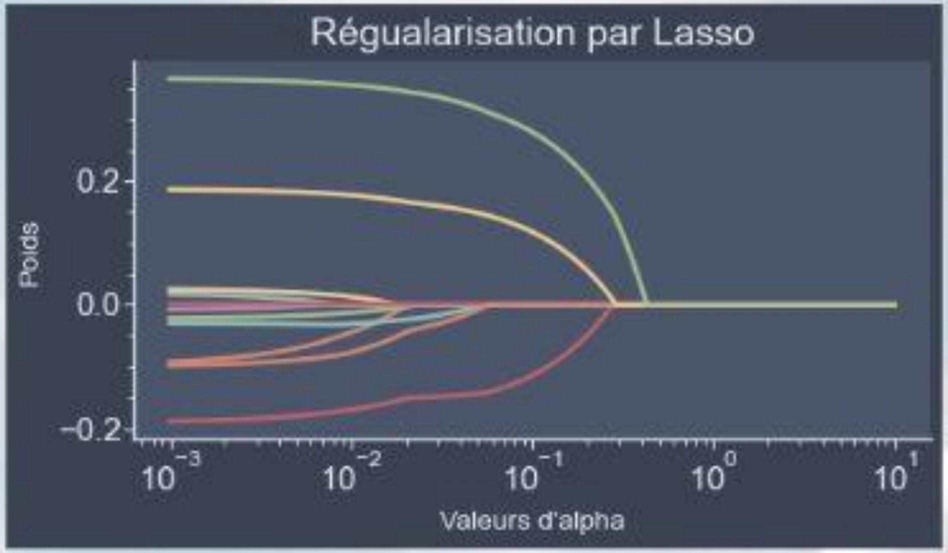
Paramètre de régularisation

$$J(w) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)^2 + \boxed{\lambda \sum_{j=1}^{n} w_j^2}\right]$$

Régularisation

Lasso vs Ridge





D. La régression Elasticnet

Régression Elasticnet ou régularisation L1 et L2

Paramètre de régularisation

$$J(w) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right)^2 \boxed{+\lambda_1 \sum_{j=1}^{n} |W_J| + \lambda_2 \sum_{j=1}^{n} w_j^2}\right]$$

Paramètre de régularisation

$$J(w) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right)^2 \boxed{+\lambda\left[\alpha \sum_{j=1}^{n} |W_J| + \frac{1-\alpha}{2} \sum_{j=1}^{n} w_j^2\right]}\right]$$

Mix entre la régularisation L1 et L2

Régression Elasticnet ou régularisation L1 et L2

$\alpha = 0$

$$J(w) = \frac{1}{2m}\left[\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})^2 + \lambda\left[\cancel{\alpha\sum_{j=1}^{n}|W_j|} + \frac{1-\alpha}{2}\sum_{j=1}^{n}w_j^2\right]\right]$$

L2

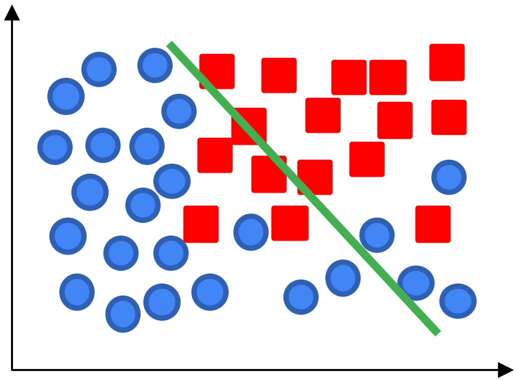
$\alpha = 1$

$$J(w) = \frac{1}{2m}\left[\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})^2 + \lambda\left[\alpha\sum_{j=1}^{n}|W_j| + \cancel{\frac{1-\alpha}{2}\sum_{j=1}^{n}w_j^2}\right]\right]$$

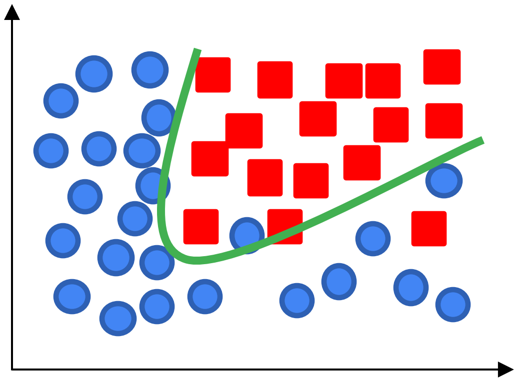
L1

E. Pour la régression logistique

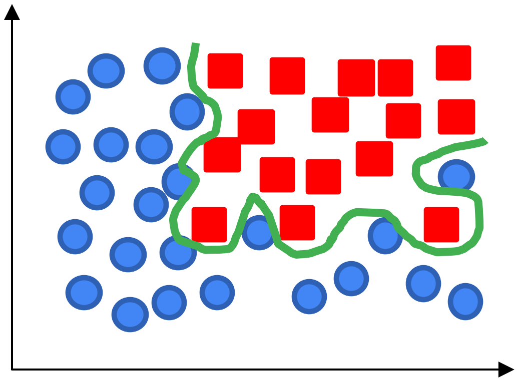
Pour la classification



Sous-entraînement



Entraînement correct



Sur-entraînement

Régression logistique pénalisée

Ridge

$$J(w) = -\left[\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log\left(h\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \log\left(1 - h\left(x^{(i)}\right)\right)\right] + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2$$

Lasso

$$J(w) = -\left[\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log\left(h\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \log\left(1 - h\left(x^{(i)}\right)\right)\right] + \frac{\lambda}{2m}\sum_{j=1}^{n} |W_J|$$

Elasticnet

$$J(w) = -\left[\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log\left(h\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right) \log\left(1 - h\left(x^{(i)}\right)\right)\right] + \lambda\left[\alpha\sum_{j=1}^{n} |W_J| + \frac{1-\alpha}{2}\sum_{j=1}^{n} w_j^2\right]$$

Résumé

- Lasso/pénalisation L1
  - Sélection des variables les plus importantes
- Ridge/pénalisation L2
  - Homogénéisation des paramètres
- Elasticnet/pénalisation L1 & L2
  - Sélection des variables et homogénéisation des paramètres.