Correction exercices model selection

Exercice n° 1 :

1. La validation croisée k-Fold consiste à découper un jeu de données en une partition en k éléments de tailles ≈ égales.

| Fold 1 | Fold 2 | ... | Fold k | } Dataset

Ensuite, pour tout fold, il est possible d'entraîner un modèle sur son complémentaire, puis d'évaluer les performances sur ce dernier.

Répéter cette procédure sur les k folds donne une distribution de scores de test.

On peut ensuite regarder des statistiques sur cette distribution (moyenne, médiane, ...) pour décider si le modèle est bon ou non.

2. Choisir λ en se basant sur l'erreur d'entraînement n'est pas une bonne idée car λ est là pour éviter l'overfitting. L'erreur d'entraînement biaise λ vers des petites valeurs.

3. $k = 1$: Un seul fold, il n'est pas possible d'avoir un jeu de validation et tout se base sur l'erreur d'entraînement.

$k = n$: Tous les folds sont de taille 1

4. Cas extrêmes :

$k = 1 \Rightarrow$ Très optimiste car erreur d'entraînement uniquement. Fort biais mais faible variance.

$k = n \Rightarrow$ Jeu de validation de taille 1 : forte variance.

Exercice n° 2 :

Moindres carrés : $\beta^* \in \underset{\beta}{\operatorname{argmin}} \underbrace{\|Y - X\beta\|_2^2}_{*}$.

1. Soit $\beta_0 \in \operatorname{argmin} *$,

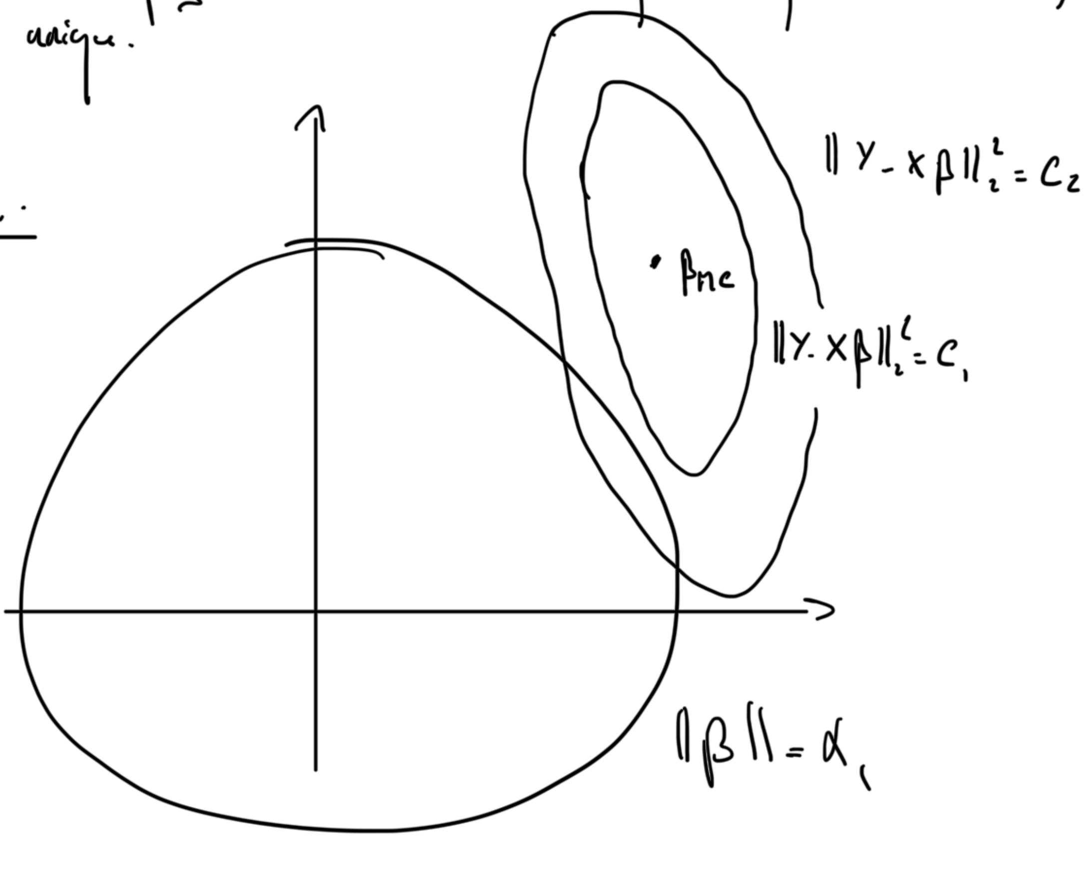
Alors $S = \{ \beta_0 + \Delta\beta \; ; \; \Delta\beta \in \operatorname{Ker}(X) \}$ est l'ensemble des solutions du problème des moindres carrés.

Or, $\dim(\operatorname{Ker}(X)) + \underbrace{\dim(\operatorname{Im}(X))}_{\leq n} = p \implies \dim(\operatorname{Ker}(X)) \geq p - n \gg 1$.

2. $$\beta_\lambda \in \underset{\beta}{\operatorname{argmin}} \; \underbrace{\|Y - X\beta\|_2^2}_{\text{convexe}} + \underbrace{\lambda \|\beta\|_2^2}_{\text{fortement convexe}}.$$

3. Donc β_λ est le minimum d'une fonction fortement convexe, et est donc unique.

4.



Exercice n°3:

1. $$\hat{\beta} \in \min_{\beta} \underbrace{\|Y - X\beta\|_2^2 + \lambda \|\beta\|_1}_{*}$$

$\|Y - X\beta\|_2^2 = \|Y\|_2^2 - 2\beta^T v + \|\beta\|_2^2$ avec $v = X^T Y$.

Donc $\min_\beta *$ est équivalent à $\min_\beta \|\beta - v\|_2^2 + \lambda\|\beta\|_1$

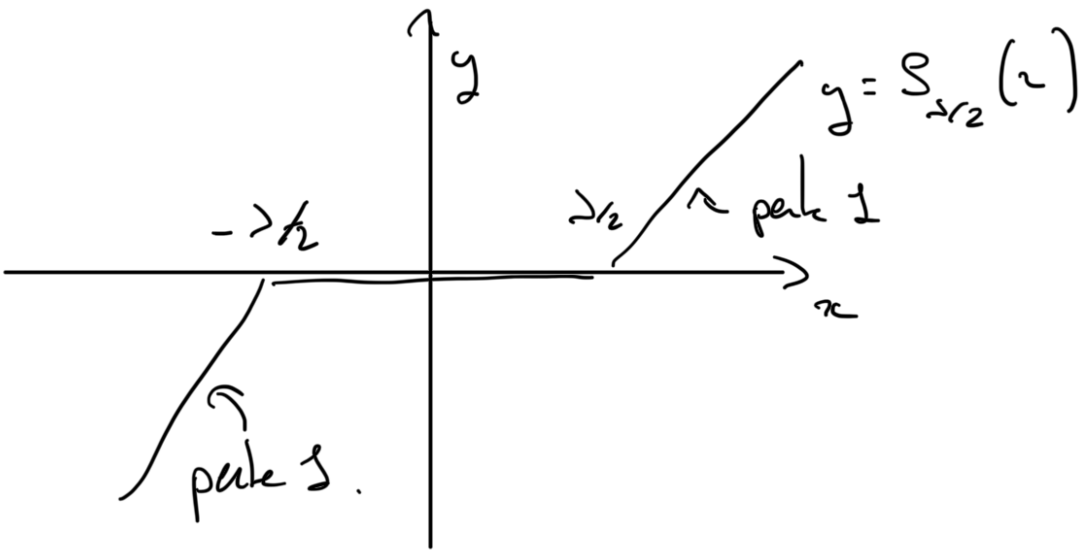
Le problème est séparable en $\beta_1, \dots, \beta_p$ avec $\forall j$,

$$\hat\beta_j \in \arg\min_b \underbrace{(b - v_j)^2 + \lambda|b|}_{f(b)} .$$

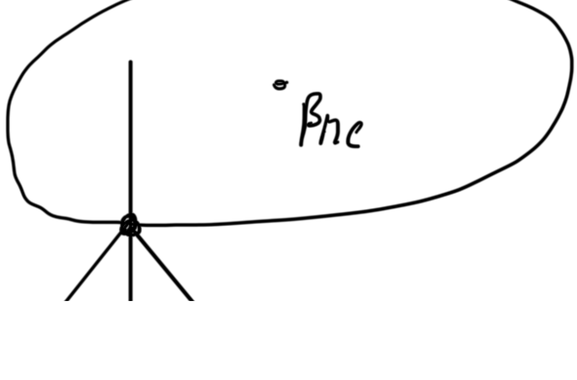
On cherche $0 \in \partial f(\hat\beta_j)$.

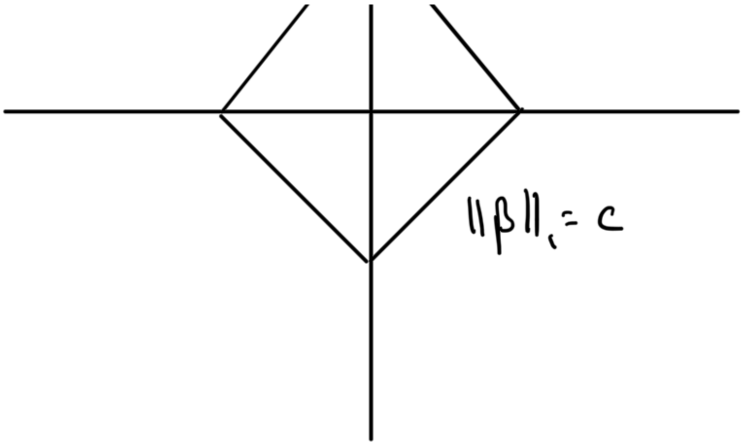
- $\hat\beta_j > 0$: $2(b - v_j) + \lambda = 0 \implies \hat\beta_j = v_j - \lambda/2$
- $\hat\beta_j < 0$: $\hat\beta_j = v_j + \frac{\lambda}{2}$
- $\hat\beta_j = 0$: $0 \in 2(0 - v_j) + \lambda[-1, 1] \iff |v_j| \leq \lambda/2$.

Donc $\hat\beta_j = S_{\lambda/2}(v_j)$ avec



2.





3. Le lasso effectue une selection de modèle car il selectionne un ensemble de variables qui expliquent bien le Y et met à 0 les autres.