



# École Polytechnique de Montréal

## Département de génie informatique et génie logiciel

---

### Rapport Assignment 1

#### MACHINE LEARNING

---

Antoine Pichon, 2489005

## Table des matières

<b>1</b>	<b>Part 1 : Linear and Weighted Linear Regression</b>	<b>2</b>
<b>2</b>	<b>Part 2 : Cross-Validated Model Selection</b>	<b>3</b>
<b>3</b>	<b>Part 3 : Gradient Descent for Ridge Regression with Learning Rate Schedules</b>	<b>3</b>
<b>4</b>	<b>Références</b>	<b>5</b>

# 1 Part 1 : Linear and Weighted Linear Regression

Dans cette partie, nous avons implémenté une régression linéaire. Commençons par démontrer, comme il est demandé dans le sujet, que la solution qui minimise la *weighted Ridge Regression Loss Function* est donnée par la formule fermée suivante :

$$\omega^* = (X^T X + \Lambda)^{-1} X^T Y \quad (1)$$

*Démonstration.* Premièrement, on note que la fonction est la somme de deux fonctions strictement convexes sur  $R^d$  (car  $\Lambda$  est symétrique) donc elle admet un minimum global unique. Soit  $L$  la fonction dérivable à minimiser :

$$L(\omega) = \|X\omega - Y\|_2^2 + \omega^T \Lambda \omega \quad (2)$$

Dérivons cette pour  $w \in R^d$  fonction terme à terme (en partant du principe qu'on ne redémontre pas la dérivée de la norme 2) :

$$\frac{d}{d\omega} \|X\omega - Y\|_2^2 = 2X^T(X\omega - Y) \quad (3)$$

et, sachant que  $\Lambda$  est symétrique :

$$\frac{d}{d\omega} \omega^T \Lambda \omega = (\Lambda + \Lambda^T)\omega = 2\Lambda\omega \quad (4)$$

d'où on déduit de (3) et (4) que :

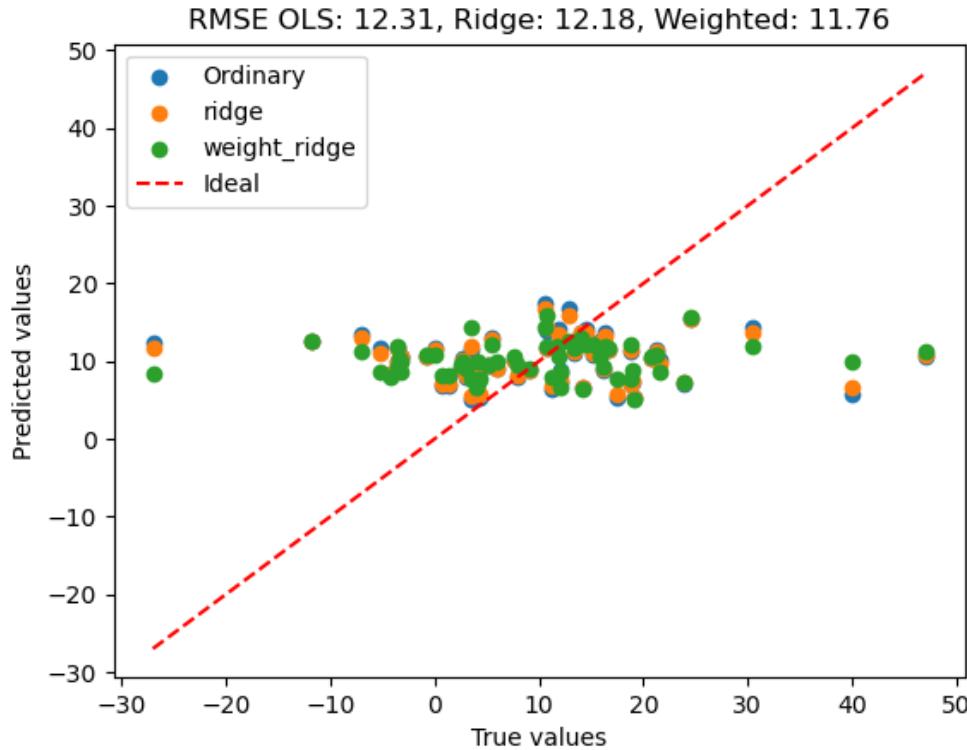
$$\frac{dL}{d\omega} = 2X^T(X\omega - Y) + 2\Lambda\omega \quad (5)$$

On cherche maintenant le minimum de  $L$  en résolvant l'équation  $\frac{dL}{d\omega} = 0$  :

$$\begin{aligned} 0 &= 2X^T(X\omega^* - Y) + 2\Lambda\omega^* \\ &\Leftrightarrow (2X^T X + 2\Lambda)\omega^* = 2X^T Y \\ &\Leftrightarrow \omega^* = (X^T X + \Lambda)^{-1} X^T Y \end{aligned}$$

ce  $\omega$  est donc l'unique minimum global de  $R^d$ . □

On entraîne ensuite notre modèle et on calcule le RMSE et on compare pour chaque régression (ordinaire, ridge regression et weighted ridge regression) les résultats obtenus sur le jeu de données de validation. On obtient les graphiques suivants :



On remarque ici que les valeurs prédites semblent loin des vraies valeurs. En effet, comme le montre le RMSE qui est de l'ordre de grandeur de 12 sachant que les valeurs des données sont plutôt entre -50 et 50, on peut en conclure que le modèle n'est pas très performant. Nous pouvons tout de même dire que le meilleur modèle selon le RMSE est le weighted ridge regression.

## 2 Part 2 : Cross-Validated Model Selection

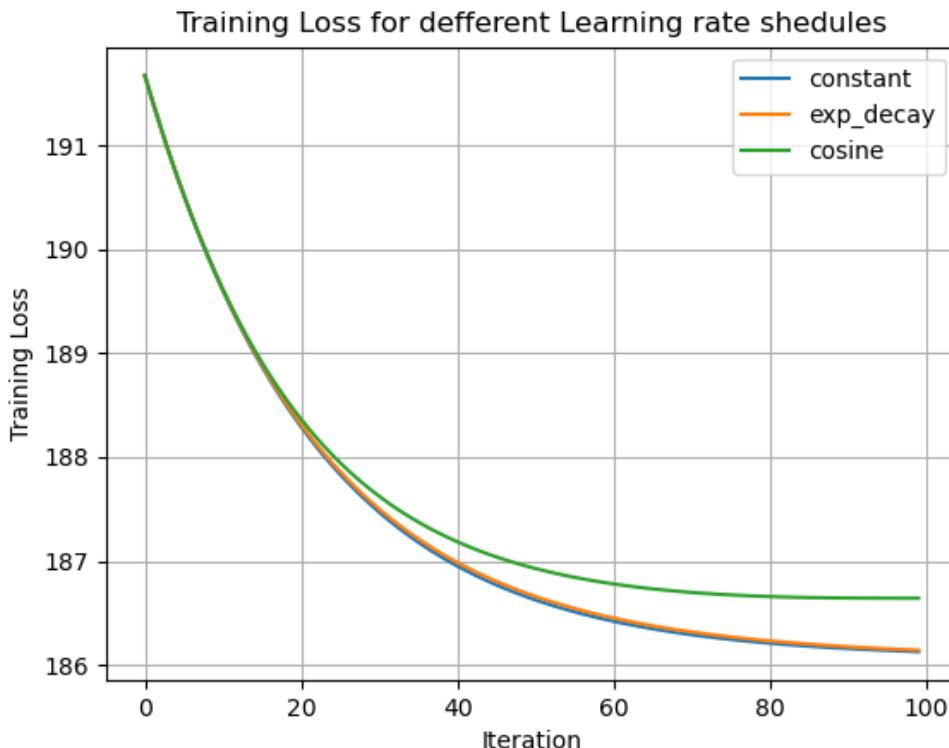
Dans cette partie, nous avons implémenté la validation croisée pour choisir le meilleur paramètre  $\lambda$  pour la ridge regression. Nous avons testé les valeurs de  $\lambda$  suivantes : [0.01, 0.1, 1, 10, 100]. On obtient les résultats suivants :

- Metric : RMSE, Best lambda : 1, Score : 9.48
- Metric : MAE, Best lambda : 1, Score : 9.47
- Metric : MaxError, Best lambda : 10, Score : 9.59

On remarque que le meilleur  $\lambda$  est 1 pour les métriques RMSE et MAE et 10 pour la métrique MaxError.

## 3 Part 3 : Gradient Descent for Ridge Regression with Learning Rate Schedules

Dans cette partie, nous avons implémenté la descente de gradient pour la ridge regression. Nous avons comparé les différentes schedules et avons tracé les loss en fonction du nombre d'itérations. On obtient les graphiques suivants :



On remarque que la schedule Exponential decay et constant sont les plus performantes mais ce graphique ne nous permet pas de comparer clairement ces deux modèles. En effet, ces deux modèles ont une vitesse de convergence plus rapide et leurs loss finaux sont plus bas que pour l'autre modèle. On peut ensuite essayer de comparer les modèles grâce au RMSE que nous renvoie le terminal :

- RMSE for constant : 14.835124597042826
- RMSE for expdecay : 14.837655857740783
- RMSE for cosine : 14.886059843854753

On remarque encore que les modèles constant et expdecay sont les plus performants et ont des RMSE très proches. Nos expériences ne nous permettent pas de comparer ces deux modèles de manière claire.

## 4 Références

- Le template L<sup>A</sup>T<sub>E</sub>X a été adapté de celui donné pour le cours INF8085 Cybersécurité.
- L'utilisation d'une IA générative pour la correction grammaticale et orthographique a été faite comme cela est autorisé dans le sujet.