UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Departamento de Matemática

Tarea 4 de Optimización No Lineal (MAT-279)
Segundo semestre de 2022

En aplicaciones relacionadas con *machine learning*, se dispone de una serie de datos $\{(w_i, y_i) : i \in I\}$ con los cuales nos interesa ajustar un modelo para explicar la variable Y en términos de W, usualmente de manera lineal (i.e. $Y \sim X^\top W$). En general, estos problemas se pueden escribir de la forma

$$\min_{x\in\mathbb{R}^d} F(x) := \sum_{i\in I} f_i(x) + \lambda\mathcal{R}(x),$$

donde f_i representa el costo de utilizar el i-ésimo dato, y $\mathcal{R}$ es una función de regularización para mantener la simpleza del modelo. En este problema, se piden estudiar tres modelos

Mínimos cuadrados $f_i(x) = (y_i - x^\top w_i)^2$, y $\mathcal{R} \equiv 0$.

Ridge $f_i(x) = (y_i - x^\top w_i)^2$, y $\mathcal{R}(x) = \|\cdot\|_2$.

Se pide

1. Justifique la convexidad de estos problemas.
2. Implemente el **Método del gradiente** (en el lenguaje de su gusto, de preferencia Python) descrito en clases (Apunte, pág.40). Su rutina debe considerar como entradas X, W, λ y debe entregar el vector de coeficientes óptimos x^*. El pseudo-código se escribe de la siguiente forma Considere

Algorithm 1 Método del Gradiente con paso de Armijo

1: **procedure** GRAD($Y, W, \lambda, \epsilon, \alpha, \beta$)
2: Defina un punto inicial x^0
3: **for** $\nu = 0, \ldots$ **do**
4: Calcule $g^\nu = \nabla F(x^\nu)$
5: **while** $\|g^\nu\| \geq \epsilon$ **do**
6: Defina $d^\nu = -g^\nu$
7: Calcule

$$\lambda_\nu = \max_{k=0,1,\ldots} \left\{\beta^k \;\middle|\; F(x^\nu + \beta^k d^\nu) - F(x^\nu) \leq \alpha g^\nu \cdot d^\nu \beta^k\right\}$$

8: Defina $x^{\nu+1} \leftarrow x^\nu + \lambda_\nu d^\nu$ y vuelva al paso (4)
9: **end while**
10: **end for**
11: **end procedure**

$\alpha, \beta \approx 0,8$. Considere los datos en `data.csv`, correspondiente a las variaciones de una superficie y sometida a una sistema de fuerzas que la deforman, y una colección de variables que representan los materiales presentes en esta superficie $x_1, \ldots, x_{15}$

3. Estime los coeficientes para los tres modelos con $\lambda = 1$. Reporte las primeras 10 iteraciones de su método y compare las soluciones correspondientes.
4. Finalmente, estime regresiones Ridge con parámetro $\lambda \in \{0{,}01; 0{,}1; 1; 10; 1000\}$. Comente en las soluciones obtenidas y realice un gráfico que muestre la diferencia en magnitud de los coeficientes x.