Curso: Métodos basados en Kernels para aprendizaje automático
Profesora: Tamara Fernández

Tarea 1

Instrucciones de entrega:

- **Entregue un informe en formato pdf** con sus resultados, pseudo codigo (si corresponde) y conclusiones. Además incluya su codigo en R.
- **Fecha de entrega: 31 de agosto 2023.**

Ejercicio 1. (20 puntos)

1. Cree un vector de 100 elementos de manera que los primeros 20 elementos sean $1, 2, \ldots, 20$, los siguientes 10 elementos sean $10, 20, 30, \ldots, 100$, y los últimos 70 elementos sean $31, 32, \ldots, 100$.

   Hint: Utilice la función `seq()`. Para obtener más información sobre esta función, en la ventana de la consola, escribe `?seq`.

2. Utilice R para calcular:

   2.1) $\begin{pmatrix} 2 \\ 5 \\ -6 \\ 1 \end{pmatrix}^{\intercal} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$

   2.2) $\left(\begin{pmatrix} -1 \\ 3 \\ -6 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ -2 \\ 3 \\ -4 \end{pmatrix} \right)^{\intercal} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 2 \end{pmatrix}$

   2.3) $\begin{pmatrix} 0 \\ -1 \\ 2 \\ 0 \end{pmatrix}^{\intercal} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 8 & 7 & 6 & 5 \\ -4 & -3 & -2 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 2 \end{pmatrix}$

   para responder **cree los vectores y matrices involucrados en R**.

3. Escriba una función en R que, utilizando ciclos `for`, calcule la norma de un vector $x \in \mathbb{R}^n$ de largo n arbitrario dada por $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$. Utilice su función en 5 vectores cualesquiera que cumplan con tener dimensión $n = 50, 500, 1000, 15000$ y 200000 respectivamente.

4. Repita el ejercicio anterior sin utilizar ciclos `for` o ningún tipo de ciclo iterativo como `while`.

5. Escriba una función en R que reciba como argumento un vector $x \in \mathbb{R}^n$ de largo n arbitrario y retorne

$$\texttt{output} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}),$$

   donde $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ representa el promedio de x.

5. Escriba una función que reciba como argumento la matriz

$$\begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{pmatrix} \quad \text{y retorne} \quad \begin{pmatrix} a! & b! & c! & d! \\ e! & f! & g! & h! \\ i! & j! & k! & l! \\ m! & n! & o! & p! \end{pmatrix}$$

   donde $a, b \ldots, p$ son números enteros positivos. Su función **no debe usar la función** `factorial`.

6. Considere una matriz $\boldsymbol{M}$ de dimensiones $n \times m$ donde $\boldsymbol{M}_{i,j} = (i+j)(i-j)$ para cualquier $i \in \{1, \ldots, n\}$ y $j \in \{1, \ldots, m\}$. Además, defina

$$a = \frac{1}{n} \sum_{\ell=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{n} \boldsymbol{y}_j \boldsymbol{x}_\ell \boldsymbol{x}_k \boldsymbol{M}_{\ell,j}.$$

Encuentre una expresión para a en terminos de operaciones vectoriales entre los vectores columna $\boldsymbol{x}$, $\boldsymbol{y}$ y la matriz $\boldsymbol{M}$. Implemente en R un algoritmo que use su solución y pruébelo en un ejemplo.

Ejercicio 2. (10 puntos) En este ejercicio introducimos la función de kernel llamada **Rational Quadratic**, definida por:

$$K_{RQ}(\boldsymbol{x}, \boldsymbol{y}) = \sigma^2 \left(1 + \frac{\|\boldsymbol{x} - \boldsymbol{y}\|}{2\alpha\ell^2}\right)^{-\alpha}, \quad \text{donde} \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2 \quad \text{y} \quad \sigma, \ell, \alpha > 0$$

1. Escriba una función en R que le permita evaluar el kernel Rational Quadratic.
2. Grafique el comportamiento de la función anterior en terminos de $\|\boldsymbol{x} - \boldsymbol{y}\|$. Para esto use $\sigma = 1$ y $\ell = 1$ y $\alpha = 1$.
3. Efectúa nuevamente la tarea anterior, manteniendo constantes los valores de σ y ℓ, pero esta vez considerando distintos valores para α: $\alpha = 5$, $\alpha = 10$, y $\alpha = 50$.

Ejercicio 3. (10 puntos) Considere dos vectores $\boldsymbol{x}$ e $\boldsymbol{y}$ de dimensión n cada uno, y considere la matriz $\mathbf{A}$ de dimensión $n \times n$.

1. Escriba un método/función basado en ciclos `for` que le permita calcular $\boldsymbol{x}^\intercal \mathbf{A} \boldsymbol{y}$. El método debe recibir como parámetros de entrada $\boldsymbol{x}, \boldsymbol{y}$ y $\mathbf{A}$.
2. Investigue la función `system.time` para calcular el tiempo de ejecución para valores de n entre 100 y 100000. Grafique n vs el tiempo de ejecución, y compárelo con el tiempo de ejecución obtenido al usar `t(x)%*%A%*%y` (multiplicación nativa de R). Discuta sus resultados.

Ejercicio 4. (10 puntos) En las siguientes preguntas trabajaremos con la base de datos `starbucks` del paquete `openintro`, que reune datos nutricionales para 77 productos vendidos por Starbucks. Esta base de datos contiene la siguiente información:

- `item`: Nombre del producto vendido.
- `calories`: Número de calorias.
- `fat`: Grasas (gramos).
- `carb`: Carbohidratos (gramos).
- `fiber`: Fibra (gramos).
- `protein`: Proteinas (gramos).
- `type`: Tipo de producto: `bakery, bistro box, hot breakfast, parfait, petite, salad,` y `sandwich`

Utilizando los datos anteriores, cree en R una matriz que contenga la información de las calorias, grasas, carbohidratos, fibra y proteinas. Llame a esta matriz `Z`.

a) Usando los datos en Z obtenga la matriz de Gram usando el kernel lineal denotado por

$$K(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + c$$

donde $c = 1$.

b) Usando los datos en Z obtenga la matriz de Gram usando el kernel Gaussiano denotado por

$$K(\boldsymbol{x}, \boldsymbol{y}) = \sigma^2 \exp\left\{-\frac{1}{\ell^2}(\boldsymbol{x}-\boldsymbol{y})^\intercal(\boldsymbol{x}-\boldsymbol{y})\right\}.$$

Use $\sigma = 2, \ell = 0.5$.

Ejercicio 5. (10 puntos)

1. Muestre que si la matriz $\mathbf{X}$ es de rango completo, entonces $\mathbf{X}^\intercal\mathbf{X}$ también tiene rango completo.

2. Sea $\mathbf{A}$ una matriz de $d \times d$ cuyas entradas dependen en un escalar θ. La derivada de $\mathbf{A}$ respecto a θ se define mediante

$$\frac{\partial}{\partial\theta}\mathbf{A} = \begin{pmatrix} \frac{\partial}{\partial\theta}\mathbf{A}_{11} & \frac{\partial}{\partial\theta}\mathbf{A}_{12} & \dots & \frac{\partial}{\partial\theta}\mathbf{A}_{1d} \\ \frac{\partial}{\partial\theta}\mathbf{A}_{21} & \frac{\partial}{\partial\theta}\mathbf{A}_{22} & \dots & \frac{\partial}{\partial\theta}\mathbf{A}_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial\theta}\mathbf{A}_{d1} & \frac{\partial}{\partial\theta}\mathbf{A}_{d2} & \dots & \frac{\partial}{\partial\theta}\mathbf{A}_{dd} \end{pmatrix}$$

Suponga que $\mathbf{A}$ tiene inversa dada por $\mathbf{A}^{-1}$. Muestre que

$$\frac{\partial}{\partial\theta}\mathbf{A}^{-1} = -\mathbf{A}^{-1}\left(\frac{\partial}{\partial\theta}\mathbf{A}\right)\mathbf{A}^{-1}.$$

Hint 1: Para responder la primera pregunta utilice la definicón de rango completo que establece que la matriz $\mathbf{X}$ es de rango completo si se cumple lo siguiente:

$$\mathbf{X}\boldsymbol{v} = 0 \qquad \text{si y solo si} \qquad \boldsymbol{v} = 0.$$

Utilizando lo anterior, deduzca que si $\mathbf{X}$ es de rango completo, entonces $\mathbf{X}^\intercal\mathbf{X}$ también lo es.