Matemáticas para Ciencias de los Datos: Trabajo práctico 0

M. Sc. Saúl Calderón Ramírez
Instituto Tecnológico de Costa Rica,
Escuela de Computación, bachillerato en Ingeniería en Computación,
PAttern Recongition and MAchine Learning Group (PARMA-Group)

5 de septiembre de 2020

Nombre integrantes: Esteban García Solís

1. Sistemas lineales (40 puntos)

1. Demuestre si los siguientes sistemas $L\left\{\cdot\right\}$ (con entrada $u(t)$ y salida $g(t)$, y $h(t)$ una función cualquiera) son lineales o no lineales. Además, muestrelo con una implementación en Pytorch, usando como entrada un arreglo de 50 valores generados al azar. Si va a demostrar por contraejemplo, muestre las entradas y salidas de la corrida en Pytorch que demuestran el no cumplimiento de la propiedad.

   a) $g\left(t\right)=u(t)+5$

   $$\begin{aligned} L\left\{u_1(t)+u_2(t)\right\} &= u_1(t)+u_2(t)+5 \\ &\neq (u_1(t)+5)+(u_2(t)+5) \\ &= L\left\{u_1(t)\right\}+L\left\{u_2(t)\right\} \end{aligned}$$

   No se cumple la superposición por lo tanto no es lineal. (ver colab)

   b) $g\left(t\right)=u(t)\,h(t)$

   $$\begin{aligned} L\left\{\alpha u_1(t)+\beta u_2(t)\right\} &= (\alpha u_1(t)+\beta u_2(t))h(t) \\ &= \alpha h(t)u_1(t)+\beta h(t)u_2(t) \\ &= L\left\{\alpha u_1(t)\right\}+L\left\{\beta u_2(t)\right\} \end{aligned}$$

   Se cumple superposición y homogenidad por lo tanto es lineal.

c) $g(t) = \text{máx}(u(t))$

$$\begin{aligned} L\{u_1(t) + u_2(t)\} &= max(u_1(t) + u_2(t)) \\ L\{u_1(t)\} + L\{u_2(t)\} &= max(u_1(t)) + max(u_2(t)) \\ max(u_1(t) + u_2(t)) &\leq max(u_1(t)) + max(u_2(t)) \end{aligned}$$

Por desigualdad triangular no siempre se cumple la superposción por lo tanto no es lineal (ver colab)

d) $g(t) = u'(t) = \text{lím}_{h\to\infty} \frac{u(t+h)-u(t)}{h}$

$$\begin{aligned} L\{u_1(t) + u_2(t)\} &= \lim_{h\to\infty} \frac{u_1(t+h) + u_2(t+h) - (u_1(t) + u_2(t))}{h} \\ &= \lim_{h\to\infty} \frac{u_1(t+h) - u_1(t)}{h} + \lim_{h\to\infty} \frac{(u_2(t+h) - u_2(t)}{h} \\ &= L\{u_1(t)\} + L\{u_2(t)\} \end{aligned}$$

Se cumple la superposición.

$$\begin{aligned} L\{\alpha u_1(t)\} &= \lim_{h\to\infty} \frac{\alpha u_1(t+h) - \alpha u_1(t)}{h} \\ &= \alpha \lim_{h\to\infty} \frac{u_1(t+h) - u_1(t)}{h} \\ &= \alpha L\{u_1(t)\} \end{aligned}$$

Tambien se cumple la homogenidad por lo tanto es lineal

e) $g(t) = |u(t)|$, suponga un $\alpha = -1$

$$\begin{aligned} L\{\alpha u(t)\} &= |\alpha u(t)| = |-u(t)| \\ \neq \alpha L\{u(t)\} &= \alpha |u(t)| = -|u(t)| \end{aligned}$$

No se cumple la homogenidad por lo tanto no es lineal. (ver colab)

2. Demuestre si los siguientes sistemas con múltiples variables de entrada $L\{\cdot\}$ (con entrada vectorial $\vec{u}$ y salida escalar $s \in \mathbb{R}$), cumplen la condición de homogeneidad absoluta, superposición. Además, muestrelo con una implementación en Pytorch, mostrando la propiedad con 50 vectores generados al azar. Si realiza la demostración por contraejemplo, muestre las entradas y salidas de la corrida en Pytorch que demuestran el no cumplimiento de la propiedad.

   a) Norma de Manhattan ℓ_1, prueba de homogenidad absoluta y superposición (ver colab):

$$
\begin{aligned}
\left\|t\vec{x}\right\|_1 &= \sum_{i=i}^{n} |tx_i| \\
&= |t| \sum_{i=i}^{n} |x_i| \\
&= |t| \left\|t\vec{x}\right\|_1
\end{aligned}
$$

$\ell_1$cumple la homogenidad absoluta.

$$
\begin{aligned}
\|\vec{x}+\vec{y}\|_1 &= \sum_{i=i}^{n} |x_i + y_i| \\
&\leq \sum_{i=i}^{n} |x_i| + |y_i| \\
&= \sum_{i=i}^{n} |x_i| + \sum_{i=i}^{n} |y_i| \\
&= \|\vec{x}\|_1 + \|\vec{y}\|_1
\end{aligned}
$$

$\ell_1$no cumple la superpocisión.

b) Norma Euclidiana ℓ_2, prueba de homogenidad absoluta y superposición (ver colab):

$$
\begin{aligned}
\left\|t\vec{x}\right\|_2 &= \sqrt{\sum_{i=i}^{n} (tx_i)^2} \\
&= \sqrt{\sum_{i=i}^{n} t^2 x_i^2} \\
&= |t| \sqrt{\sum_{i=i}^{n} x_i^2} \\
&= |t| \left\|\vec{x}\right\|_2
\end{aligned}
$$

ℓ_2 cumple con la homogenidad absoluta.

$$
\begin{aligned}
\|\vec{x}+\vec{y}\|_2 &= \sqrt{\sum_{i=i}^{n}(x_i+y_i)^2} \\
&= \sqrt{\sum_{i=i}^{n} x_i^2 + 2x_iy_i + y_i^2} \\
&= \sqrt{\sum_{i=i}^{n} x_i^2} + \sqrt{\sum_{i=i}^{n} y_i^2} + \sqrt{\sum_{i=i}^{n} 2x_iy_i} \\
&= \sum_{i=i}^{n} |x_i| + \sum_{i=i}^{n} |y_i| + \sqrt{\sum_{i=i}^{n} 2x_iy_i} \\
&= \|\vec{x}\|_2 + \|\vec{y}\|_2 + \sqrt{\sum_{i=i}^{n} 2x_iy_i}
\end{aligned}
$$

ℓ_2 no cumple con la superposición

c) Norma ℓ_∞, prueba de homogenidad absoluta y superposición (ver colab):

$$
\begin{aligned}
\left\|t\vec{x}\right\|_\infty &= \left(\sum_{i=i}^{n} |tx_i|^\infty\right)^{\frac{1}{\infty}} \\
&= \left(|t|^\infty \sum_{i=i}^{n} |x_i|^\infty\right)^{\frac{1}{\infty}} \\
&= |t| \left(\sum_{i=i}^{n} |x_i|^\infty\right)^{\frac{1}{\infty}} \\
&= |t| \left\|\vec{x}\right\|_\infty
\end{aligned}
$$

ℓ_∞cumple con la homogenidad absoluta.

$$
\begin{aligned}
\|\vec{x}+\vec{y}\|_\infty &= \left(\sum_{i=i}^{n} |x_i+y_i|^\infty\right)^{\frac{1}{\infty}}\\
&= \sqrt[\infty]{\sum_{i=i}^{n} |x_i+y_i|^\infty}\\
&\neq \sqrt[\infty]{\sum_{i=i}^{n} |x_i|^\infty} + \sqrt[\infty]{\sum_{i=i}^{n} |y_i|^\infty}\\
&= \|\vec{x}\|_\infty + \|\vec{y}\|_\infty
\end{aligned}
$$

ℓ_∞ no cumple con la superposición

d) $g(\vec{x}) = \vec{w}\cdot\vec{x}+b = \vec{w}^T\vec{x}+b = w_1x_1+w_2x_2+\ldots+w_nx_n+b$ con $\vec{w}\in\mathbb{R}^n$ y $b\in\mathbb{R}$ coeficientes conocidos, demuestre si es un sistema lineal respecto al arreglo de entradas $\vec{x}$.

$$
\begin{aligned}
g(\vec{x}+\vec{y}) &= \vec{w}\cdot(\vec{x}+\vec{y})+b\\
&= \left(\sum_{i=i}^{n} w_i(x_i+y_i)\right)+b\\
&= \left(\sum_{i=i}^{n} w_ix_i+w_iy_i\right)+b\\
&= \sum_{i=i}^{n} w_ix_i+\sum_{i=i}^{n} w_iy_i+b\\
&\neq \left(\sum_{i=i}^{n} w_ix_i\right)+b+\left(\sum_{i=i}^{n} w_iy_i\right)+b\\
&= (\vec{w}\cdot\vec{x}+b)+(\vec{w}\cdot\vec{y}+b)
\end{aligned}
$$

No cumple la superposición por tanto el sistema no es lineal.(ver colab)

1) Explique qué modificaciones al arreglo de entradas $\vec{x}$ y de pesos $\vec{w}$ pueden realizarse para que la salida del sistema sea la misma, y sea además un sistema lineal.

   R/puede concatenarse b al vector $\vec{w}$,$\vec{w} = \begin{bmatrix} w_1\\ w_2\\ \vdots\\ w_n\\ b\end{bmatrix}$ y concatenar un

1al vector $\vec{x}$, $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ 1 \end{bmatrix}$, el sistema se modifica de la siguiente manera $g(\vec{x}) = \vec{w} \cdot \vec{x}$ pues $b$ahora es parte del vector de pesos. Como podemos ver a continuacion la salida del sistema es la misma:

$$\begin{aligned} g(\vec{x}) &= \vec{w} \cdot \vec{x} \\ &= w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + b1 \\ &= w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + b \\ &= \vec{w} \cdot \vec{x} + b \end{aligned}$$

2. Vectores (20 puntos)

1. Graficación y propiedades de los vectores

   a) Usando Python grafique los siguientes vectores $\vec{v}_1 = \begin{bmatrix} -0{,}3 \\ 0{,}8 \\ 0{,}1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 0{,}5 \\ 0{,}2 \\ 0{,}4 \end{bmatrix}$ y $\vec{v}_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \\ 0 \end{bmatrix}$. (ver colab)

   b) Demuestre cuáles de los vectores anteriores son unitarios.(ver colab)

   c) Calcule el ángulo en grados, entre los vectores $\vec{v}_1$y $\vec{v}_2$, $\vec{v}_2$ y $\vec{v}_3$ y $\vec{v}_1$ y $\vec{v}_3$, implementando la fórmula en Pytorch, sin usar las funciones correspondientes de la biblioteca. (ver colab)

   d) Calcule la distancia en ℓ_1, ℓ_2, y ℓ_∞ entre los vectores $\vec{v}_1$y $\vec{v}_2$, $\vec{v}_2$ y $\vec{v}_3$ y $\vec{v}_1$ y $\vec{v}_3$, implementando la fórmula en en Pytorch, sin usar las funciones correspondientes de la biblioteca. Compare el resultado obtenido con el uso de la función *torch.norm.*(ver colab)

2. Propiedades del producto punto: demuestre lo siguiente. Además, muestrelo con una implementación en Pytorch, usando como entrada un arreglo de 50 arreglos generados al azar, adjunte un pantallazo con la salida de la comparación del resultado a ambos lados de la igualdad, o en su defecto, demuestre el no cumplimiento de la propiedad con un contraejemplo.

   a) Conmutatividad del producto punto $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$

$$\begin{aligned}\vec{u}\cdot\vec{v} &= \sum_{i=1}^{n} u_i v_i \\ &= \sum_{i=1}^{n} v_i u_i \\ &= \vec{v}\cdot\vec{u}\end{aligned}$$

por lo tanto el producto punto cumple la conmutatividad.(ver colab)

b) No asociatividad del producto punto $\vec{u}\cdot(\vec{v}\cdot\vec{w}) \neq (\vec{u}\cdot\vec{v})\cdot\vec{w}$
R/ La expresión $\vec{u}\cdot(\vec{v}\cdot\vec{w})$ no tiene sentido pues $(\vec{v}\cdot\vec{w})$ es un escalar y la definición del producto punto no permite operandos escalar-vector, de igual manera ocurre con $(\vec{u}\cdot\vec{v})$, por lo tanto no se puede evaluar la asociatividad en terminos de la definición del producto punto.

3. (40 puntos) Funciones multivariable

1. Funciones lineales multivariable: un hiperplano definido en un espacio $\mathbb{R}^{n+1}$ se puede expresar como una función con dominio $\vec{x}\in\mathbb{R}^n$ y codominio en $\mathbb{R}$ como sigue: $z = f(\vec{x}) = \vec{x}\cdot\vec{w}$, con $\vec{w}\in\mathbb{R}^n$ el arreglo de coeficientes de tal funcional.

   a) Tómese $\vec{w}_1 = \begin{bmatrix}0{,}5\\0{,}2\end{bmatrix}$ para la función f_1 y $\vec{w}_2 = \begin{bmatrix}-0{,}1\\0{,}05\end{bmatrix}$ para la función f_2, (funciones con dominio en $\mathbb{R}^2$ y codominio en). Grafique ambos planos en Pytorch. (ver colab)

   b) Exprese ambos planos en su forma vectorial $\vec{n}\cdot\overrightarrow{P_0P_1} = 0$, con $\vec{n}$ el vector normal a tal hiperplano, P_0 y P puntos sobre el mismo. Un vector normal $\vec{n}\in\mathbb{R}^m$ a dos vectores $\vec{v}\in\mathbb{R}^m$ y $\vec{u}\in\mathbb{R}^m$ se obtiene en términos del producto cruz como $\vec{n} = \vec{v}\times\vec{u}$, el cual se define a continuación (con $A_{\setminus i}$ el resultado de eliminar la columna i de la matriz A):

$$\vec{n} = \text{det}\left(A_{\setminus 1}\right)\hat{i} - \text{det}\left(A_{\setminus 2}\right)\hat{j} + \text{det}\left(A_{\setminus 3}\right)\hat{k}$$

   con $A = \begin{bmatrix} - & \vec{v} & - \\ - & \vec{u} & - \end{bmatrix}$
   R/ para $\vec{w_1}$obtenemos un P_0usando $(x,y) = (2,3)$ y P usando $(x,y) = (3,2)$

$$\begin{aligned} f_1(x,y) &= 0{,}5x + 0{,}2y \\ f_1(2,3) &= 0{,}5(2) + 0{,}2(3) = 1{,}6 \\ f_1(3,2) &= 0{,}5(3) + 0{,}2(2) = 1{,}9 \end{aligned}$$

Tenemos $P_0 = (2,3,1{,}6)$ y $P = (3,2,1{,}9)$ que usamos para calcular $\vec{n}$:

$$\vec{n} = \begin{bmatrix} 2 \\ 3 \\ 1{,}6 \end{bmatrix} \times \begin{bmatrix} 3 \\ 2 \\ 1{,}9 \end{bmatrix} = \begin{bmatrix} 3(1{,}9) - 1{,}6(2) \\ 1{,}6(3) - 2(1{,}9) \\ 2(2) - 3(3) \end{bmatrix}$$
$$= \begin{bmatrix} 2{,}5 \\ 1 \\ -5 \end{bmatrix}$$

obtenemos el vector $\overrightarrow{P_0P_1}$:

$$\overrightarrow{P_0P_1} = \begin{bmatrix} 2 \\ 3 \\ 1{,}6 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \\ 1{,}9 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ -0{,}3 \end{bmatrix}$$

El plano dado por $\vec{w_1}$se expresa en forma vectorial como:

$$\begin{bmatrix} 2{,}5 \\ 1 \\ -5 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 1 \\ -0{,}3 \end{bmatrix} = 0$$

R/ para $\vec{w_2}$obtenemos un P_0usando $(x,y) = (2,3)$ y P usando $(x,y) = (3,2)$

$$f_1(x,y) = -0{,}1x + 0{,}05y$$
$$f_1(2,3) = -0{,}1(2) + 0{,}05(3) = -0{,}05$$
$$f_1(3,2) = -0{,}1(3) + 0{,}05(2) = -0{,}2$$

Tenemos $P_0 = (2,3,1{,}6)$ y $P = (3,2,1{,}9)$ que usamos para calcular $\vec{n}$:

$$\vec{n} = \begin{bmatrix} 2 \\ 3 \\ -0{,}05 \end{bmatrix} \times \begin{bmatrix} 3 \\ 2 \\ -0{,}2 \end{bmatrix} = \begin{bmatrix} 3(-0{,}2) - -0{,}05(2) \\ -0{,}05(3) - 2(-0{,}2) \\ 2(2) - 3(3) \end{bmatrix}$$
$$= \begin{bmatrix} -0{,}5 \\ 0{,}25 \\ -5 \end{bmatrix}$$

obtenemos el vector $\overrightarrow{P_0P_1}$:

$$\overrightarrow{P_0P_1} = \begin{bmatrix} 2 \\ 3 \\ -0{,}05 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \\ -0{,}2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 0{,}15 \end{bmatrix}$$

El plano dado por $\vec{w_2}$se expresa en forma vectorial como:

$$\begin{bmatrix} -0{,}5 \\ 0{,}25 \\ -5 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 1 \\ 0{,}15 \end{bmatrix} = 0$$

2. **El vector gradiente**: Para cada una de las siguientes funciones multivariable: grafique su superficie, calcule el vector gradiente manualmente, evaluelo y grafique el vector unitario en la dirección del gradiente para los dos puntos especificados (en la misma figura de la superficie) y calcule la magnitud de tal vector gradiente en cada punto. Además calcule la matriz Hessiana. En general, investigue ¿qué indica la magnitud del vector gradiente y la matriz Hessiana?

   a) $f(x, y) = x^3y^2+1$, evaluación del gradiente en los puntos $P_0 = (0, 0)$ y $P_1 = (7{,}4, -6{,}3)$. (ver colab)

$$\frac{\partial}{\partial x}\left(x^3y^2 + 1\right) = 3x^2y^2$$
$$\frac{\partial}{\partial y}\left(x^3y^2 + 1\right) = 2x^3y$$
$$\nabla f = \begin{bmatrix} 3x^2y^2 \\ 2x^3y \end{bmatrix}$$

   Derivadas para formar la matriz Hessiana:

$$\frac{\partial^2}{\partial xx}\left(x^3y^2 + 1\right) = \frac{\partial}{\partial x}\left(3x^2y^2\right) = 6xy^2$$
$$\frac{\partial^2}{\partial xy}\left(x^3y^2 + 1\right) = \frac{\partial}{\partial y}\left(3x^2y^2\right) = 6x^2y$$
$$\frac{\partial^2}{\partial y^2}\left(x^3y^2 + 1\right) = \frac{\partial}{\partial y}\left(2x^3y\right) = 2x^3$$

   La matriz Hessiana:

$$Hf(x, y) = \begin{bmatrix} 6xy^2 & 6x^2y \\ 6x^2y & 2x^3 \end{bmatrix}$$

b) $f(x,y) = e^x + e^y + 2x^2 + 4y + 3$, evaluación del gradiente en los puntos $P_0 = (3{,}8, 1{,}8)$ y $P_1 = (6{,}2, 7)$.(ver colab)

$$\begin{aligned}
\frac{\partial}{\partial x}\left(e^x + e^y + 2x^2 + 4y + 3\right) &= 4x + e^x \\
\frac{\partial}{\partial y}\left(e^x + e^y + 2x^2 + 4y + 3\right) &= e^y + 4 \\
\nabla f &= \begin{bmatrix} 4x + e^x \\ e^y + 4 \end{bmatrix}
\end{aligned}$$

Derivadas para formar la matriz Hessiana:

$$\begin{aligned}
\frac{\partial^2}{\partial xx}\left(e^x + e^y + 2x^2 + 4y + 3\right) &= \frac{\partial}{\partial x}\left(4x + e^x\right) = 4 + e^x \\
\frac{\partial^2}{\partial xy}\left(e^x + e^y + 2x^2 + 4y + 3\right) &= \frac{\partial}{\partial y}\left(4x + e^x\right) = 0 \\
\frac{\partial^2}{\partial yy}\left(e^x + e^y + 2x^2 + 4y + 3\right) &= \frac{\partial}{\partial y}\left(e^y + 4\right) = e^y
\end{aligned}$$

La matriz Hessiana:

$$Hf(x,y) = \begin{bmatrix} 4 + e^x & 0 \\ 0 & e \end{bmatrix}$$

c) $f(x,y) = \ln\left(x^2 + y^2\right)$, evaluación del gradiente en los puntos $P_0 = (-2{,}4, 6{,}7)$ y $P_1 = (0, 4{,}3)$.(ver colab)

$$\begin{aligned}
\frac{\partial}{\partial x}\left(\ln\left(x^2 + y^2\right)\right) &= \frac{1}{x^2 + y^2} \cdot \frac{\partial}{\partial x}(x^2 + y^2) \\
&= \frac{2x}{x^2 + y^2} \\
\frac{\partial}{\partial y}\left(\ln\left(x^2 + y^2\right)\right) &= \frac{1}{x^2 + y^2} \cdot \frac{\partial}{\partial y}(x^2 + y^2) \\
&= \frac{2y}{x^2 + y^2} \\
\nabla f &= \begin{bmatrix} \frac{2x}{x^2+y^2} \\ \frac{2y}{x^2+y^2} \end{bmatrix}
\end{aligned}$$

Derivadas para formar la matriz Hessiana:

$$\frac{\partial^2}{\partial xx}\ln\left(x^2+y^2\right)=\frac{\partial}{\partial x}\left(\frac{2x}{x^2+y^2}\right)=\frac{2\left(x^2+y^2\right)-2x(2x)}{\left(x^2+y^2\right)^2}=\frac{2y^2-2x^2}{\left(x^2+y^2\right)^2}$$
$$\frac{\partial^2}{\partial xy}\ln\left(x^2+y^2\right)=\frac{\partial}{\partial y}\left(\frac{2x}{x^2+y^2}\right)=\frac{\left(x^2+y^2\right)0-2x(2y)}{\left(x^2+y^2\right)^2}=\frac{4xy}{\left(x^2+y^2\right)^2}$$
$$\frac{\partial^2}{\partial yy}\ln\left(x^2+y^2\right)=\frac{\partial}{\partial x}\left(\frac{2y}{x^2+y^2}\right)=\frac{2\left(x^2+y^2\right)-2y(2y)}{\left(x^2+y^2\right)^2}=\frac{2x^2-2y^2}{\left(x^2+y^2\right)^2}$$

La matriz Hessiana:

$$Hf(x,y)=\begin{bmatrix}\frac{2y^2-2x^2}{(x^2+y^2)^2} & \frac{4xy}{(x^2+y^2)^2}\\ \frac{4xy}{(x^2+y^2)^2} & \frac{2x^2-2y^2}{(x^2+y^2)^2}\end{bmatrix}$$

R/ La magnitud del vector gradiente indica la máxima razón de cambio de la función f en el punto (x, y).

La matriz Hessina evaluada en los puntos críticos permite determinar si se trata de un punto silla, un mínimo o máximo local:

a) si $det(Hf(x,y) < 0$ es un punto silla

b) si $det(Hf(x,y) > 0$ es un máximo o minimo

c) si $det(Hf(x,y) = 0$ no se puede determinar con la segundas derivadas

4. (10 puntos extra) La distancia de Minkowski como índice de error

Tómese el siguiente problema: Usted labora en el departamento de ciencias de los datos de una cooperativa de productores de café. Los agricultores desean estimar la productividad en quintales por héctarea. Los datos más sencillos a recolectar en corto plazo, corresponden a la humedad y temperatura, componentes del vector de características a emplear por su modelo:

$$\overrightarrow{x}=\begin{bmatrix}x_1\\x_2\end{bmatrix}$$

donde x_1 y x_2 corresponden a la temperatura en grados, y la humedad relativa (%), respectivamente. El modelo y que usted construirá, tiene entonces como entrada una muestra $\overrightarrow{x}_i$, y salida la productividad estimada en quintales por hectárea:

$$\widetilde{t}_i = y\left(\overrightarrow{x}_i\right)$$

Usted dispone de un conjunto de $m = 30$ muestras con sus correspondientes productividades en quintales por hectárea:

$$X = \begin{bmatrix} - & \overrightarrow{x}_1 & - \\ & \vdots & \\ - & \overrightarrow{x}_m & - \end{bmatrix} \quad T = \begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix}$$

Usted debe elegir entre dos métricas para cuantificar la exactitud de su modelo: El **error medio absoluto** o MAE en inglés, y la **raíz del error cuadrado medio** o RMSE en inglés.

Basada en la distancia ℓ_1, el error medio absoluto o MAE en inglés se define como,

$$\text{MAE} = \frac{1}{m}\sum_{i=1}^{m} |\tilde{t}_i - t_i|$$

mientras que la raíz del error medio cuadrado (RMSE en inglés) está dada por:

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} \left(\tilde{t}_i - t_i\right)^2}$$

basado en la distancia ℓ_2.

Similitudes entre las métricas: Ambas métricas son similares en el sentido de que expresan el error promedio de predicción del modelo en unidades de la variable de interés (en este caso quintales por hectárea). Además, ambas métricas pueden variar de 0 a ∞ y son indiferentes a la dirección de los errores. Son puntuaciones negativamente orientadas, lo que significa que los valores más bajos son mejores.

Diferencias entre las métricas: Explique la diferencia entre las métricas, analizándolo desde el punto de vista de la sensibilidad a los valores atípicos. La Tabla 1 muestra 3 particiones del set de datos, las cuales le ayudarán a responder entonces la pregunta ¿cuál métrica es más sensible a los valores atípicos? Use como apoyo el cálculo en Pytorch del MAE y el RMSE de los datos pendientes (?) en la tabla 1. Realice el cálculo de forma matricial, prescindiendo al máximo de estructuras de repetición.

i	t_i	$\tilde{t}_i$	$\lvert\tilde{t}_i - t_i\rvert$	$\left(\tilde{t}_i - t_i\right)^2$
1	4	2		
2	6	4		
3	5	3		
4	6	4		
5	8	6		
6	10	8		
7	7	5		
8	4	2		
9	2	4		
10	8	10		

MAE $=?$, $\sigma_{\text{MAE}} =?$

RMSE $=?$,$\sigma_{\text{RMSE}} =?$

i	t_i	$\tilde{t}_i$	$\lvert\tilde{t}_i - t_i\rvert$	$\left(\tilde{t}_i - t_i\right)^2$
11	5	4		1
12	3	2		1
13	2	3		1
14	4	5		1
15	20	21		1
16	32	29		9
17	5	2		9
18	4	7		9
19	7	4		9
20	41	38		9

MAE $=?$, $\sigma_{\text{MAE}} =?$

RMSE $=?$,$\sigma_{\text{RMSE}} =?$

i	t_i	$\tilde{t}_i$	$\lvert\tilde{t}_i - t_i\rvert$	$\left(\tilde{t}_i - t_i\right)^2$
21	6	6		
22	20	20		
23	31	31		
24	41	41		
25	50	50		
26	62	62		
27	73	73		
28	4	4		
29	7	7		
30	40	20		

MAE $=?$, $\sigma_{\text{MAE}} =?$

RMSE $=?$,$\sigma_{\text{RMSE}} =?$

Cuadro 1: 3 particiones disjuntas del conjunto de muestras.