Meta 3.2 Caso de estudio: Modelo de regresión multivariada.

El modelo de regresión multivariada relaciona más de un predictor y más de una respuesta.

La regresión multivariada es un método utilizado para medir el grado en que más de una variable independiente (predictores) y más de una variable dependiente (respuestas) están relacionadas linealmente. El método se usa ampliamente para predecir el comportamiento de las variables de respuesta asociadas a los cambios en las variables predictoras, una vez que se ha establecido un grado de relación deseada.

Sean q observaciones $\{x_{p,i}; y_{p,j}\}_{p=1}^{q}$ con *n* predictores (**X**) y *m* respuestas (**Y**). Sea $\boldsymbol{Y}$ la matriz de respuestas q×m, $\boldsymbol{A}$ es la matriz de diseño de tamaño $q \times \varrho$ tal que las *n* primeras columnas son los predictores (**X**) y la última columna sean 1's, es decir $\boldsymbol{A} = [\boldsymbol{X}|\mathbf{1}_{q\times 1}]$. Sea $\boldsymbol{\Theta}$ una matriz $\varrho \times m$ de parámetros fijos, $\boldsymbol{\Xi}$ una matriz q×m tal que $\boldsymbol{\Xi} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$ (multivariada normalmente distribuida con matriz de covarianza $\boldsymbol{\Sigma}$) y $\varrho = n + 1$ es el numero de parámetros.

Dado el siguiente conjunto de datos de entrenamiento (challenge00_dataset22.txt) de dos atributos predictores (X=[$x_{p,i}$] para i=1,2 y p=1…4) y dos atributos de respuesta (Y=[$y_{p,j}$] para j=1,2 y p=1...4).

$$X = \begin{bmatrix} 4.7 & 6.0 \\ 6.1 & 3.9 \\ 2.9 & 4.2 \\ 7.0 & 5.5 \end{bmatrix} \qquad Y = \begin{bmatrix} 3.52 & 4.02 \\ 5.43 & 6.23 \\ 4.95 & 5.76 \\ 4.70 & 4.28 \end{bmatrix}$$

Realizar un análisis para demostrar lo siguiente:

1. Función de la hipótesis del modelo de regresión lineal múltiple con múltiples salidas.
2. Función de costo (Métricas de regresión SSE, MSE, RMSE)
3. Función gradiente, $\nabla E(\Theta) \equiv \frac{\partial E}{\partial \boldsymbol{\Theta}}$
4. Función jacobiana, $\boldsymbol{J}_{\vec{\Xi}} \equiv \frac{\partial \vec{\Xi}}{\partial \vec{\boldsymbol{\Theta}}}$
5. Encontrar los parámetros del modelo *lineal múltiple con multiples salidas*.
6. Implementar las funciones de los puntos 1-4.

1. Función de la hipótesis del modelo de regresión multivariada.

La hipótesis de regresión para este caso de estudio se define como;

$\hat{y}_{p,j} = \theta_{1,j} x_{p,1} + \theta_{2,j} x_{p,2} + \theta_{3,j}$ para j=1,2 y p=1...4

$e_{p,j} = y_{p,j} - \hat{y}_{p,j}$

Calcular para todas las p=1,2,3,4 y j=1,2

$$\hat{y}_{1,1} = \theta_{1,1}x_{1,1} + \theta_{2,1}x_{1,2} + \theta_{3,1} \qquad \hat{y}_{1,2} = \theta_{1,2}x_{1,1} + \theta_{2,2}x_{1,2} + \theta_{3,2}$$
$$\hat{y}_{2,1} = \theta_{1,1}x_{2,1} + \theta_{2,1}x_{2,2} + \theta_{3,1} \qquad \hat{y}_{2,2} = \theta_{1,2}x_{2,1} + \theta_{2,2}x_{2,2} + \theta_{3,2}$$
$$\hat{y}_{3,1} = \theta_{1,1}x_{3,1} + \theta_{2,1}x_{3,2} + \theta_{3,1} \qquad \hat{y}_{3,2} = \theta_{1,2}x_{3,1} + \theta_{2,2}x_{3,2} + \theta_{3,2}$$
$$\hat{y}_{4,1} = \theta_{1,1}x_{4,1} + \theta_{2,1}x_{4,2} + \theta_{3,1} \qquad \hat{y}_{4,2} = \theta_{1,2}x_{4,1} + \theta_{2,2}x_{4,2} + \theta_{3,2}$$

$$\begin{bmatrix} \hat{y}_{1,1} & \hat{y}_{1,2} \\ \hat{y}_{2,1} & \hat{y}_{2,2} \\ \hat{y}_{3,1} & \hat{y}_{3,2} \\ \hat{y}_{4,1} & \hat{y}_{4,2} \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & 1 \\ x_{2,1} & x_{2,2} & 1 \\ x_{3,1} & x_{3,2} & 1 \\ x_{4,1} & x_{4,2} & 1 \end{bmatrix} \begin{bmatrix} \theta_{1,1} & \theta_{1,2} \\ \theta_{2,1} & \theta_{2,2} \\ \theta_{3,1} & \theta_{3,2} \end{bmatrix}$$

$$\hat{Y} = A\Theta$$

donde $\hat{Y}$ es la hipótesis del modelo de regresión lineal múltiple con múltiples salidas, A es la matriz de diseño y Θ es la matriz de parámetros del modelo de regresión lineal. A continuación se muestran las variables matriciales.

$$\hat{Y} = \begin{bmatrix} \hat{y}_{1,1} & \hat{y}_{1,2} \\ \hat{y}_{2,1} & \hat{y}_{2,2} \\ \hat{y}_{3,1} & \hat{y}_{3,2} \\ \hat{y}_{4,1} & \hat{y}_{4,2} \end{bmatrix},\ A = \begin{bmatrix} x_{1,1} & x_{1,2} & 1 \\ x_{2,1} & x_{2,2} & 1 \\ x_{3,1} & x_{3,2} & 1 \\ x_{4,1} & x_{4,2} & 1 \end{bmatrix} = [X \mid 1_{4x1}]\ ,\ \Theta = \begin{bmatrix} \theta_{1,1} & \theta_{1,2} \\ \theta_{2,1} & \theta_{2,2} \\ \theta_{3,1} & \theta_{3,2} \end{bmatrix}$$

Las medidas de error se muestran a continuación:

$$e_{1,1} = y_{1,1} - \hat{y}_{1,1} \qquad e_{1,2} = y_{1,2} - \hat{y}_{1,2}$$
$$e_{2,1} = y_{2,1} - \hat{y}_{2,1} \qquad e_{2,2} = y_{2,2} - \hat{y}_{2,2}$$
$$e_{3,1} = y_{3,1} - \hat{y}_{3,1} \qquad e_{3,2} = y_{3,2} - \hat{y}_{3,2}$$
$$e_{4,1} = y_{4,1} - \hat{y}_{4,1} \qquad e_{4,2} = y_{4,2} - \hat{y}_{4,2}$$

La matriz de las medidas de error es:

$$\begin{bmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \\ e_{3,1} & e_{3,2} \\ e_{4,1} & e_{4,2} \end{bmatrix} = \begin{bmatrix} y_{1,1} & y_{1,2} \\ y_{2,1} & y_{2,2} \\ y_{3,1} & y_{3,2} \\ y_{4,1} & y_{4,2} \end{bmatrix} - \begin{bmatrix} \hat{y}_{1,1} & \hat{y}_{1,2} \\ \hat{y}_{2,1} & \hat{y}_{2,2} \\ \hat{y}_{3,1} & \hat{y}_{3,2} \\ \hat{y}_{4,1} & \hat{y}_{4,2} \end{bmatrix}$$

donde

$$\boldsymbol{\Xi} = \begin{bmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \\ e_{3,1} & e_{3,2} \\ e_{4,1} & e_{4,2} \end{bmatrix}, \qquad \boldsymbol{Y} = \begin{bmatrix} y_{1,1} & y_{1,2} \\ y_{2,1} & y_{2,2} \\ y_{3,1} & y_{3,2} \\ y_{4,1} & y_{4,2} \end{bmatrix}, \qquad \widehat{\boldsymbol{Y}} = \begin{bmatrix} \hat{y}_{1,1} & \hat{y}_{1,2} \\ \hat{y}_{2,1} & \hat{y}_{2,2} \\ \hat{y}_{3,1} & \hat{y}_{3,2} \\ \hat{y}_{4,1} & \hat{y}_{4,2} \end{bmatrix}$$

$\boldsymbol{\Xi} = Y - \hat{Y}$

2. Función de costo (Métricas de regresión SSE, MSE, RMSE)

La suma de los errores al cuadrado (SSE) se define como:

$$SSE = e_{1,1}^2 + e_{2,1}^2 + e_{3,1}^2 + e_{4,1}^2 + e_{1,2}^2 + e_{2,2}^2 + e_{3,2}^2 + e_{4,2}^2$$

$$SSE \equiv \sum_{p=1}^{q} \sum_{j=1}^{m} e_{p,j}^2$$

Entonces la media de la suma de los errores al cuadrado (MSE) es:

$$MSE \equiv SSE/(mq)$$

y la raíz de la media de la suma de los errores al cuadrado (RMSE) se evalúa como:

$$RMSE \equiv \sqrt{MSE}$$

Si las matrices $\boldsymbol{\Xi}$ y $\boldsymbol{\Theta}$ son vectorizadas en columna denotadas como:

$$\vec{\boldsymbol{\Xi}} = \begin{bmatrix} e_{1,1} \\ e_{2,1} \\ e_{3,1} \\ e_{4,1} \\ e_{1,2} \\ e_{2,2} \\ e_{3,2} \\ e_{4,2} \end{bmatrix}, \qquad \vec{\boldsymbol{\Theta}} = \begin{bmatrix} \theta_{1,1} \\ \theta_{2,1} \\ \theta_{3,1} \\ \theta_{1,2} \\ \theta_{2,2} \\ \theta_{3,2} \end{bmatrix}$$

entonces el SSE se evalúa como:

$$SSE = \begin{bmatrix} e_{1,1} & e_{2,1} & e_{3,1} & e_{4,1} & e_{1,2} & e_{2,2} & e_{3,2} & e_{4,2} \end{bmatrix} \begin{bmatrix} e_{1,1} \\ e_{2,1} \\ e_{3,1} \\ e_{4,1} \\ e_{1,2} \\ e_{2,2} \\ e_{3,2} \\ e_{4,2} \end{bmatrix}$$

$$SSE = \vec{\boldsymbol{\Xi}}'\vec{\boldsymbol{\Xi}}$$

teta - todos las variables

3. Función gradiente, $\boldsymbol{\nabla E}(\boldsymbol{\Theta}) \equiv \frac{\partial E}{\partial \theta_{i,j}}$

Asumiendo que $E = SSE$, se evalúan las derivadas $\frac{\partial E}{\partial \theta_{i,j}}$ para i=1,…,n+1 y j=1,…,m

Matriz gradiente: $\frac{\partial E}{\partial \theta}$

$$\frac{\partial E}{\partial \theta_{1,1}} = \frac{\partial E}{\partial e_{1,1}}\frac{\partial e_{1,1}}{\partial \hat{y}_{1,1}}\frac{\partial \hat{y}_{1,1}}{\partial \theta_{1,1}} + \frac{\partial E}{\partial e_{2,1}}\frac{\partial e_{2,1}}{\partial \hat{y}_{2,1}}\frac{\partial \hat{y}_{2,1}}{\partial \theta_{1,1}} + \frac{\partial E}{\partial e_{3,1}}\frac{\partial e_{3,1}}{\partial \hat{y}_{3,1}}\frac{\partial \hat{y}_{3,1}}{\partial \theta_{1,1}} + \frac{\partial E}{\partial e_{4,1}}\frac{\partial e_{4,1}}{\partial \hat{y}_{4,1}}\frac{\partial \hat{y}_{4,1}}{\partial \theta_{1,1}}$$

$$\frac{\partial E}{\partial \theta_{1,1}} = (2e_{1,1})\,(-1)\,(x_{1,1}) + (2e_{2,1})\,(-1)\,(x_{2,1}) + (2e_{3,1})\,(-1)\,(x_{3,1}) + (2e_{4,1})\,(-1)\,(x_{4,1})$$

$$\frac{\partial E}{\partial \theta_{1,1}} = -2e_{1,1}x_{1,1} - 2e_{2,1}x_{2,1} - 2e_{3,1}x_{3,1} - 2e_{4,1}x_{4,1}$$

$$\frac{\partial E}{\partial \theta_{1,1}} = -2\sum_{p=1}^{4} e_{p,1}x_{p,1}$$

$$\frac{\partial E}{\partial \theta_{2,1}} = \frac{\partial E}{\partial e_{1,1}}\frac{\partial e_{1,1}}{\partial \hat{y}_{1,1}}\frac{\partial \hat{y}_{1,1}}{\partial \theta_{2,1}} + \frac{\partial E}{\partial e_{2,1}}\frac{\partial e_{2,1}}{\partial \hat{y}_{2,1}}\frac{\partial \hat{y}_{2,1}}{\partial \theta_{2,1}} + \frac{\partial E}{\partial e_{3,1}}\frac{\partial e_{3,1}}{\partial \hat{y}_{3,1}}\frac{\partial \hat{y}_{3,1}}{\partial \theta_{2,1}} + \frac{\partial E}{\partial e_{4,1}}\frac{\partial e_{4,1}}{\partial \hat{y}_{4,1}}\frac{\partial \hat{y}_{4,1}}{\partial \theta_{2,1}}$$

$$\frac{\partial E}{\partial \theta_{2,1}} = (2e_{1,1})\,(-1)\,(x_{1,2}) + (2e_{2,1})\,(-1)\,(x_{2,2}) + (2e_{3,1})\,(-1)\,(x_{3,2}) + (2e_{4,1})\,(-1)\,(x_{4,2})$$

$$\frac{\partial E}{\partial \theta_{2,1}} = -2e_{1,1}x_{1,2} - 2e_{2,1}x_{2,2} - 2e_{3,1}x_{3,2} - 2e_{4,1}x_{4,2}$$

$$\frac{\partial E}{\partial \theta_{2,1}} = -2\sum_{p=1}^{4} e_{p,1}x_{p,2}$$

$$\frac{\partial E}{\partial \theta_{3,1}} = \frac{\partial E}{\partial e_{1,1}}\frac{\partial e_{1,1}}{\partial \hat{y}_{1,1}}\frac{\partial \hat{y}_{1,1}}{\partial \theta_{3,1}} + \frac{\partial E}{\partial e_{2,1}}\frac{\partial e_{2,1}}{\partial \hat{y}_{2,1}}\frac{\partial \hat{y}_{2,1}}{\partial \theta_{3,1}} + \frac{\partial E}{\partial e_{3,1}}\frac{\partial e_{3,1}}{\partial \hat{y}_{3,1}}\frac{\partial \hat{y}_{3,1}}{\partial \theta_{3,1}} + \frac{\partial E}{\partial e_{4,1}}\frac{\partial e_{4,1}}{\partial \hat{y}_{4,1}}\frac{\partial \hat{y}_{4,1}}{\partial \theta_{3,1}}$$

$$\frac{\partial E}{\partial \theta_{3,1}} = (2e_{1,1})(-1)(1) + (2e_{2,1})(-1)(1) + (2e_{3,1})(-1)(1) + (2e_{4,1})(-1)(1)$$

$$\frac{\partial E}{\partial \theta_{3,1}} = -2e_{1,1} - 2e_{2,1} - 2e_{3,1} - 2e_{4,1}$$

$$\frac{\partial E}{\partial \theta_{3,1}} = -2\sum_{p=1}^{4} e_{p,1}$$

$$\frac{\partial E}{\partial \theta_{1,2}} = \frac{\partial E}{\partial e_{1,2}}\frac{\partial e_{1,2}}{\partial \hat{y}_{1,2}}\frac{\partial \hat{y}_{1,2}}{\partial \theta_{1,2}} + \frac{\partial E}{\partial e_{2,2}}\frac{\partial e_{2,2}}{\partial \hat{y}_{2,2}}\frac{\partial \hat{y}_{2,2}}{\partial \theta_{1,2}} + \frac{\partial E}{\partial e_{3,2}}\frac{\partial e_{3,2}}{\partial \hat{y}_{3,2}}\frac{\partial \hat{y}_{3,2}}{\partial \theta_{1,2}} + \frac{\partial E}{\partial e_{4,2}}\frac{\partial e_{4,2}}{\partial \hat{y}_{4,2}}\frac{\partial \hat{y}_{4,2}}{\partial \theta_{1,2}}$$

$$\frac{\partial E}{\partial \theta_{1,2}} = (2e_{1,2})(-1)(x_{1,1}) + (2e_{2,2})(-1)(x_{2,1}) + (2e_{3,2})(-1)(x_{3,1}) + (2e_{4,2})(-1)(x_{4,1})$$

$$\frac{\partial E}{\partial \theta_{1,2}} = -2e_{1,2}x_{1,1} - 2e_{2,2}x_{2,1} - 2e_{3,2}x_{3,1} - 2e_{4,2}x_{4,1}$$

$$\frac{\partial E}{\partial \theta_{1,2}} = -2\sum_{p=1}^{4} e_{p,2}x_{p,1}$$

$$\frac{\partial E}{\partial \theta_{2,2}} = \frac{\partial E}{\partial e_{1,2}}\frac{\partial e_{1,2}}{\partial \hat{y}_{1,2}}\frac{\partial \hat{y}_{1,2}}{\partial \theta_{2,2}} + \frac{\partial E}{\partial e_{2,2}}\frac{\partial e_{2,2}}{\partial \hat{y}_{2,2}}\frac{\partial \hat{y}_{2,2}}{\partial \theta_{2,2}} + \frac{\partial E}{\partial e_{3,2}}\frac{\partial e_{3,2}}{\partial \hat{y}_{3,2}}\frac{\partial \hat{y}_{3,2}}{\partial \theta_{2,2}} + \frac{\partial E}{\partial e_{4,2}}\frac{\partial e_{4,2}}{\partial \hat{y}_{4,2}}\frac{\partial \hat{y}_{4,2}}{\partial \theta_{2,2}}$$

$$\frac{\partial E}{\partial \theta_{2,2}} = (2e_{1,2})(-1)(x_{1,2}) + (2e_{2,2})(-1)(x_{2,2}) + (2e_{3,2})(-1)(x_{3,2}) + (2e_{4,2})(-1)(x_{4,2})$$

$$\frac{\partial E}{\partial \theta_{2,2}} = -2e_{1,2}x_{1,2} - 2e_{2,2}x_{2,2} - 2e_{3,2}x_{3,2} - 2e_{4,2}x_{4,2}$$

$$\frac{\partial E}{\partial \theta_{2,2}} = -2\sum_{p=1}^{4} e_{p,2}x_{p,2}$$

$$\frac{\partial E}{\partial \theta_{3,2}} = \frac{\partial E}{\partial e_{1,2}} \frac{\partial e_{1,2}}{\partial \hat{y}_{1,2}} \frac{\partial \hat{y}_{1,2}}{\partial \theta_{3,2}} + \frac{\partial E}{\partial e_{2,2}} \frac{\partial e_{2,2}}{\partial \hat{y}_{2,2}} \frac{\partial \hat{y}_{2,2}}{\partial \theta_{3,2}} + \frac{\partial E}{\partial e_{3,2}} \frac{\partial e_{3,2}}{\partial \hat{y}_{3,2}} \frac{\partial \hat{y}_{3,2}}{\partial \theta_{3,2}} + \frac{\partial E}{\partial e_{4,2}} \frac{\partial e_{4,2}}{\partial \hat{y}_{4,2}} \frac{\partial \hat{y}_{4,2}}{\partial \theta_{3,2}}$$

$$\frac{\partial E}{\partial \theta_{3,2}} = \left(2e_{1,2}\right)(-1)\,(1) + \left(2e_{2,2}\right)(-1)\,(1) + \left(2e_{3,2}\right)(-1)\,(1) + \left(2e_{4,2}\right)(-1)\,(1)$$

$$\frac{\partial E}{\partial \theta_{3,2}} = -2e_{1,2} - 2e_{2,2} - 2e_{3,2} - 2e_{4,2}$$

$$\frac{\partial E}{\partial \theta_{3,2}} = -2\sum_{p=1}^{4} e_{p,2}$$

El gradiente matricial $\frac{\partial E}{\partial \mathbf{\Theta}} = [\frac{\partial E}{\partial \theta_{i,j}}]$ para i=1,…3 y j=1,2 del caso de estudio se puede simplificar con las siguientes sumatorias obtenidas anteriormente;

$$\frac{\partial E}{\partial \theta_{i,j}} = -2\sum_{p=1}^{4} e_{p,j} x_{p,i}$$

$$\frac{\partial E}{\partial \theta_{3,j}} = -2\sum_{p=1}^{4} e_{p,j}$$

O bien en forma matricial, sustituyendo los resultados anteriores, se expresa como:

$$\frac{\partial E}{\partial \mathbf{\Theta}} = \begin{bmatrix} \frac{\partial E}{\partial \theta_{1,1}} & \frac{\partial E}{\partial \theta_{1,2}} \\ \frac{\partial E}{\partial \theta_{2,1}} & \frac{\partial E}{\partial \theta_{2,2}} \\ \frac{\partial E}{\partial \theta_{3,1}} & \frac{\partial E}{\partial \theta_{3,2}} \end{bmatrix} = \begin{bmatrix} -2\sum_{p=1}^{4} e_{p,1} x_{p,1} & -2\sum_{p=1}^{4} e_{p,2} x_{p,1} \\ -2\sum_{p=1}^{4} e_{p,1} x_{p,2} & -2\sum_{p=1}^{4} e_{p,2} x_{p,2} \\ -2\sum_{p=1}^{4} e_{p,1} & -2\sum_{p=1}^{4} e_{p,2} \end{bmatrix}$$

Expandiendo las sumatorias se obtiene

$$\frac{\partial E}{\partial \mathbf{\Theta}} = \begin{bmatrix} -2e_{1,1}x_{1,1} - 2e_{2,1}x_{2,1} - 2e_{3,1}x_{3,1} - 2e_{4,1}x_{4,1} & -2e_{1,2}x_{1,1} - 2e_{2,2}x_{2,1} - 2e_{3,2}x_{3,1} - 2e_{4,2}x_{4,1} \\ -2e_{1,1}x_{1,2} - 2e_{2,1}x_{2,2} - 2e_{3,1}x_{3,2} - 2e_{4,1}x_{4,2} & -2e_{1,2}x_{1,2} - 2e_{2,2}x_{2,2} - 2e_{3,2}x_{3,2} - 2e_{4,2}x_{4,2} \\ -2e_{1,1} - 2e_{2,1} - 2e_{3,1} - 2e_{4,1} & -2e_{1,2} - 2e_{2,2} - 2e_{3,2} - 2e_{4,2} \end{bmatrix}$$

Factorizando la matriz anterior, obtenemos

$$\frac{\partial E}{\partial \mathbf{\Theta}} = 2\begin{bmatrix} -x_{1,1} & -x_{2,1} & -x_{3,1} & -x_{4,1} \\ -x_{1,2} & -x_{2,2} & -x_{3,2} & -x_{4,2} \\ -1 & -1 & -1 & -1 \end{bmatrix}\begin{bmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \\ e_{3,1} & e_{3,2} \\ e_{4,1} & e_{4,2} \end{bmatrix}$$

Entonces $\frac{\partial E}{\partial \mathbf{\Theta}}$ es 2 veces la matriz de diseño transpuesta negativa (-$\boldsymbol{A}'$) multiplicado por la matriz $\mathbf{\Xi}$, es decir;

$$\frac{\partial E}{\partial \mathbf{\Theta}} = 2(-\boldsymbol{A}')\mathbf{\Xi}$$

4. Función jacobiana, $\boldsymbol{J}_{\vec{\Xi}} \equiv \frac{\partial \vec{\Xi}}{\partial \vec{\Theta}}$

Desarrollando el vector gradiente $\frac{\partial E}{\partial \vec{\mathbf{\Theta}}}$, obtenemos:

$$\frac{\partial E}{\partial \vec{\mathbf{\Theta}}} = 2\, e_{1,1}\frac{\partial e_{1,1}}{\partial \vec{\mathbf{\Theta}}} + 2\, e_{2,1}\frac{\partial e_{2,1}}{\partial \vec{\mathbf{\Theta}}} + 2\, e_{3,1}\frac{\partial e_{3,1}}{\partial \vec{\mathbf{\Theta}}} + 2\, e_{4,1}\frac{\partial e_{4,1}}{\partial \vec{\mathbf{\Theta}}} +$$

$$2\, e_{1,2}\frac{\partial e_{1,2}}{\partial \vec{\mathbf{\Theta}}} + 2\, e_{2,2}\frac{\partial e_{2,2}}{\partial \vec{\mathbf{\Theta}}} + 2\, e_{3,2}\frac{\partial e_{3,2}}{\partial \vec{\mathbf{\Theta}}} + 2\, e_{4,2}\frac{\partial e_{4,2}}{\partial \vec{\mathbf{\Theta}}}$$

$$\frac{\partial E}{\partial \vec{\mathbf{\Theta}}} = 2\begin{bmatrix} \frac{\partial e_{1,1}}{\partial \vec{\mathbf{\Theta}}} & \frac{\partial e_{2,1}}{\partial \vec{\mathbf{\Theta}}} & \frac{\partial e_{3,1}}{\partial \vec{\mathbf{\Theta}}} & \frac{\partial e_{4,1}}{\partial \vec{\mathbf{\Theta}}} & \frac{\partial e_{1,2}}{\partial \vec{\mathbf{\Theta}}} & \frac{\partial e_{2,2}}{\partial \vec{\mathbf{\Theta}}} & \frac{\partial e_{3,2}}{\partial \vec{\mathbf{\Theta}}} & \frac{\partial e_{4,2}}{\partial \vec{\mathbf{\Theta}}} \end{bmatrix}\begin{bmatrix} e_{1,1} \\ e_{2,1} \\ e_{3,1} \\ e_{4,1} \\ e_{1,2} \\ e_{2,2} \\ e_{3,2} \\ e_{4,2} \end{bmatrix}$$

$$\frac{\partial E}{\partial \vec{\mathbf{\Theta}}} = 2\, \boldsymbol{J}_{\vec{\Xi}}'\, \vec{\mathbf{\Xi}}$$

Entonces el vector gradiente $\frac{\partial E}{\partial \vec{\mathbf{\Theta}}}$ es 2 veces la jacobiana traspuesta de $\vec{\mathbf{\Xi}}$ ($\boldsymbol{J}_{\vec{\Xi}}$) multiplicada por el vector $\vec{\mathbf{\Xi}}$. La deriva parcial del vector $\vec{\mathbf{\Xi}}$ con respecto al vector $\vec{\mathbf{\Theta}}$ denotada como $\frac{\partial \vec{\Xi}}{\partial \vec{\Theta}}$, es la matriz jacobiana de $\vec{\mathbf{\Xi}}$ definida como $\boldsymbol{J}_{\vec{\Xi}} \equiv \frac{\partial \vec{\Xi}}{\partial \vec{\Theta}}$, es decir;

$$\boldsymbol{J}_{\vec{\Xi}} \equiv \frac{\boldsymbol{\partial}\vec{\boldsymbol{\Xi}}}{\boldsymbol{\partial}\vec{\boldsymbol{\Theta}}} = \begin{bmatrix} \frac{\partial e_{1,1}}{\partial\vec{\boldsymbol{\Theta}}} \\ \frac{\partial e_{2,1}}{\partial\vec{\boldsymbol{\Theta}}} \\ \frac{\partial e_{3,1}}{\partial\vec{\boldsymbol{\Theta}}} \\ \frac{\partial e_{4,1}}{\partial\vec{\boldsymbol{\Theta}}} \\ \frac{\partial e_{1,2}}{\partial\vec{\boldsymbol{\Theta}}} \\ \frac{\partial e_{2,2}}{\partial\vec{\boldsymbol{\Theta}}} \\ \frac{\partial e_{3,2}}{\partial\vec{\boldsymbol{\Theta}}} \\ \frac{\partial e_{4,2}}{\partial\vec{\boldsymbol{\Theta}}} \end{bmatrix} = \begin{bmatrix} \frac{\partial e_{1,1}}{\partial\theta_{1,1}} & \frac{\partial e_{1,1}}{\partial\theta_{2,1}} & \frac{\partial e_{1,1}}{\partial\theta_{3,1}} & 0 & 0 & 0 \\ \frac{\partial e_{2,1}}{\partial\theta_{1,1}} & \frac{\partial e_{2,1}}{\partial\theta_{2,1}} & \frac{\partial e_{2,1}}{\partial\theta_{3,1}} & 0 & 0 & 0 \\ \frac{\partial e_{3,1}}{\partial\theta_{1,1}} & \frac{\partial e_{3,1}}{\partial\theta_{2,1}} & \frac{\partial e_{3,1}}{\partial\theta_{3,1}} & 0 & 0 & 0 \\ \frac{\partial e_{4,1}}{\partial\theta_{1,1}} & \frac{\partial e_{4,1}}{\partial\theta_{2,1}} & \frac{\partial e_{4,1}}{\partial\theta_{3,1}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\partial e_{1,2}}{\partial\theta_{1,2}} & \frac{\partial e_{1,2}}{\partial\theta_{2,2}} & \frac{\partial e_{1,2}}{\partial\theta_{3,2}} \\ 0 & 0 & 0 & \frac{\partial e_{2,2}}{\partial\theta_{1,2}} & \frac{\partial e_{2,2}}{\partial\theta_{2,2}} & \frac{\partial e_{2,2}}{\partial\theta_{3,2}} \\ 0 & 0 & 0 & \frac{\partial e_{3,2}}{\partial\theta_{1,2}} & \frac{\partial e_{3,2}}{\partial\theta_{2,2}} & \frac{\partial e_{3,2}}{\partial\theta_{3,2}} \\ 0 & 0 & 0 & \frac{\partial e_{4,2}}{\partial\theta_{1,2}} & \frac{\partial e_{4,2}}{\partial\theta_{2,2}} & \frac{\partial e_{4,2}}{\partial\theta_{3,2}} \end{bmatrix}$$

$$\boldsymbol{J}_{\vec{\Xi}} \equiv \frac{\boldsymbol{\partial}\vec{\boldsymbol{\Xi}}}{\boldsymbol{\partial}\vec{\boldsymbol{\Theta}}} = \begin{bmatrix} -x_{1,1} & -x_{1,2} & -1 & 0 & 0 & 0 \\ -x_{2,1} & -x_{2,2} & -1 & 0 & 0 & 0 \\ -x_{3,1} & -x_{3,2} & -1 & 0 & 0 & 0 \\ -x_{4,1} & -x_{4,2} & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -x_{1,1} & -x_{1,2} & -1 \\ 0 & 0 & 0 & -x_{2,1} & -x_{2,2} & -1 \\ 0 & 0 & 0 & -x_{3,1} & -x_{3,2} & -1 \\ 0 & 0 & 0 & -x_{4,1} & -x_{4,2} & -1 \end{bmatrix} = -\begin{bmatrix} \boldsymbol{A}_{4\times3} & \boldsymbol{0}_{4\times3} \\ \boldsymbol{0}_{4\times3} & \boldsymbol{A}_{4\times3} \end{bmatrix}$$

Entonces la jacobiana $\boldsymbol{J}_{\vec{\Xi}}$ para m variables dependientes y ϱ parámetros es:

$$\boldsymbol{J}_{\vec{\Xi}} \equiv \frac{\boldsymbol{\partial}\vec{\boldsymbol{\Xi}}}{\boldsymbol{\partial}\vec{\boldsymbol{\Theta}}} = -\begin{bmatrix} \boldsymbol{A}_{q\times\varrho} & \boldsymbol{0}_{q\times\varrho} & \cdots & \boldsymbol{0}_{q\times\varrho} \\ \boldsymbol{0}_{q\times\varrho} & \boldsymbol{A}_{q\times\varrho} & \cdots & \boldsymbol{0}_{q\times\varrho} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0}_{q\times\varrho} & \boldsymbol{0}_{q\times\varrho} & \cdots & \boldsymbol{A}_{q\times\varrho} \end{bmatrix}$$

Calculando el gradiente, usando la matriz jacobiana, obtenemos:

$$\frac{\partial E}{\partial \vec{\boldsymbol{\Theta}}} = 2\,\boldsymbol{J}'_{\vec{\Xi}}\,\vec{\boldsymbol{\Xi}} = 2\begin{bmatrix} -x_{1,1} & -x_{2,1} & -x_{3,1} & -x_{4,1} & 0 & 0 & 0 & 0 \\ -x_{1,2} & -x_{2,2} & -x_{3,2} & -x_{4,2} & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -x_{1,1} & -x_{2,1} & -x_{3,1} & -x_{4,1} \\ 0 & 0 & 0 & 0 & -x_{1,2} & -x_{2,2} & -x_{3,2} & -x_{4,2} \\ 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \end{bmatrix}\begin{bmatrix} e_{1,1} \\ e_{2,1} \\ e_{3,1} \\ e_{4,1} \\ e_{1,2} \\ e_{2,2} \\ e_{3,2} \\ e_{4,2} \end{bmatrix}$$

Multiplicando las matrices, obtenemos;

$$\frac{\partial E}{\partial \vec{\boldsymbol{\Theta}}} = 2\,\boldsymbol{J}'_{\vec{\Xi}}\,\vec{\boldsymbol{\Xi}} = \begin{bmatrix} -2e_{1,1}x_{1,1} - 2e_{2,1}x_{2,1} - 2e_{3,1}x_{3,1} - 2e_{4,1}x_{4,1} \\ -2e_{1,1}x_{1,2} - 2e_{2,1}x_{2,2} - 2e_{3,1}x_{3,2} - 2e_{4,1}x_{4,2} \\ -2e_{1,1} - 2e_{2,1} - 2e_{3,1} - 2e_{4,1} \\ -2e_{1,2}x_{1,1} - 2e_{2,2}x_{2,1} - 2e_{3,2}x_{3,1} - 2e_{4,2}x_{4,1} \\ -2e_{1,2}x_{1,2} - 2e_{2,2}x_{2,2} - 2e_{3,2}x_{3,2} - 2e_{4,2}x_{4,2} \\ -2e_{1,2} - 2e_{2,2} - 2e_{3,2} - 2e_{4,2} \end{bmatrix}$$

Si el vector gradiente $\frac{\partial E}{\partial \vec{\boldsymbol{\Theta}}}$ se desvectoriza obtenemos la matriz gradiente $\frac{\partial E}{\partial \boldsymbol{\Theta}}$, denotada como:

$$\frac{\partial E}{\partial \boldsymbol{\Theta}} = \begin{bmatrix} -2e_{1,1}x_{1,1} - 2e_{2,1}x_{2,1} - 2e_{3,1}x_{3,1} - 2e_{4,1}x_{4,1} & -2e_{1,2}x_{1,1} - 2e_{2,2}x_{2,1} - 2e_{3,2}x_{3,1} - 2e_{4,2}x_{4,1} \\ -2e_{1,1}x_{1,2} - 2e_{2,1}x_{2,2} - 2e_{3,1}x_{3,2} - 2e_{4,1}x_{4,2} & -2e_{1,2}x_{1,2} - 2e_{2,2}x_{2,2} - 2e_{3,2}x_{3,2} - 2e_{4,2}x_{4,2} \\ -2e_{1,1} - 2e_{2,1} - 2e_{3,1} - 2e_{4,1} & -2e_{1,2} - 2e_{2,2} - 2e_{3,2} - 2e_{4,2} \end{bmatrix}$$

Factorizando la matriz anterior, es decir;

$$\frac{\partial E}{\partial \boldsymbol{\Theta}} = 2\begin{bmatrix} -x_{1,1} & -x_{2,1} & -x_{3,1} & -x_{4,1} \\ -x_{1,2} & -x_{2,2} & -x_{3,2} & -x_{4,2} \\ -1 & -1 & -1 & -1 \end{bmatrix}\begin{bmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \\ e_{3,1} & e_{3,2} \\ e_{4,1} & e_{4,2} \end{bmatrix}$$

Entonces $\frac{\partial E}{\partial \boldsymbol{\Theta}}$ es 2 veces la matriz de diseño transpuesta negativa ($-\boldsymbol{A}'$) multiplicado por la matriz $\boldsymbol{\Xi}$, es decir;

$$\frac{\partial E}{\partial \Theta} = 2(-\boldsymbol{A}')\boldsymbol{\Xi}$$

Ambas demostraciones nos llevan al mismo resultado del gradiente en forma matricial.

5. **Encontrar los parámetros del modelo *de* regresión multivariada.**

Optimización de la función de costo, $E(\Theta)$:

$$\widehat{\Theta} = \arg\min_{\Theta} E$$

$$\nabla E(\widehat{\Theta}) = -2\boldsymbol{A}'\boldsymbol{\Xi} = 0$$

$$A'\boldsymbol{\Xi} = 0$$

$$A'(Y - A\widehat{\Theta}) = 0$$

$$A'Y - A'A\widehat{\Theta} = 0$$

$$A'Y = A'A\widehat{\Theta}$$

$$A'A\widehat{\Theta} = A'Y$$

$$\widehat{\Theta} = (A'A)^{-1}A'Y$$

$\widehat{\Theta} = (A'A)^{+}A'Y$, donde la pseudoinversa de Moore-Penrose, $(A'A)^{+}$ de una matriz $A'A$ es una generalización de la matriz inversa.

6. **Implementar las funciones de los puntos 1-4.**

Resumen

Hipótesis del modelo de regresión multivariada

$$\hat{y}_{p,j} = \sum_{i=1}^{n} \theta_{i,j} x_{p,i} + \theta_{n+1,j}$$

$$\hat{Y} = A\Theta$$

donde la matriz de diseño para este caso es, $\boldsymbol{A} = [\boldsymbol{X} \mid \boldsymbol{1}]$

Término de error*:*

$$\boldsymbol{\Xi} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}$$

Funciones de costo para regresión multivariada

$$SSE = \vec{\boldsymbol{\Xi}}'\vec{\boldsymbol{\Xi}}$$

$$MSE = \frac{SSE}{qm}$$

$$RMSE = \sqrt{MSE}$$

Función gradiente para regresión multivariada

$\frac{\partial E}{\partial \theta_{i,j}} = -2\sum_{p=1}^{q} e_{p,j} x_{p,i}$, para i=1,…,n+1; j=1,…,m. Asimismo $x_{p,n+1} = 1$

$$\frac{\partial E}{\partial \boldsymbol{\Theta}} = 2(-\boldsymbol{A}')\boldsymbol{\Xi}$$

$$\frac{\partial E}{\partial \vec{\boldsymbol{\Theta}}} = 2\,\boldsymbol{J}'_{\vec{\Xi}}\,\vec{\boldsymbol{\Xi}}$$

donde $\boldsymbol{J}_{\vec{\Xi}} = -\begin{bmatrix} \boldsymbol{A}_{q\times\varrho} & \boldsymbol{0}_{q\times\varrho} & \cdots & \boldsymbol{0}_{q\times\varrho} \\ \boldsymbol{0}_{q\times\varrho} & \boldsymbol{A}_{q\times\varrho} & \cdots & \boldsymbol{0}_{q\times\varrho} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0}_{q\times\varrho} & \boldsymbol{0}_{q\times\varrho} & \cdots & \boldsymbol{A}_{q\times\varrho} \end{bmatrix}$

Métricas de regresión para evaluar la precisión

a. Coeficiente de determinación (R^2)

$$R^2 = 1 - \frac{\sum_{p=1}^{q}\left(y_p - \hat{y}_p\right)^2}{\sum_{p=1}^{q}\left(y_p - \bar{y}\right)^2}$$

b. Coeficiente de determinación (R^2_{ajus})

$$R^2_{ajus} = 1 - \left(\frac{q-1}{q-\varrho-1}\right)(1-R^2)$$

donde q es el número de datos de entrenamiento y ϱ es el número de parámetros del modelo.