Matemática para ciencias de los datos: Trabajo práctico 1

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)

15 de mayo de 2019

Fecha de entrega: Domingo 26 de Mayo del 2019

Entrega: Un archivo .zip con el código fuente LaTeX o Lyx, el pdf, y un jupyter notebook en Pytorch, debidamente documentado, con una función definida por ejercicio. A través del TEC-digital.

Modo de trabajo: Grupos de 2 personas.

Resumen

En el presente trabajo práctico se repasarán aspectos básicos del algebra lineal, relacionados con los conceptos a desarrollar a lo largo del curso, mezclando aspectos teóricos y prácticos, usando el lenguaje Python con la librería Pytorch.

1. (**20 puntos**) Implemente la función *calcularTrazaMatriz* la cual calcule la traza de una matriz usando únicamente operaciones básicas en pytorch (multiplicación, multiplicación por elemento, matriz identidad, etc.), prescindiendo de estructuras de repetición como el *for* el *while*.

   a) Documente su correcto funcionamiento con matrices arbitrarias $A, B \in \mathbb{R}^{3\times 3}$ y haciendo el cálculo manual de su traza correspondiente.

2. **(20 puntos)** Para la siguiente matriz:

$$A = s\begin{bmatrix}-1 & 2 & 2\\ 2 & -1 & 2\\ 2 & 2 & 2\end{bmatrix}$$

   defina un valor de s que haga la matriz ortonormal, de forma que $U^T U = I = U\,U^T$.

3. **(20 puntos)** Con las matrices no singulares y por ende invertibles $A, X, Y \in \mathbb{R}^{n\times n}$, suponga que:

$$XA = I_n \qquad AY = I_n$$

   con $I_n \in \mathbb{R}^{n\times n}$ la matriz identidad. Demuestre que $X = Y$.

4. **(20 puntos)** Muestre con un ejemplo numérico que para un vector $\overrightarrow{x} \in \mathbb{R}^n$ y una matriz simétrica $A \in \mathbb{R}^{n\times n}$

$$\left(\vec{x}^T A\, \vec{x}\right)^T = \vec{x}^T A\, \vec{x}$$

   a) Incluya el código en Pytorch que permita corroborar tal igualdad para cualquier matriz simétrica A y vector $\overrightarrow{x} \in \mathbb{R}^n$, generado aleatoriamente. Recuerde que a partir de cualquier matriz cuadrada A generada aleatoriamente puede calcularse una matriz simétrica, haciendo $\frac{1}{2}A + \frac{1}{2}A^T$.

5. **(20 puntos)** Demuestre que la siguiente ecuación matricial:

$$\left\| A\, \vec{x} - \vec{b} \right\|^2 + \|\vec{x}\|$$

   con $\overrightarrow{x} \in \mathbb{R}^n$, $\overrightarrow{b} \in \mathbb{R}^m$ y $A \in \mathbb{R}^{m\times n}$,se puede reescribir como sigue:

$$\vec{x}^T A^T A\, \vec{x} - 2\, \vec{b}^T A\, \vec{x} + \vec{b}^T \vec{b} + \sqrt{\vec{x}^T \vec{x}}$$

6. **(20 puntos)** La Matriz de covarianza

   Para dos variables aleatorias X e Y se la covarianza como el valores esperado de la diferencia de una variable aleatoria y su esperanza (media):

$$\Sigma_{X,Y} = \mathrm{cov}\,(X,Y) = \mathbb{E}\left[(X - \mathbb{E}[X])\,(Y - \mathbb{E}[Y])\right]$$

   y mide la variación conjunta de tales variables aleatorias. Para el caso de contar con arreglos de muestras $h\,[u]$ y $g\,[u]$ para las variables aleatorias X e Y respectivamente, se tiene que la covarianza de tales variables aleatorias está dada por:

$$\Sigma_{X,Y} \cong \frac{1}{N-1} \sum_{u=1}^{N} \left(h\,[u] - \mu_X\right)\left(g\,[u] - \mu_Y\right).$$

   con las medias o esperanzas $\mu_X = \mathbb{E}[X]$ y $\mu_Y = \mathbb{E}[Y]$, componentes del vector medio

$$\overrightarrow{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$$

   La matriz de covarianza para n variables aleatorias $X_1, X_2, \ldots, X_n$ se define como:

$$\Sigma = \begin{bmatrix} \mathbb{E}\left[(X_1 - \mathbb{E}[X_1])\,(X_1 - \mathbb{E}[X_1])\right] & \ldots & \mathbb{E}\left[(X_1 - \mathbb{E}[X_1])\,(X_n - \mathbb{E}[X_n])\right] \\ \vdots & \ddots & \vdots \\ \mathbb{E}\left[(X_n - \mathbb{E}[X_n])\,(X_1 - \mathbb{E}[X_1])\right] & \ldots & \mathbb{E}\left[(X_n - \mathbb{E}[X_n])\,(X_n - \mathbb{E}[X_n])\right] \end{bmatrix},$$

observe que en la diagonal de la matriz Σ (entrada $\Sigma_{i,i}$) se tiene que

$$\mathbb{E}\left[\left(X_i - \mathbb{E}\left[X_i\right]\right)\left(X_i - \mathbb{E}\left[X_i\right]\right)\right] = \sigma_{X_i}^2,$$

por lo que entonces la matriz de covarianza se puede reescribir como:

$$\Sigma = \begin{bmatrix} \sigma_{X_1}^2 & \dots & \mathbb{E}\left[\left(X_1 - \mathbb{E}\left[X_1\right]\right)\left(X_n - \mathbb{E}\left[X_n\right]\right)\right] \\ \vdots & \ddots & \vdots \\ \mathbb{E}\left[\left(X_n - \mathbb{E}\left[X_n\right]\right)\left(X_1 - \mathbb{E}\left[X_1\right]\right)\right] & \dots & \sigma_{X_n}^2 \end{bmatrix}.$$

Además, la matriz de covarianza Σ presenta la propiedad de ser simétrica, puesto que $\mathbb{E}\left[\left(X_i - \mathbb{E}\left[X_i\right]\right)\left(X_j - \mathbb{E}\left[X_j\right]\right)\right] = \mathbb{E}\left[\left(X_j - \mathbb{E}\left[X_j\right]\right)\left(X_i - \mathbb{E}\left[X_i\right]\right)\right] \Rightarrow \Sigma_{X_i,X_j} = \Sigma_{X_j,X_i}$.

Ejemplo

Suponga que se desea encontrar la matriz de covarianza para tres variables aleatorias X_1, X_2 y X_3, para las cuales se han recabado los siguientes arreglos de muestras para $N = 4$ experimentos, respectivamente:

$$\begin{aligned} h_1 &= \begin{bmatrix} 2 & 4 & 6 & 8 \end{bmatrix} \\ h_2 &= \begin{bmatrix} 4 & 8 & 12 & 16 \end{bmatrix} \\ h_3 &= \begin{bmatrix} 12 & 10 & 5 & 9 \end{bmatrix} \end{aligned}$$

En términos de muestras se tienen 4 muestras

$$U = \{\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4\} = \begin{bmatrix} | & | & | & | \\ \vec{u}_1 & \vec{u}_2 & \vec{u}_3 & \vec{u}_4 \\ | & | & | & | \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 & 8 \\ 4 & 8 & 12 & 16 \\ 12 & 10 & 5 & 9 \end{bmatrix}$$

con $u_i \in \mathbb{R}^3$, donde cada dimensión es una variable aleatoria, y $U \in \mathbb{R}^{3\times 4}$.

Observe en estos datos, que la dimensión 1 y 2 son combinación lineal para todas las muestras, por lo que la covarianza de ambas dimensiones debe ser alta, no así la dimensión 1 con la 3 o la 2 con la 3. Además

Se procede entonces a calcular las entradas Σ_{X_1,X_2}, Σ_{X_1,X_3} y Σ_{X_2,X_3}, además de los valores de la diagonal $\sigma_{X_1}^2$, $\sigma_{X_2}^2$ y $\sigma_{X_3}^2$, teniendo en cuenta que $\mu_{X_1} = 5$, $\mu_{X_2} = 10$ y $\mu_{X_3} = 9$, con lo que entonces:

$$\overrightarrow{\mu} = \begin{bmatrix} \mu_{X_1} = 5 \\ \mu_{X_2} = 10 \\ \mu_{X_3} = 9 \end{bmatrix}$$

y haciendo los cálculos respectivos:

$$\begin{aligned}
\Sigma_{X_1,X_2} &= \tfrac{1}{4-1}\left((5-2)(10-4)+(5-4)(10-8)+(5-6)(10-12)+(5-8)(10-16)\right)\\
\Sigma_{X_1,X_3} &= \tfrac{1}{4-1}\left((5-2)(9-12)+(5-4)(9-10)+(5-6)(9-5)+(5-8)(9-9)\right)\\
\Sigma_{X_2,X_3} &= \tfrac{1}{4-1}\left((10-4)(9-12)+(10-8)(9-10)+(10-12)(9-5)+(10-16)(9-9)\right)\\
\sigma^2_{X_1} &= \tfrac{1}{4-1}\left((5-2)^2+(5-4)^2+(5-6)^2+(5-8)^2\right)\\
\sigma^2_{X_2} &= \tfrac{1}{4-1}\left((10-4)^2+(10-8)^2+(10-12)^2+(10-16)^2\right)\\
\sigma^2_{X_3} &= \tfrac{1}{4-1}\left((9-12)^2+(9-10)^2+(9-5)^2+(9-9)^2\right)
\end{aligned}$$

lo cual desarrollado corresponde a:

$$\begin{aligned}
\Sigma_{X_1,X_2} &= \tfrac{1}{3}\left(3\cdot 6+1\cdot 2+-1\cdot -2+-3\cdot -6\right)=\tfrac{40}{3}=13{,}333\\
\Sigma_{X_1,X_3} &= \tfrac{1}{3}\left(3\cdot -3+1\cdot -1+-1\cdot 4+-3\cdot 0\right)=-\tfrac{14}{3}=-4{,}667\\
\Sigma_{X_2,X_3} &= \tfrac{1}{3}\left(6\cdot -3+2\cdot -1+-2\cdot 4+-6\cdot 0\right)=-\tfrac{28}{3}=-9{,}333\\
\sigma^2_{X_1} &= \tfrac{1}{3}\left(9+1+1+9\right)=\tfrac{20}{3}=6{,}667\\
\sigma^2_{X_2} &= \tfrac{1}{3}\left(36+4+4+36\right)=\tfrac{80}{3}=26{,}667\\
\sigma^2_{X_3} &= \tfrac{1}{3}\left(9+1+16+0\right)=\tfrac{14}{3}=8{,}667.
\end{aligned}$$

Por lo que se obtiene la matriz de covarianza:

$$\Sigma=\begin{bmatrix}\frac{20}{3} & \frac{40}{3} & -\frac{14}{3}\\ \frac{40}{3} & \frac{80}{3} & -\frac{28}{3}\\ -\frac{14}{3} & -\frac{28}{3} & \frac{14}{3}\end{bmatrix}=\begin{bmatrix}6{,}667 & 13{,}333 & -4{,}667\\ 13{,}333 & 26{,}666 & -9{,}333\\ -4{,}667 & -9{,}333 & 8{,}667\end{bmatrix}.$$

También la matriz de covarianza se puede escribir, para un conjunto de muestras $X=\{\vec{x}_1,\vec{x}_2,\ldots,\vec{x}_m\}$, con $\vec{x}_i\in\mathbb{R}^{n\times 1}$, como:

$$\Sigma=\frac{1}{m-1}\sum_{i=1}^{m}\left(\vec{x}_i-\vec{\mu}\right)\left(\vec{x}_i-\vec{\mu}\right)^T$$

donde $\vec{\mu}\in\mathbb{R}^{n\times 1}$ es la muestra promedio del conjunto de datos X (donde cada componente es el valor medio de cada dimensión).

a) Escriba la función en Pytorch sin usar estructuras de repetición *for* o *while*, usando únicamente las funciones para cálculo de vector medio, y multiplicaciones matriciales. Documente su uso con el ejemplo descrito en este documento.