Repaso de Álgebra lineal

Reconocimiento de patrones

Gamaliel Moreno

gamalielmch@uaz.edu.mx
http://pds.uaz.edu.mx/

Enero-Julio 2021

Contenido

1 Vectores y matrices
- Defnciones

2 Operaciones matriciales
- Defnciones
- Interpretaciones

3 Gradientes Matriciales
- Gradiente y traza
- Derivación de ecuaciones normales

Escalar

Vectores y matrices
Defınciones
Operaciones matriciales
Defınciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

Un escalar es un número real

$$s \in \mathbb{R}$$

Vector

Vectores y matrices
Defínciones
Operaciones matriciales
Defínciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

Vector de m dimensiones $\boldsymbol{x} \in \mathbb{R}^m$ (m escalares)

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

Vector fila

Vectores y matrices
Definciones
Operaciones matriciales
Definciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

Vector de n dimensiones $\boldsymbol{x} \in \mathbb{R}^n$ (n escalares)

$$\boldsymbol{x}^T = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$$

Matriz

Matriz de $m \times n$ $\boldsymbol{A} \in \mathbb{R}^{m \times n}$

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

- m: files
- n: columnas
- En notación a_{ij} primer subíndice i es la fila y el segundo la columna
- Todas las matrices en $\mathbb{R}^{m \times n}$ conforman un espacio lineal. Esto es $B = \{B_i | B_i \in \mathbb{R}^{m \times n}\}$, entonces $M = \sum_i c_i B_i$

Tensor

Vectores y matrices
Definciones
Operaciones matriciales
Definciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

Generalización de conceptos anteriores

- Escalar $s \in \mathbb{R}$: tensor 0
- Vector $\boldsymbol{v} \in \mathbb{R}^n$: tensor 1
- Matriz $\boldsymbol{M} \in \mathbb{R}^{m \times n}$: tensor 2
- En general $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots n_q}$ es un tensor de orden q

Matriz en vector

Vectores y matrices
Defínciones
Operaciones matriciales
Defínciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

Matriz $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ se descompone en m vectores fila

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \boxed{\boldsymbol{a}_{1,:}{}^{T}} \\ \boxed{\boldsymbol{a}_{2,:}{}^{T}} \\ \vdots \\ \boxed{\boldsymbol{a}_{m,:}{}^{T}} \end{bmatrix}$$

o en vectores columna

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \boxed{\boldsymbol{a}_{:,1}} & \boxed{\boldsymbol{a}_{:,2}} & \vdots & \boxed{\boldsymbol{a}_{:,n}} \end{bmatrix}$$

Vectores como matrices

Vectores y matrices
Definciones
Operaciones matriciales
Definciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

- Observe que todo vector es un tipo particular de matriz
  - Vector fila: matriz de dimensiones $1 \times n$
  - Vector columna: matriz de dimensiones $m \times 1$
- Propiedades de matrices aplicarán a vectores

Matriz transpuesta 1

Si

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

entonces

$$\boldsymbol{A}^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}$$

- En otras palabras, si $\boldsymbol{B} = \boldsymbol{A}^T$ entonces $b_{ij} = a_{ij}$
- $(\boldsymbol{A}^T)^T = \boldsymbol{A}$
- $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$
- $(\boldsymbol{A}+\boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$

Simetría y anti-simetría 1

Matriz simétrica si $\boldsymbol{A} = \boldsymbol{A}^T$

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{21} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

es decir $a_{ij} = a_{ji}$

- La matriz es anti-simétrica si $\boldsymbol{A} = -\boldsymbol{A}^T$
- Para cualquier matriz cuadrada $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ se cumple
  - $\boldsymbol{A} + \boldsymbol{A}^T$ es simétrica
  - $\boldsymbol{A} - \boldsymbol{A}^T$ es anti-simétrica
  - $\boldsymbol{A} = \boldsymbol{A}_e + \boldsymbol{A}_o = \frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^T) + \frac{1}{2}(\boldsymbol{A} - \boldsymbol{A}^T)$
- $\mathbb{S}^n$ es el conjunto de todas las matrices simétricas $n \times n$

Matriz diagonal

Vectores y matrices
Defínciones
Operaciones matriciales
Defínciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

Una matriz es diagonal si todos los elementos son cero excepto aquellos en la diagonal

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{nn} \end{bmatrix}$$

Se utiliza la notación $\boldsymbol{A} = diag(\underline{\boldsymbol{a}})$ y $diag(\boldsymbol{A}) = \underline{\boldsymbol{a}}$ con $\underline{\boldsymbol{a}} = \begin{bmatrix} a_{11} & a_{22} & \cdots & a_{nn} \end{bmatrix}^T$

Matriz identidad

Vectores y matrices
Defínciones
Operaciones matriciales
Defínciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

Es una matriz diagonal cuyos elementos no nulos son iguales a uno

$$\boldsymbol{I} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

- Se cumple $\boldsymbol{AI} = \boldsymbol{A} = \boldsymbol{IA}$ ($\boldsymbol{I}$ es el elemento nuestro del producto matricial)
- $\boldsymbol{I} = diag([1, 1, \ldots, 1]^T)$

Suma de matrices

Vectores y matrices
Defininciones
Operaciones matriciales
Defininciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

El producto $s\boldsymbol{A}$ es otra matriz con todos los componentes escalados

$$s\boldsymbol{A} = s \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} sa_{11} & sa_{12} & \dots & sa_{1n} \\ sa_{21} & sa_{22} & \dots & sa_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ sa_{m1} & sa_{m2} & \dots & sa_{mn} \end{bmatrix}$$

Producto escalar-matriz

Vectores y matrices
Defínciones
Operaciones matriciales
Defínciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

Suma definida para dos matrices de idéntico tamaño:

$$\boldsymbol{c} = \boldsymbol{A} + \boldsymbol{B} \Leftrightarrow c_{ij} = a_{ij} + b_{ij}$$

$$\begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{bmatrix} =$$

$$\begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{bmatrix}$$

Combinación lineal

Vectores y matrices
Defınciones
Operaciones matriciales
Defınciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

Una matriz $\boldsymbol{A}$ es la combinación lineal de un conjunto de m matrices si

$$\boldsymbol{A} = \sum_{i=1}^{m} s_i \boldsymbol{B}_i$$

donde s_i son los coeficientes de la combinación.

- El conjunto $\mathcal{B} = \{\boldsymbol{B}_i | i = 1 \ldots m\}$ engendra un espacio lineal $\mathcal{V}$ compuesto de todas las combinaciones lineales posibles de las matrices $\mathcal{B}$

Producto punto entre vectores

- El producto punto está definido para dos vectores de dimensión n, y es un valor escalar calculado con:

$$\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^T \boldsymbol{y} = x_1 y_1 + x_2 + y_2 + \cdots x_n y_n = \sum_{i=1}^{n} x_i y_i$$

- Interpretación relacionada con proyección entre vectores
- El producto punto es un tipo de producto interno y por tanto se puede denotar también $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$
- Para matrices, el producto interno de Frobenius se define como:

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle_F = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}$$

Ángulo entre dos vectores

Vectores y matrices

Defínciones

Operaciones matriciales

Defínciones

Interpretaciones

Gradientes Matriciales

Gradiente y traza

Derivación de ecuaciones normales

- El ángulo α entre dos vectores $\boldsymbol{x}$ y $\boldsymbol{y}$ está dado por

$$\alpha = arcos\left(\frac{\boldsymbol{x}^T\boldsymbol{y}}{||\boldsymbol{x}||| \cdot ||\boldsymbol{y}||}\right)$$

- La proyección ortogonal de $\boldsymbol{y}$ sobre la dirección de $\boldsymbol{x}$ está dada por

$$y_{\boldsymbol{x}} = \frac{\boldsymbol{x}^T\boldsymbol{y}}{||\boldsymbol{x}||}$$

- Si $\boldsymbol{x}^T\boldsymbol{y} = 0$ ambos vectores son ortogonales
- Si $||\boldsymbol{x}|| = 1$ se dice que $\boldsymbol{x}$ está normalizado
- Vectores son ortonormales si son ortogonales y normalizados

Producto externo

Vectores y matrices
Defninciones
Operaciones matriciales
Defninciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

El producto externo está definido para dos vectores y es una matriz de dimensiones $m \times n$ con m el tamaño del primer vector y n el tamaño del segundo vector

$$\boldsymbol{x}\boldsymbol{y}^T = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}$$

Ejemplo: Uso de operaciones aritméticas 1

Dado el vector columna m-dimensional $\boldsymbol{x}$ ¿con qué operaciones puede expresarse la réplica de ese vector en n columnas de una matriz? ¿con qué operaciones puede expresarse la réplica de un vector fila n-dimensional $\boldsymbol{x}$ en m filas de una matriz?

$$\boldsymbol{A} = \begin{bmatrix} x_1 & x_1 & \dots & x_1 \\ x_2 & x_2 & \dots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_m & x_m & \dots & x_m \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ x_1 & x_2 & \dots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1 & x_2 & \dots & x_n \end{bmatrix}$$

Ejemplo: Uso de operaciones aritméticas 1

Vectores y matrices
Definciones
Operaciones matriciales
Definciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

$$\boldsymbol{A} = [\boldsymbol{x} \quad \boldsymbol{x} \cdots \boldsymbol{x}] = \boldsymbol{x}\boldsymbol{1}^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} x_1 & x_1 & \dots & x_1 \\ x_2 & x_2 & \dots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_m & x_m & \dots & x_m \end{bmatrix}$$

Ejemplo: Uso de operaciones aritméticas 1

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{x}^T \\ \boldsymbol{x}^T \\ \vdots \\ \boldsymbol{x}^T \end{bmatrix} = \mathbf{1}\boldsymbol{x}^T = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \cdots x_n \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ x_1 & x_2 & \dots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1 & x_2 & \dots & x_n \end{bmatrix}$$

Producto matricial

El producto entre una matriz $\boldsymbol{A}$ de dimensiones $m \times n$ por otra matriz de $\boldsymbol{B}$ de dimensión $n \times l$ es la matriz

$$\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B} = \begin{bmatrix} \boldsymbol{a}_{1,:}^{T} \\ \boldsymbol{a}_{2,:}^{T} \\ \vdots \\ \boldsymbol{a}_{m,:}^{T} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_{:,1}^{T} & \boldsymbol{b}_{:,2}^{T} \cdots \boldsymbol{b}_{:,l}^{T} \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_{1,:} \cdot \boldsymbol{b}_{:,1} & \boldsymbol{a}_{1,:} \cdot \boldsymbol{b}_{:,2} & \cdots & \boldsymbol{a}_{1,:} \cdot \boldsymbol{b}_{:,l} \\ \boldsymbol{a}_{2,:} \cdot \boldsymbol{b}_{:,1} & \boldsymbol{a}_{2,:} \cdot \boldsymbol{b}_{:,2} & \cdots & \boldsymbol{a}_{2,:} \cdot \boldsymbol{b}_{:,l} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_{m,:} \cdot \boldsymbol{b}_{:,1} & \boldsymbol{a}_{m,:} \cdot \boldsymbol{b}_{:,2} & \cdots & \boldsymbol{a}_{m,:} \cdot \boldsymbol{b}_{:,l} \end{bmatrix}$$

Propiedades el producto matricial

Vectores y matrices
Defninciones
Operaciones matriciales
Defninciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

- El producto matricial NO es conmutativo

$$\boldsymbol{AB} \neq \boldsymbol{BA}$$

- Si las dimensiones permiten, el producto matricial SÍ es asociativo

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$$

- Si las dimensiones permiten, el producto matricial SÍ es distributivo

$$(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$$

Interpretaciones del producto matricial 1

- Observe primero el producto matriz vector

$$\boldsymbol{c} = \boldsymbol{A}\boldsymbol{b} = \begin{bmatrix} \boldsymbol{a}_{1,:}^T \\ \boldsymbol{a}_{2,:}^T \\ \vdots \\ \boldsymbol{a}_{m,:}^T \end{bmatrix} \boldsymbol{b} = \begin{bmatrix} \boldsymbol{a}_{1,:}^T \boldsymbol{b} \\ \boldsymbol{a}_{2,:}^T \boldsymbol{b} \\ \vdots \\ \boldsymbol{a}_{m,:}^T \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_{1,:} \cdot \boldsymbol{b} \\ \boldsymbol{a}_{2,:} \cdot \boldsymbol{b} \\ \vdots \\ \boldsymbol{a}_{m,:} \cdot \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}$$

- el producto matriz-vector mapea un vector en otro
- En el producto matriz-matriz $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}$, si $\boldsymbol{b}_{:,j}$ es la j-ésima columna de $\boldsymbol{B}$ entonces el patrón anterior se cumple para la j-ésima columna $\boldsymbol{c}_{:,j}$ del resultado.

Producto como combinación lineal de columnas 1

Vectores y matrices
Defninciones
Operaciones matriciales
Defninciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

- Otra forma de ver el producto matriz-vector es como combinación lineal de las columnas de la matriz:

$$\boldsymbol{c} = \boldsymbol{A}\boldsymbol{b} = \begin{bmatrix} \boldsymbol{a}_{:,1} & \boldsymbol{a}_{:,2} & \cdots & \boldsymbol{a}_{:,n} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$= b_1 \boldsymbol{a}_{:,1} + b_2 \boldsymbol{a}_{:,2} + \cdots + b_n \boldsymbol{a}_{:,n}$$

- Observe la similitud con el producto punto
- El espacio engendrado por las columnas de $\boldsymbol{A}$ se denomina espacio columna de $\boldsymbol{A}$
- El espacio columna se conoce también como el alcance columna de $\boldsymbol{A}$ (range)
- La dimensión del alcance es el rango (rank)

Producto vector matriz

- Observe ahora el producto vector-matriz:

$$\boldsymbol{c}^T = \boldsymbol{b}^T\boldsymbol{A} = \boldsymbol{b}^T \begin{bmatrix} \boldsymbol{a}_{:,1} & \boldsymbol{a}_{:,2} & \cdots & \boldsymbol{a}_{:,n} \end{bmatrix}$$
$$= \begin{bmatrix} \boldsymbol{b}\cdot\boldsymbol{a}_{:,1} & \boldsymbol{b}\cdot\boldsymbol{a}_{:,2} & \cdots & \boldsymbol{b}\cdot\boldsymbol{a}_{:,n} \end{bmatrix}$$

- En el producto matriz-matriz $\boldsymbol{AB}$, si $\boldsymbol{a}_{j,:}$ es la j-ésima fila de $\boldsymbol{A}$ entonces el patrón anterior se cumple para la j-ésima fila del resultado

Producto como combinación lineal de filas 1

Vectores y matrices
Defenciones
Operaciones matriciales
Defenciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

- Otra forma de ver el producto vector-matriz es como combinación lineal de los vectores fila

$$\boldsymbol{c}^T = \boldsymbol{b}^T \boldsymbol{A} = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_{1,:}^T \\ \boldsymbol{a}_{2,:}^T \\ \cdots \\ \boldsymbol{a}_{n,:}^T \end{bmatrix} =$$

$$= b_1 \boldsymbol{a}_{1,:}^T + b_2 \boldsymbol{a}_{2,:}^T + \cdots + b_n \boldsymbol{a}_{n,:}^T$$

- Observe la similitud con el producto punto.

Producto como combinación lineal de filas 2

Vectores y matrices
Defininciones
Operaciones matriciales
Defininciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

Theorem

Dos posibles interpretaciones Lo anterior implica que el producto de dos matrices puede interpretarse como combinaciones lineales de las columnas de la primera matriz o de las filas de la segunda matriz

Producto de Hadamard

Vectores y matrices
Defínciones
Operaciones matriciales
Defínciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

- Producto de Hadamard, de Schur, o elemento por elemento
- Denotado por $\boldsymbol{A} \odot \boldsymbol{B}$
- aplica para dos matrices del mismo tamaño

$$\boldsymbol{A} \odot \boldsymbol{B} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \odot \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \dots & a_{1n}b_{1n} \\ a_{21}b_{21} & a_{22}b_{22} & \dots & a_{2n}b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{m1} & a_{m2}b_{m2} & \dots & a_{mn}b_{mn} \end{bmatrix}$$

Derivadas con matrices y la traza

- Sea $\boldsymbol{A}$ una matriz de tamano $m \times n$ y $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ una función que mapea matrices como $\boldsymbol{A}$ avalores reales
- El gradiente de la función escalar f es la matriz

$$\nabla_{\boldsymbol{A}} f(\boldsymbol{A}) = \begin{bmatrix} \frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} & \cdots & \frac{\partial f}{\partial a_{1n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial a_{m1}} & \frac{\partial f}{\partial a_{m2}} & \cdots & \frac{\partial f}{\partial a_{mn}} \end{bmatrix} \boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- La traza de la matriz cuadrada $\boldsymbol{A}$ de tamaño $n \times n$ es

$$tr\boldsymbol{A} = \sum_{i=1}^{n} a_{ii}$$

- Un escalar s (matriz 1×1) tiene $trs = s$

Vectores y matrices
Defínciones
Operaciones matriciales
Defínciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

Propiedades de la traza

- $tr(\boldsymbol{A}+\boldsymbol{B}) = tr\boldsymbol{A} + tr\boldsymbol{B}$
- $tr(c\boldsymbol{A}) = ctr\boldsymbol{A}$
- $tr\boldsymbol{A} = tr\boldsymbol{A}^T$
- Si $\boldsymbol{AB}$ es cuadrada entonces $tr\boldsymbol{AB} = \boldsymbol{BA}$

$$tr\boldsymbol{AB} = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}b_{ij} = \sum_{j=1}^{m}\sum_{i=1}^{n} b_{ij}a_{ij} = tr\boldsymbol{BA}$$

- Corolario
  - $tr\boldsymbol{ABC} = tr\boldsymbol{CAB} = tr\boldsymbol{BCA}$
  - $tr\boldsymbol{ABCD} = tr\boldsymbol{DABC} = tr\boldsymbol{CDAB} = tr\boldsymbol{BCDA}$

Combinando traza con derivadas

1. $\nabla_{\boldsymbol{A}} tr \boldsymbol{AB} = \boldsymbol{B}^T$
2. $\nabla_{\boldsymbol{A}^T} f(\boldsymbol{A}) = (\nabla_{\boldsymbol{A}} f(\boldsymbol{A}))^T$
3. $\nabla_{\boldsymbol{A}} tr \boldsymbol{ABA}^T\boldsymbol{C} = \boldsymbol{CAB} + \boldsymbol{C}^T\boldsymbol{AB}^T$
4. $\nabla_{\boldsymbol{A}} |\boldsymbol{A}| = |\boldsymbol{A}|(\boldsymbol{A}^{-1})^T$

Combinando 2 y 3

- $\nabla_{\boldsymbol{A}^T} tr \boldsymbol{ABA}^T\boldsymbol{C} = \boldsymbol{B}^T\boldsymbol{A}^T\boldsymbol{C}^T + \boldsymbol{BA}^T\boldsymbol{C}$

Ecuaciones normales en regresión lineal

Vectores y matrices
Defineiones
Operaciones matriciales
Defineiones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

- Caso particular de regresión lineal se resuelve con ecuaciones normales
- Usaremos lo anterior para derivación directa

Vectores y matrices
Defínciones
Operaciones matriciales
Defínciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

Replanteando mínimos cuadrados

- Buscamos $\boldsymbol{\theta}$ que minimiza $J(\boldsymbol{\theta})$
- Reescribiendo $J(\boldsymbol{\theta})$ en forma matricial

Sea $\boldsymbol{X}$ la matriz de diseño de tamaño $m \times n+1$

$$\boldsymbol{X} = \begin{bmatrix} x^{(1)T} \\ x^{(2)T} \\ \vdots \\ x^{(m)T} \end{bmatrix}$$

sea $\boldsymbol{y}$ el vector de valores objetivo:

$$\boldsymbol{y} = \begin{bmatrix} y^{(1)} & y^{(2)} & \dots & y^{(m)} \end{bmatrix}$$

Replanteando mínimos cuadrados

Vectores y matrices
Defínciones
Operaciones matriciales
Defínciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

Puesto que

$$\boldsymbol{X}\theta - \boldsymbol{y} = \begin{bmatrix} x^{(1)T}\boldsymbol{\theta} \\ x^{(2)T}\boldsymbol{\theta} \\ \vdots \\ x^{(m)T}\boldsymbol{\theta} \end{bmatrix} - \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \cdots \\ y^{(m)} \end{bmatrix} = \begin{bmatrix} x^{(1)T}\boldsymbol{\theta} - y^{(1)} \\ x^{(2)T}\boldsymbol{\theta} - y^{(2)} \\ \vdots \\ x^{(m)T}\boldsymbol{\theta} - y^{(m)} \end{bmatrix}$$

Entonces usando $\boldsymbol{v}^T\boldsymbol{v} = ||\boldsymbol{v}||^2 = \sum_i v_i^2$

$$J(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{m}(x^{(i)T}\boldsymbol{\theta} - y^{(i)})^2 = \frac{1}{2}||\boldsymbol{X\theta} - \boldsymbol{y}||^2$$
$$= \frac{1}{2}(\boldsymbol{X\theta} - \boldsymbol{y})^T(\boldsymbol{X\theta} - \boldsymbol{y})$$

$\frac{1}{2}(\boldsymbol{X\theta} - \boldsymbol{y})^T(\boldsymbol{X\theta} - \boldsymbol{y})$ esto es un escalar por lo tanto
$\frac{1}{2}(\boldsymbol{X\theta} - \boldsymbol{y})^T(\boldsymbol{X\theta} - \boldsymbol{y}) = \frac{1}{2}tr(\boldsymbol{X\theta} - \boldsymbol{y})^T(\boldsymbol{X\theta} - \boldsymbol{y})$

Replanteando mínimos cuadrados

Vectores y matrices
Defínciones
Operaciones matriciales
Defínciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

- El mínimo se encuentra buscando $\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) = 0$

$$\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \frac{1}{2} (\boldsymbol{X\theta} - \boldsymbol{y})^T (\boldsymbol{X\theta} - \boldsymbol{y}) \doteq 0$$

$$= \frac{1}{2} \nabla_{\boldsymbol{\theta}} (\boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{X\theta} - \boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{y} - \boldsymbol{y}^T \boldsymbol{X\theta} + \boldsymbol{y}\boldsymbol{y}^T)$$

y puesto que el término entre paréntesis es un escalar real

$$= \frac{1}{2} \nabla_{\boldsymbol{\theta}} tr(\boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{X\theta} - \boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{y} - \boldsymbol{y}^T \boldsymbol{X\theta} + \boldsymbol{y}\boldsymbol{y}^T)$$

$$= \frac{1}{2} \nabla_{\boldsymbol{\theta}} (tr \boldsymbol{\theta}^T \boldsymbol{X}^T \boldsymbol{X\theta} - 2 tr \boldsymbol{y}^T \boldsymbol{X\theta})$$

Replanteando mínimos cuadrados

Vectores y matrices
Defínciones
Operaciones matriciales
Defínciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

usando $\nabla_{\boldsymbol{A}^T} tr\boldsymbol{ABA}^T\boldsymbol{C} = \boldsymbol{B}^T\boldsymbol{A}^T\boldsymbol{C}^T + \boldsymbol{BA}^T\boldsymbol{C}$ con $\boldsymbol{A} = \boldsymbol{\theta}^T$, $\boldsymbol{B} = \boldsymbol{X}^t\boldsymbol{X}$ y $\boldsymbol{C} = I$

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) &= \frac{1}{2}\nabla_{\boldsymbol{\theta}}(tr\boldsymbol{\theta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta} - 2tr\boldsymbol{y}^T\boldsymbol{X}\boldsymbol{\theta}) \\
&= \frac{1}{2}(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta} - 2\boldsymbol{X}^T\boldsymbol{y}) \\
&= \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{X}^T\boldsymbol{y} \\
&= 0
\end{aligned}$$

Con lo que se obtienen las ecuaciones normales

$$\begin{aligned}
\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta} &= \boldsymbol{X}^T\boldsymbol{y} \\
\boldsymbol{\theta} &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}
\end{aligned}$$

- $\boldsymbol{\theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ es la solución cerrada

Replanteando mínimos cuadrados

Vectores y matrices
Defínciones
Operaciones matriciales
Defínciones
Interpretaciones
Gradientes Matriciales
Gradiente y traza
Derivación de ecuaciones normales

- $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ se conoce como la matriz seudoinversa de $\boldsymbol{X}$o seudoinversa de Moore-Penrose denotada con $\boldsymbol{X}^\dagger$:

$$\boldsymbol{\theta} = \boldsymbol{X}^\dagger \boldsymbol{y}$$