Tarea #1

1. Murphy Cap 1 : Matriz, determinantes, propiedades básicas derivadas gradiente de matrices

* **Matriz:** Arreglo rectangular de números con filas y columnas, tal que $A = (a_{ij})$ con a_{ij} el elemento en la fila i, columna j. Si $A \in \mathbb{R}^{m \times n}$ tiene m filas y n columnas.

Como datos: cada fila es una muestra y cada columna una característica

$$\mathbb{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{bmatrix} \longrightarrow x_n \in \mathbb{R}^p$$

(muestra)

$\downarrow$

$\xi \in \mathbb{R}^N$ (características)

→ **Operaciones básicas:**

* Suma: $(A+B)_{ij} = a_{ij} + b_{ij}$
* Producto por escalar: $(\alpha A)_{ij} = \alpha a_{ij}$
* Producto matricial: $(AB)_{ik} = \sum_j a_{ij} b_{jk}$ (no conmutativo en general)
* Transpuesta: A^T
* Identidad I_n : $I_n x = x$
* Inversa (si existe): A^{-1} tal que $A^{-1}A = I$ (para matrices cuadradas invertibles)
* Simétrica $A = A^T$
* Ortogonal : $Q^T Q = I$

→ **Rango, espacio nulo y columnas**

* El rango (rank) es el número de columnas linealmente independientes
* El espacio columna (column space) es el subespacio generado por las columnas
* El espacio nulo (nullspace) son los vectores x tales que $Ax = 0$

* **Determinantes:** si A es cuadrada $(n \times n)$, su determinante $\det(A)$ nos dice varias cosas:

→ **Factor de escala de volumen:** bajo la transformación lineal $x \mapsto Ax$, un conjunto en $\mathbb{R}^n$ ve su volumen multiplicado por $|\det(A)|$

→ **Invertibilidad:** $\det(A) = 0 \longleftrightarrow A$ no es invertible

→ **Signo:** el signo de $\det(A)$ indica si la orientación se preserva o invierte

* FORMULAS

→ PARA 2x2 : $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $\det(A) = ad - bc$

→ PARA TRIANGULAR (SUPERIOR O INFERIOR), $\det(A)$ ES EL PRODUCTO DE LAS ENTRADAS DIAGONALES.

→ PRODUCTO: $\det(AB) = \det(A)\det(B)$

→ TRANSPOSICIÓN: $\det(A^T) = \det(A)$

→ SI P ES MATRIZ DE PERMUTACIÓN, $\det(P) = \pm 1$ SEGÚN PARIDAD DE LA PERMUTACIÓN

* RELACIÓN CON EIGENVALORES SI A TIENE EIGENVALORES $\lambda_1, \ldots, \lambda_n$ (CONTANDO MULTIPLICIDADES), ENTONCES ENTONCES $\det(A) = \prod_i \lambda_i$. ESTO JUNTA DETERMINANTE Y COMPORTAMIENTO DINÁMICO DE A

* CÁLCULO DEL DETERMINANTE (JACOBI / FÓRMULA PRÁCTICA) PARA DERIVADAS Y ANÁLISIS ES ÚTIL LA FÓRMULA DE JACOBI (O SU DERIVADA):

$$d(\det A) = \det(A)\, tr(A^{-1} dA),$$

EN DONDE INMEDIATAMENTE:

$$d(\log \det A) = tr(A^{-1} dA)$$

ESA IDENTIDAD ES CENTRAL PARA DERIVAR LOG-VEROSIMILITUD DE GAUSSIANAS MULTIVARIADAS

* DERIVADAS Y GRADIENTES: EN ML SE TRABAJA CON FUNCIONES DE VARIAS FORMAS: ESCALARES, VECTORES, MATRICES, ESTAS SON LAS CONVENCIONES QUE SE USAN:

* PARA $f: \mathbb{R}^n \to \mathbb{R}$, EL GRADIENTE $\nabla f(x)$ ES UN VECTOR COLUMNA DE TAMAÑO n CON $(\nabla f)_i = \frac{\partial f}{\partial x_i}$

* PARA $g: \mathbb{R}^n \to \mathbb{R}^m$ DEFINIMOS LA JACOBIANA $J \in \mathbb{R}^{m \times n}$ CON $J_{ij} = \frac{\partial g_i}{\partial x_j}$

* PARA $f: \mathbb{R}^{m \times n} \to \mathbb{R}$ LA DERIVADA $\frac{\partial f}{\partial A}$ LA INTERPRETAMOS COMO UNA MATRIZ $m \times n$ CON $\left[\frac{\partial f}{\partial A}\right] = \frac{\partial f}{\partial a_{ij}}$

EN NOTACIÓN DIFERENCIAL:

$$df = tr\left(\left(\frac{\partial f}{\partial A}\right)^T dA\right)$$

ESTO RELACIONA EL DIFERENCIAL ESCALAR df CON EL "PRODUCTO INTERNO" MATRICIAL $tr(X^T Y)$. A PARTIR DE UNA EXPRESIÓN PARA df EN TÉRMINOS DE dA PODEMOS LEER $\partial f / \partial A$

→ PROPIEDADES BÁSICAS

1) DERIVADA DE $y = Ax$ RESPECTO A x

$y \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$

$$\frac{\partial (Ax)}{\partial x} = A$$ → Ax ES LINEAL EN x, CADA COMPONENTE $y_i = \sum_j a_{ij} x_j$, ASÍ $\partial y_i / \partial x_j = a_{ij}$. LA JACOBIANA J ES EXACTAMENTE A.

2) Derivada de la forma cuadrática $f(x) = x^T A x$ respecto a x

Sea $A \in \mathbb{R}^{m \times n}$, entonces:

$$\nabla_x (x^T A x) = (A + A^T) x$$

Derivación por diferencial:

$$f(x) = x^T A x \quad \rightarrow \quad df = (dx)^T A x + x^T A dx$$

El primer término es $(dx)^T Ax$, que es escalar y equivalente a su transpuesta $x^T A^T dx$, sumando,

$$df = x^T (A^T + A)\, dx$$

Por la definición de gradiente $df = (\nabla f)^T dx$, así $\nabla f = (A + A^T)x$

Caso simétrico: si $A = A^T$, entonces $\nabla f = 2Ax$

Derivada respecto a A (manteniendo x fijo):

$$\frac{\partial (x^T A x)}{\partial A} = x x^T$$

Porque $x^T A x = \sum_{ij} a_{ij}\, x_i\, x_j$ y $\partial / \partial a_{ij} = x_i x_j$

3) Derivada de la traza $tr(AB)$

Para $f(A) = tr(AB)$ (B fijo),

$$\frac{\partial tr(AB)}{\partial A} = B^T$$

Esto $tr(AB) = \sum_{ij} a_{ij} b_{ji}$ así $\partial / \partial a_{ij} = b_{ji} = (B^T)_{ij}$

De forma dual, $\partial\, tr(AB) / \partial B = A^T$

4) Derivada del log-determinante

Si A es cuadrada e invertible,

$$d(\log \det A) = tr(A^{-1} dA)$$

De aquí se extrae la derivada (matriz gradiente):

$$\frac{\partial \log \det A}{\partial A} = (A^{-1})^T$$

5) Derivada del determinante

Aplicando lo anterior

$$d(\det A) = \det(A)\, tr(A^{-1} dA)$$

De modo que la matriz gradiente es:

$$\frac{\partial \det A}{\partial A} = \det(A)(A^{-1})^T$$

Esto coincide con la adjunta: $\operatorname{adj}(A) = \det(A)A^{-1}$ y $\frac{\partial \det A}{\partial a_{ij}} = (\operatorname{adj}(A))_{ji}$

6) Derivada de la inversa

Si $Y = A^{-1}$, entonces:

$$dY = -A^{-1}(dA)A^{-1}$$

Derivación rápida: diferenciar $I = AA^{-1}$ y reordenar.

En términos de vectores (con vec y producto Kronecker):

$$\frac{\partial \operatorname{vec}(A^{-1})}{\partial \operatorname{Vec}(A)} = -(A^{-T} \otimes A^{-1})$$

7) Identidad útil

Se usa para expresar derivadas matriciales usando el operador vec y productos Kronecker

$$\operatorname{Vec}(AXB) = (B^T \otimes A)\operatorname{Vec}(X)$$