3b) Tensores y operaciones sobre tensores

- Nuestros parámetros (y el grafo computacional) están compuestos por "arreglos multidimensionales": matrices, vectores o más generalmente "tensores"

Tensor (para nosotros): simplemente una generalización de las matrices a más de dos dimensiones

¿Por qué nos interesa?

1) generalmente usamos "paquetes" (batch) de ejemplos para entrenar a una red

1 ejemplo x

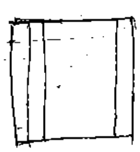
F_1	F_2	F_3	...	F_n

B ejemplos

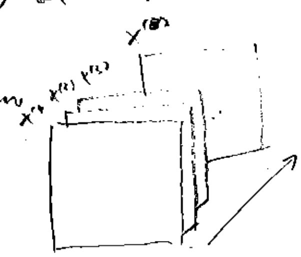
$x^{(1)}$	$f_1^{(1)}$	$f_2^{(1)}$		$f_n^{(1)}$
$x^{(2)}$	$f_1^{(2)}$	$f_2^{(2)}$		$f_n^{(2)}$
$x^{(B)}$	$f_1^{(B)}$	$f_2^{(B)}$		$f_n^{(B)}$

2) muchos ejemplos muchas veces son (naturalmente) de más de una dimensión:

Ej 1) Imagen como mapa de bits

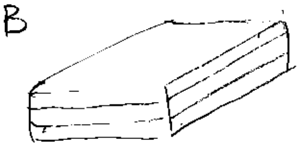

(m, n)

batch de imágenes

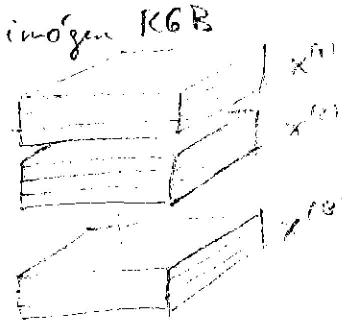


(B, m, n)

Ej 2) Imagen RGB


(m, n, 3)

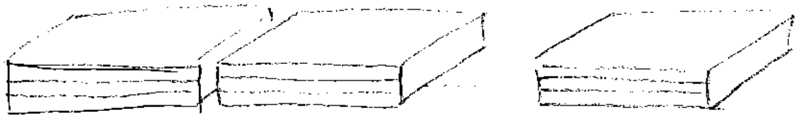
batch de imágenes RGB



(B, m, n, 3)

Ej 3) Video en RGB

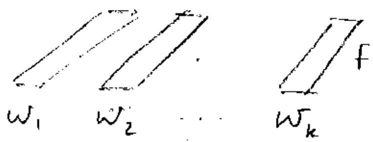
Ej3) Video en RGB



$(m, n, 3, t)$

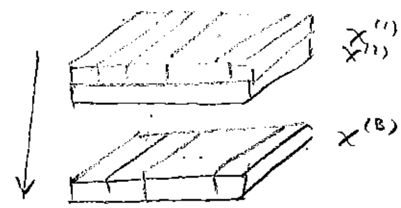
Batch de videos $(B, m, n, 3, t)$

Ej4) Texto (por palabra)



(k, f)

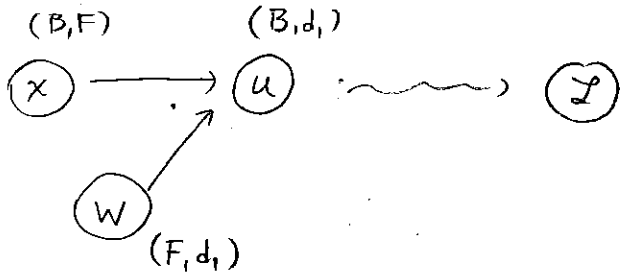
Batch de texto



(B, k, f)

3) En el algoritmo de Back Propagation necesitamos derivar estos objetos.

Ej:



si fueran escalares haríamos simplemente $\frac{\partial \mathcal{L}}{\partial W} = \frac{\partial \mathcal{L}}{\partial u} \cdot \frac{\partial u}{\partial W}$

¿qué es esto?

¿qué es esto?

Operaciones básicas entre tensores:

primero tensor T → $\mathrm{rank}(T)$ = cantidad de dimensiones

supongamos T con k dimensiones, lo denotamos por $T_{(d_1, d_2, \dots, d_k)}$

el elemento en la posición $i_1, i_2, \dots, i_k$ es simplemente $T_{i_1 i_2 \dots i_k}$

¿cuáles son filas, cuáles columnas? Realmente no nos importa!!!

Los podemos ordenar como queramos (aunque en el computador tendrán un orden (numpy.array, torch.tensor)

slice (pedazo): $T_{(n_1, n_2, n_3, n_4)}$ $\quad T_{:1:3}$ ← tensor de 2 dimensiones

Operaciones Básicas:

A, B tensores $A(n_1, n_2, \dots, n_k)$ $B(m_1, m_2, \dots, m_l)$

Aquí!! $+$ y $*$ suma y mult. punto a punto

1) $A \otimes B = X \implies X_{i_1 i_2 \dots i_k j_1 j_2 \dots j_l} = A_{i_1 i_2 \dots i_k} B_{j_1 j_2 \dots j_l}$

2) $A^{T(a,b)}$ simplemente permutar las dimensiones a y b

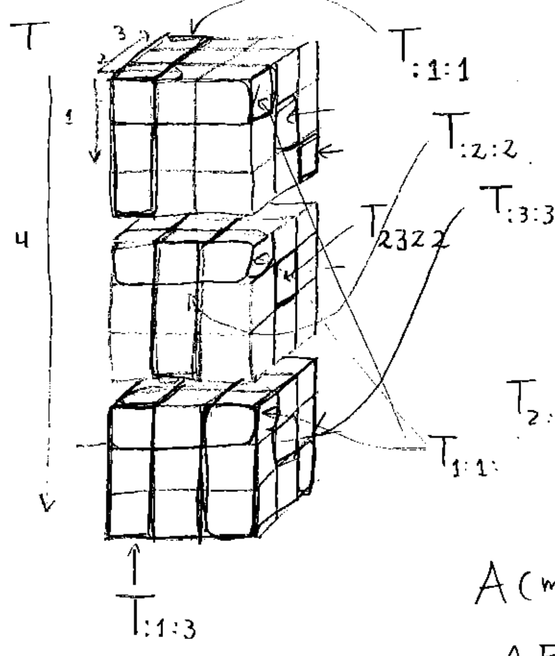
$$A^{T(a,b)}_{i_1 i_2 \dots i_k} = A_{i_1 \dots i_{a-1} i_b i_{a+1} \dots i_{b-1} i_a i_{b+1} \dots i_k} \quad \left(t_{a,b}(A) \right)$$

3) contracción: unir las diagonales de dos dimensiones del mismo tamaño

$A(m_1, m_2, m_3, m_4)$ y $m_2 = m_3 = m$ $\quad A^{C(2,3)}_{ij} = \sum_{k=1}^{m} A_{ikkj} \quad \left(cont_{23}(A) \right)$

Note que una contracción hace perder dos dimensiones

(Esto puede ser la operación más rara)



$$cont_{24}(T) = T_{:1:1} + T_{:2:2} + T_{:3:3}$$

$$cont_{24}(T)_{ij} = \sum_{k=1}^{3} T_{ikjk}$$

$$cont_{13}(T) = T_{1:1:} + T_{2:2:} + T_{3:3:}$$

$$cont_{13}(T)_{ij} = \sum_{k=1}^{3} T_{kikj}$$

Ej): multiplicación de matrices (producto interno)

$A(m,n)$ $B(n,p)$

$$A \cdot B = cont_{23}(A \otimes B)$$

$AB = T \quad T_{ij} = cont_{23}(A \otimes B)_{ij} = \sum_k (A \otimes B)_{ikkj} = \sum_k A_{ik} B_{kj}$

Ej 2): $A(m,n,p)$ $B(n,p,r)$

$A \cdot B = cont_{24,35}(A \otimes B) = \sum_l \sum_k A_{ikl} B_{klj}$ ← muchas veces tendrán que hacer este tipo de cálculos cuando hagáis back prop a mano.

Notación de Einstein: si un índice libre se repite en dos lugares distintos de una expresión $\Rightarrow$ implico una sumatoria sobre los posibles valores de los índices

- $A(m,n)$ $B(n,p)$ $\longrightarrow$ $(AB)_{ij} = A_{ik}B_{kj}$ $\longleftarrow$ suma implícita
- $A(m,n,p)$ $B(n,p,r)$ $\longrightarrow$ $T_{ij} = A_{ikl}B_{klj}$ $\longleftarrow$ suma implícita

$\boxed{A(n_1,\dots,n_k)\ B(n_1,\dots,n_k) \quad (A+B)_{i_1\dots i_k} = A_{i_1\dots i_k} + B_{i_1\dots i_k} \quad (A*B)_{i_1\dots i_k} = A_{i_1\dots i_k}B_{i_1\dots i_k}}$ $\longleftarrow$ componente a componente

★ nos simplifica considerablemente la vida !!!

Propiedades

$$cont_{ab}(A) \otimes cont_{cd}(B) = cont_{ab}(cont_{c'd'}(A \otimes B))$$

con $a,b \leq rank(A)$ $\quad c,d \leq rank(B)$ y $\quad c' = c + rank(A)$, $d' = d + rank(A)$

similar para trasponer

asociatividad:

$A \otimes B \otimes C$

$A \cdot B \otimes C$

$A \otimes B \cdot C$

$$\begin{aligned}(A \cdot B) \otimes C &= cont_{ab}(A \otimes B) \otimes C \\ &= cont_{ab}(A \otimes B \otimes C) \\ &= cont_{ab}(A \otimes T) \quad \text{con } T = B \otimes C \\ &= A \cdot T = A \cdot (B \otimes C)\end{aligned}$$

Derivadas de Tensores

$A(n_1,n_2,\dots,n_k)$

$B(m_1,m_2,\dots,m_l)$

$\dfrac{\partial A}{\partial B}(n_1,n_2,\dots,n_k,m_1,m_2,\dots,m_l)$ tq

$$\frac{\partial A}{\partial B}_{i_1\dots i_k j_1\dots j_l} = \frac{\partial A_{i_1\dots i_k}}{\partial B_{j_1\dots j_l}}$$

Propiedades:

$$\frac{\partial\, cont_{ab}(A)}{\partial B} = cont_{ab}\left(\frac{\partial A}{\partial B}\right)$$

$$\frac{\partial\, A \otimes B}{\partial C} = A \otimes \frac{\partial B}{\partial C} + \left(\frac{\partial A}{\partial C} \otimes B\right)^{T(\text{apropiada})}$$

para dejar los índices de C después de los de B

$$\frac{\partial\, A \cdot B}{\partial C} = A \cdot \frac{\partial B}{\partial C} + \left(\left(\frac{\partial A}{\partial C}\right)^{T(\text{apropiada})} \cdot B\right)^{T(\text{apropiada})}$$

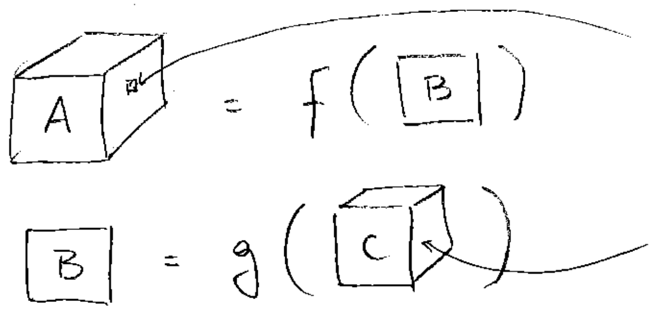
para dejar los índices en orden $A \cdot B \otimes C$

para dejar los índices de A junto a los de B

Lo más complicado es la regla de la cadena para funciones tensoriales

A, B, C tensores y $\begin{array}{l} A = f(B) \\ B = g(C) \end{array}$ ¿ $\dfrac{\partial A}{\partial C}$?

para ejemplificar, supongamos $\mathrm{rank}(A) = 3$, $\mathrm{rank}(B) = 2$, $\mathrm{rank}(C) = 3$



$A_{i_1 i_2 i_3}$

$C_{j_1 j_2 j_3}$

¿ $\dfrac{\partial A_{i_1 i_2 i_3}}{\partial C_{j_1 j_2 j_3}}$?

$$A_{i_1 i_2 i_3} = f(B)_{i_1 i_2 i_3} = f\begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & B_{22} & \cdots & \vdots \\ B_{n1} & \cdots & & B_{mn} \end{pmatrix}_{i_1 i_2 i_3}$$

$B_{kl} = g(C)_{kl}$ $\longleftarrow$ depende de $C_{j_1 j_2 j_3}$

$$\frac{\partial A_{i_1 i_2 i_3}}{\partial C_{j_1 j_2 j_3}} = \frac{\partial A_{i_1 i_2 i_3}}{\partial B_{11}} \cdot \frac{\partial B_{11}}{\partial C_{j_1 j_2 j_3}} + \frac{\partial A_{i_1 i_2 i_3}}{\partial B_{12}} \cdot \frac{\partial B_{12}}{\partial C_{j_1 j_2 j_3}} + \quad + \frac{\partial A_{i_1 i_2 i_3}}{\partial B_{mn}} \frac{\partial B_{mn}}{\partial C_{j_1 j_2 j_3}}$$

$$\overset{\text{ein}}{=} \frac{\partial A_{i_1 i_2 i_3}}{\partial B_{kl}} \frac{\partial B_{kl}}{\partial C_{j_1 j_2 j_3}}$$ $\longleftarrow$ suma implícita de Einstein

$$\frac{\partial A}{\partial C} = \frac{\partial A}{\partial B} : \frac{\partial B}{\partial C}$$ contracción respecto de todos los índices de B !!! (producto interno generalizado)

---

ya tenemos todo lo necesario para hacer a mano y completar un cálculo de Back Propagation para una red FF

Broadcasting, sea $m, n \neq 1$

si tengo dos como $A(m,n)$ $B(n)$

broadcasting $\longrightarrow$ $A + B \equiv A + B'$ $\quad \underbrace{B'(m,n) \text{ tq } B'_{ij} = B_j \ \forall i}$

o sea B se copiará m veces en la dimensión que le falta

si tengo dos como $A(m,n)$ $B(m)$ $(m \neq n)$

$A + B$ falla !!!

$A(m,n)$ $B(1,n)$ $\quad A + B \equiv A + B'$ $\quad B'(m,n)$ tq $B'_{ij} = B_{1j} \ \forall i$

$A(m,n)$ $B(m,1)$ $\quad A + B \equiv A + B'$ $\quad B'(m,n)$ tq $B'_{ij} = B_{i1} \ \forall j$

broadcasting

En general: para $A(n_1, n_2, \ldots, n_k)$ $B(m_1, m_2, \ldots, m_\ell)$

la operación $A \circ B$ se puede aplicar $\iff$

$\forall i$ tq $0 \leq i < \min(k, \ell)$:

1) $n_{k-i} = m_{\ell-i}$, o
2) $n_{k-i} \neq m_{\ell-i}$ y $n_{k-i} = 1$, o
3) $n_{k-i} \neq m_{\ell-i}$ y $m_{\ell-i} = 1$

} En cada uno de estos casos el tensor donde la dimensión de tamaño 1 se copia varias veces