## BACKPROPAGATION ILLUSTRATED

INPUT

Mxno

D7046E - LTU.SE NEURAL NETWORKS AND LEARNING MACHINES

· MATRIX CALCULUS RULE

$$\frac{\partial(u \cdot v)}{\partial x} = u^{T} \frac{\partial v}{\partial x} + v^{T} \frac{\partial u}{\partial x}$$

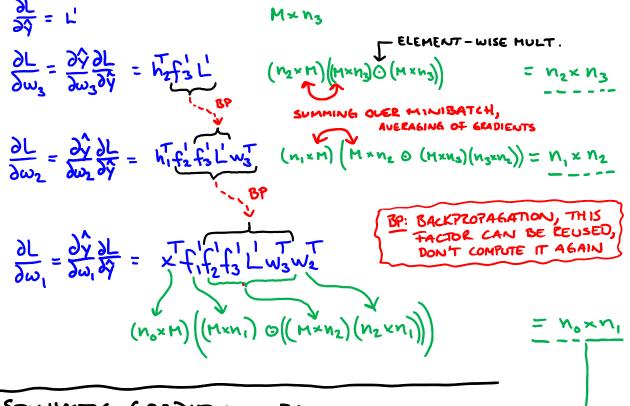
- . WEIGHT MATRICES
- w;
- · MINIBATCH OF SIZE M

. (HIPDEN) STATES

• (Ac	(NOTAUIT.	FUNCTIONS	t
-------	-----------	-----------	---

DIMENSION	NETWORK	RELATION	DERIVATIVE
M×n3	Loss L	L(9-4)	NOTATION: 13
M×n <sub>3</sub>	OUTRUT 🗘	> = f3(µ2m3)	$\frac{\partial \hat{y}}{\partial \omega_3} = h_2^{\top} f_3(h_2 w_3)$
n <sub>z</sub> ×n <sub>s</sub> M×n <sub>2</sub>	HIODEN L	$h_2 = f_2(h_1 w_2)$	$\frac{\partial \hat{y}}{\partial w_2} = \frac{\partial h_2}{\partial w_2} \frac{\partial \hat{y}}{\partial h_2} = h_1 f_2 f_3 w_3$
n <sub>1</sub> ×n <sub>2</sub> M×n <sub>1</sub>	HIPPEN h	$h_i = f_i (x w_i)$	$\frac{\partial \hat{y}}{\partial \omega_1} = \frac{\partial h_1}{\partial h_2} \frac{\partial h_2}{\partial h_1} \frac{\partial \hat{y}}{\partial h_2} = \times f_1 f_2 f_3 w_3 w_2$
n <sub>o</sub> x n <sub>l</sub>	<u>↑</u> w₁	•	gmi gmi gyi gys

## DEZIVATIVES OF LOSS WITH RESPECT TO THE WEIGHTS



STOCHASTIC GRADIENT DESCENT

Wi - Wi- J. DL CONSISTENT DIMENSIONS STEP SIZE (LEARNING RATE)

"STOCHASTIC BECAUSE THE TRUE GRADIENT UNKNOWN: THE GADIENT CALCULATED HERE IS VALID FOR THE DATA IN THE MINIBATCH