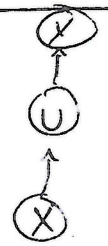
10-601 Nueral Networks

Chain Rule



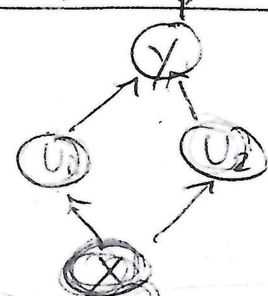
$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}$$

$y = f(u)$
$u = g(x)$

$$\frac{dy}{dx} = f'(g(x))\, g'(x)$$

---

comp. graph →



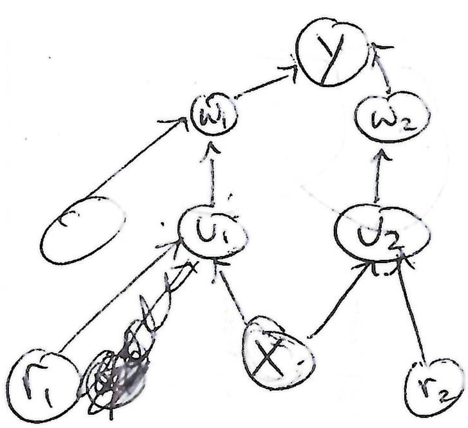
$y = f(u_1, u_2)$
$u_1 = g_1(x)$
$u_2 = g_2(x)$

$$\frac{dy}{dy} = 1$$

$$\frac{dy}{du_1} = \ldots$$

$$\frac{dy}{du_2} =$$

$$\frac{dy}{dx} = \frac{dy}{du_1}\frac{du_1}{dx} + \frac{dy}{du_2}\frac{du_2}{dx}$$



$$\frac{dy}{dx} = \frac{dy}{du_1} \cdot \frac{du_1}{dx} + \frac{dy}{du_2} \cdot \frac{du_2}{dx}$$

$$\frac{dy}{du_1} = \frac{dy}{dw_1} \cdot \frac{dw_1}{du_1}$$

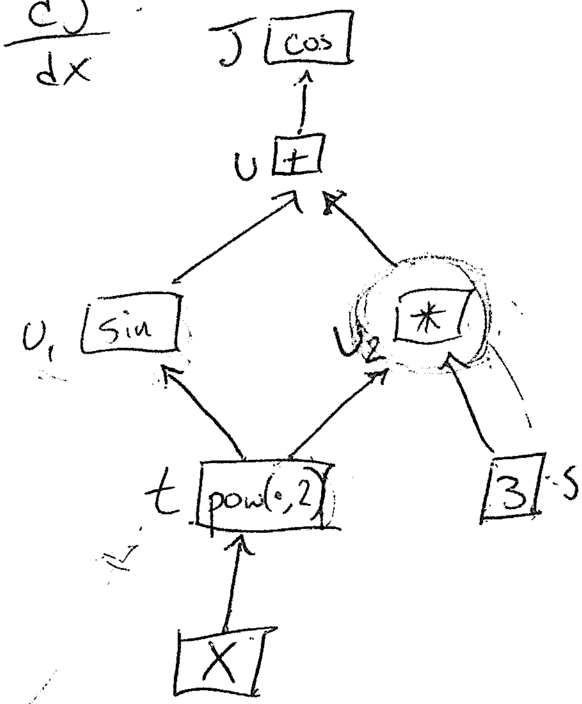
$u_1 = g_1(x, r_1)$

Simple Example

$$J = \cos(\sin(x^2) + 3x^2)$$

(s → 3)

Goal: $\frac{dJ}{dx}$



$$\frac{dJ}{dJ} = 1$$

$$\frac{dJ}{du} = \sin(u)$$

$$\frac{dJ}{du_1} = \frac{dJ}{du} \cdot \underbrace{\frac{du}{du_1}}_{1}$$

$$\frac{dJ}{du_2} = \frac{dJ}{du} \cdot \underbrace{\frac{du}{du_2}}_{1}$$

$$\frac{dJ}{dt} \leftarrow 0$$

$$\frac{dJ}{dt} \mathrel{+}= \frac{dJ}{du_1} \underbrace{\frac{du_1}{dt}}_{\cos(t)}$$

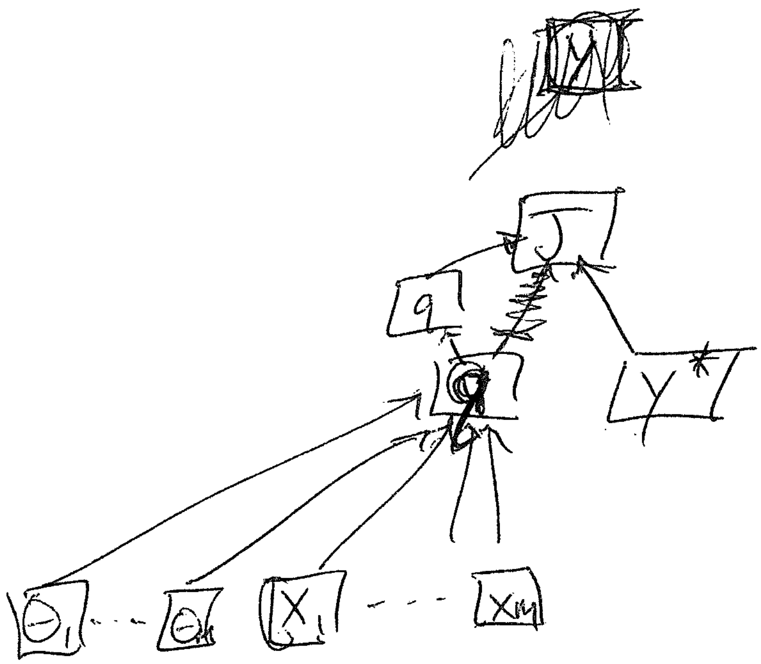
$$\frac{dJ}{dt} \mathrel{+}= \frac{dJ}{du_2} \underbrace{\frac{du_2}{dt}}_{3}$$

$$\frac{dJ}{ds} \mathrel{+}= \frac{dJ}{du_2} \cdot \frac{du_2}{ds}$$

$$\frac{dJ}{dx} \leftarrow 0$$

$$\frac{dJ}{dx} \mathrel{+}= \frac{dJ}{dt} \cdot \underbrace{\frac{dt}{dx}}_{2x}$$

LR Comp Graph



10-601

Sigmoid

$$\frac{d}{dx}\sigma(x) = (1-\sigma(x))\,\sigma(x)$$

(σ: sigmoid)

ReLU

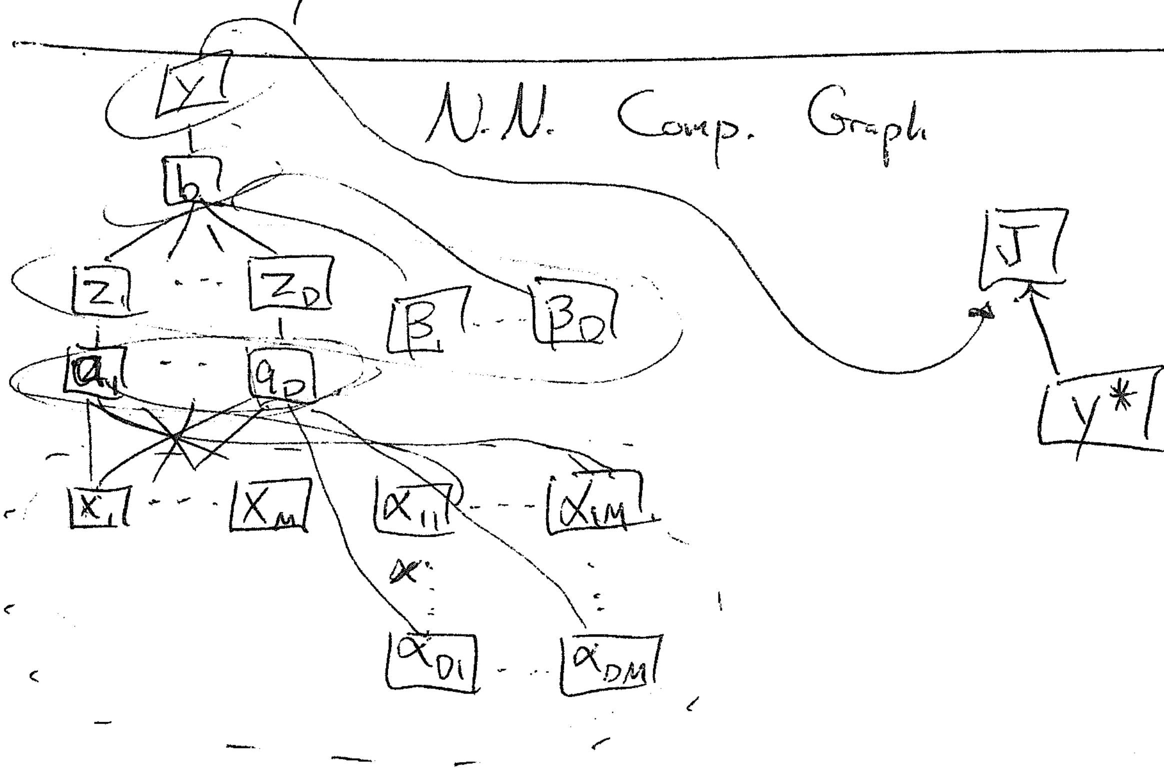
$$\frac{d}{da}\max(a,b) = \begin{cases} 1 & \text{if } a>b \\ 0 & \text{otherwise} \end{cases}$$

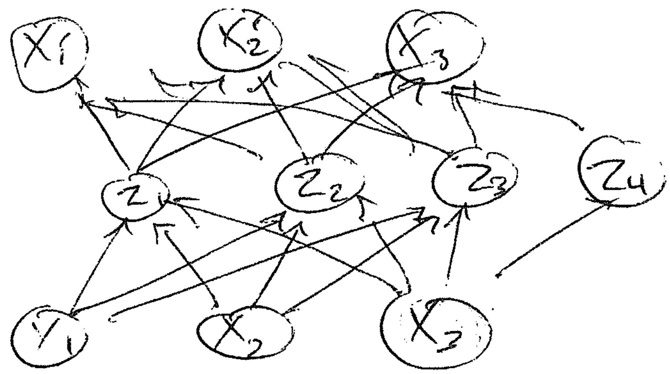
$$y \triangleq P(\cdot \mid x)$$

L.R.

$$y = f(x) + \epsilon$$

N.N. Comp. Graph




X1
X2
X3
Z1
Z2
Z3
Z4
X2
X3