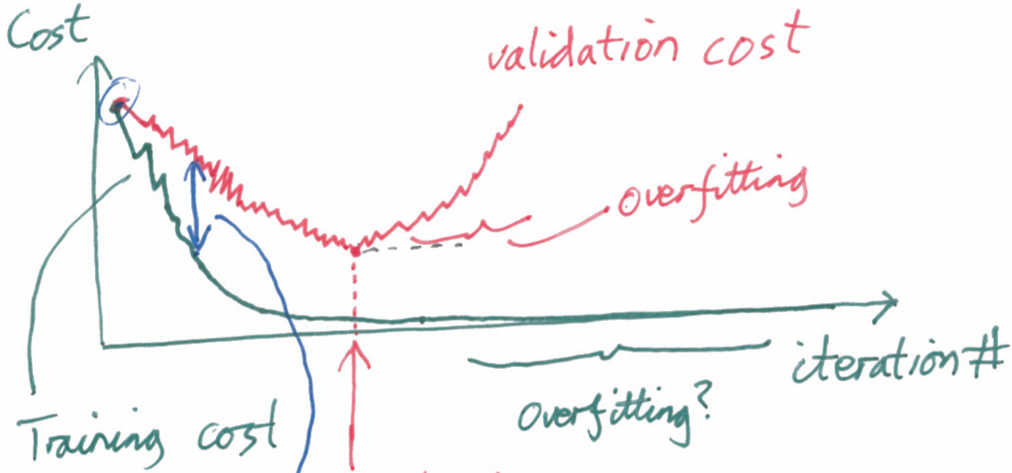
Regularization — Early-stopping

Could do L2 regularization

Set λ ... cross-validate



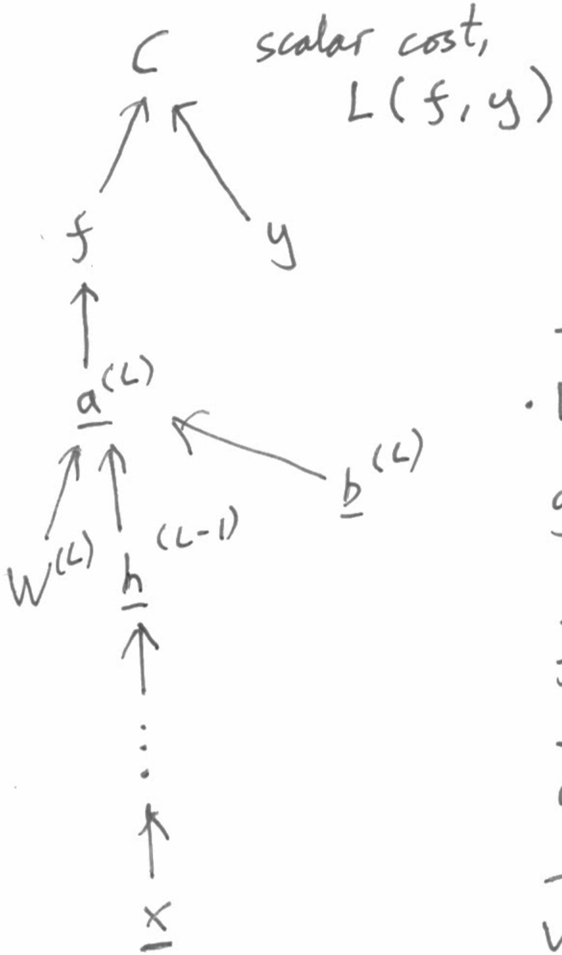
(val. error − train error)

is **not** "overfitingness"

$\Rightarrow$ do not use for model comparison

Getting gradients – Reverse-mode differentiation

"Backpropagation"



Graph
DAG

Strategy:

- For every intermediate θ

get $\bar{\theta} = \frac{\partial c}{\partial \theta}$

$$\bar{f} = \frac{\partial c}{\partial f}$$

$$\overline{a_i^{(L)}} = \frac{\partial c}{\partial a_i^{(L)}}$$

$$\overline{W_{ij}^{(L)}} = \frac{\partial c}{\partial W_{ij}^{(L)}}$$

Start at end of computation graph

Example: $c = (f-y)^2$

$$\bar{f} = \frac{\partial c}{\partial f} = 2(f-y)$$

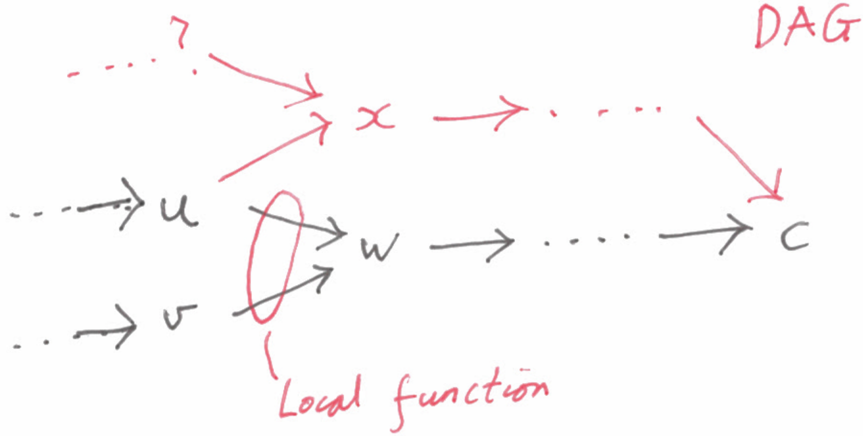
If have $\underset{N\times 1}{\underline{f}}$ and $\underset{N\times 1}{\underline{y}}$, $c = \sum_n (f_n - y_n)^2$

$$\frac{\partial c}{f_m} = 2(f_m - y_m)$$

$$\bar{\underline{f}} = 2(\underline{f} - \underline{y})$$

(I reshowed L13 6 and 8)

We combine local propagation rules



Assume we have $\bar{w} = \frac{\partial c}{\partial w}$

Want $\bar{u} = \frac{\partial c}{\partial u}$, $\bar{v} = \frac{\partial c}{\partial v}$

Chain rule

$$\bar{u} = \frac{\partial c}{\partial u} = \underbrace{\frac{\partial c}{\partial w}}_{\bar{w}} \underbrace{\frac{\partial w}{\partial u}}_{\text{Derivative of local function}} + \overbrace{\frac{\partial c}{\partial x}}^{\bar{x}} \overbrace{\frac{\partial x}{\partial u}}^{\text{Deriv local function}}$$

Derivative of local function
Depends on u, v, and/or w

Elementwise Functions

$$h_k = g(a_k)$$

$$\bar{a}_k = \frac{\partial c}{\partial a_k} = \frac{\partial c}{\partial h_k}\frac{\partial h_k}{\partial a_k}$$

$$= \bar{h}_k \, g'(a_k)$$

$$\bar{\underline{a}} = \bar{\underline{h}} \odot g'(\underline{a})$$

Hadamard product

* Python

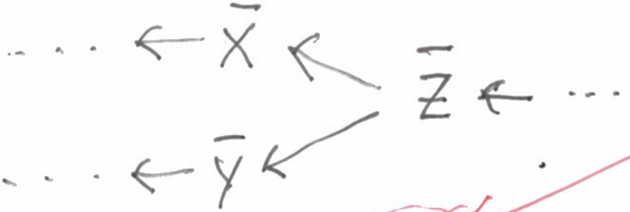
.* Matlab

Example: Matrix Multiplication

$N \times N$ matrix, N^2 numbers in matrix

$O(N^3)$, $\cancel{O(N^4)}$, $\cancel{O(N^\alpha)}$ $\cancel{2.59 < 3}$ $\cancel{2.2}$

$$Z = XY$$



N^4 settings

$$\frac{\partial c}{\partial X_{ij}} = \sum_{m,n} \underbrace{\frac{\partial c}{\partial Z_{mn}}}_{\bar{Z}_{mn}} \underbrace{\frac{\partial Z_{mn}}{\partial X_{ij}}}_{\delta_{mi} Y_{jn}} = \sum_n \bar{Z}_{in} \underbrace{Y_{jn}}_{(Y^T)_{nj}}$$

$$\boxed{\bar{X} = \bar{Z} Y^T} \qquad \boxed{\bar{Y} = X^T \bar{Z}}$$

$$Z_{mn} = \sum_{p} X_{mp} Y_{pn}$$

$$= X_{m1} Y_{1n} + X_{m2} Y_{2n} + \dots$$

$$\dots \underbrace{X_{mj} Y_{jn}} + \dots$$

$$\underset{M \times N}{Z} = \underset{M \times P}{X} \; \underset{P \times N}{Y}$$

$$\underset{P \times N}{\bar{Y}} = \underset{P \times M}{X^T} \; \underset{M \times N}{\bar{Z}}$$