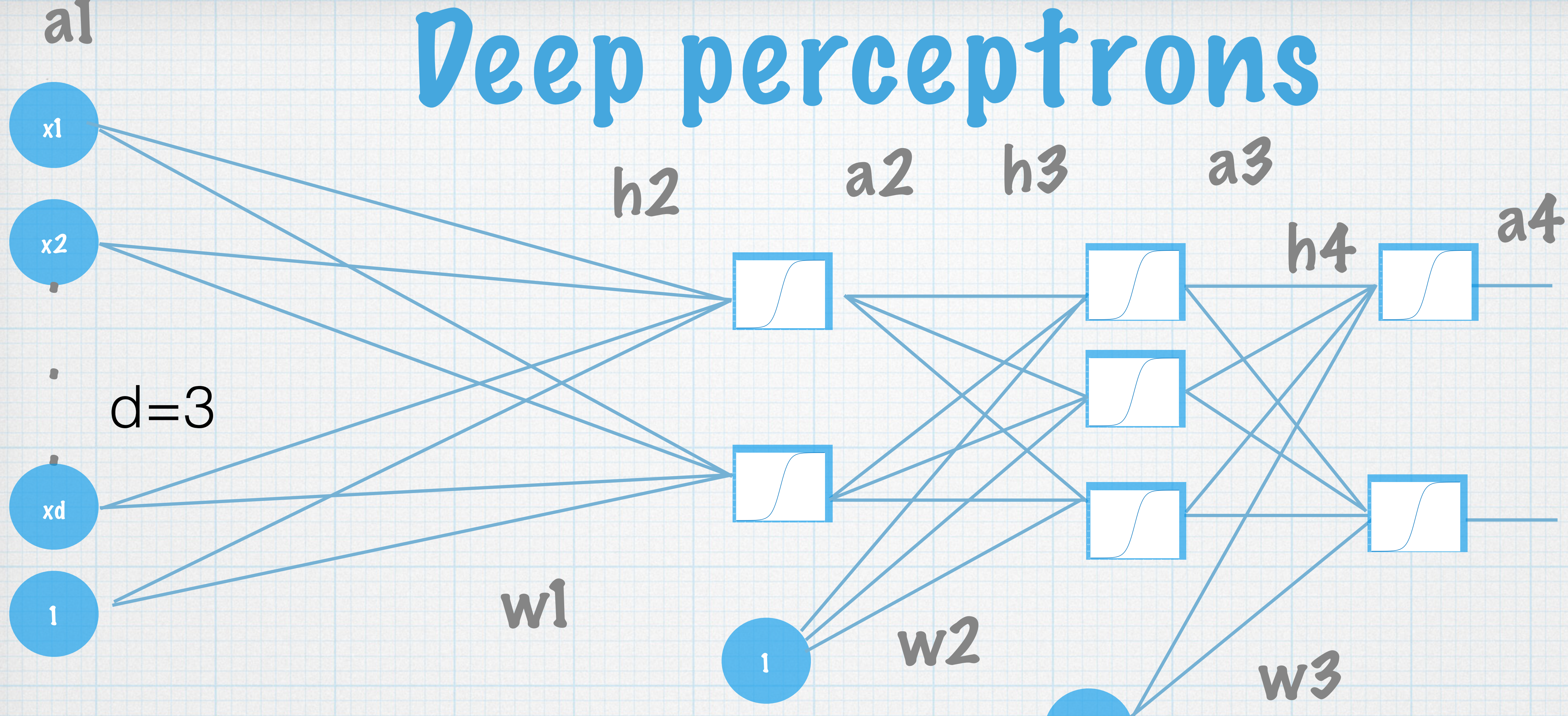
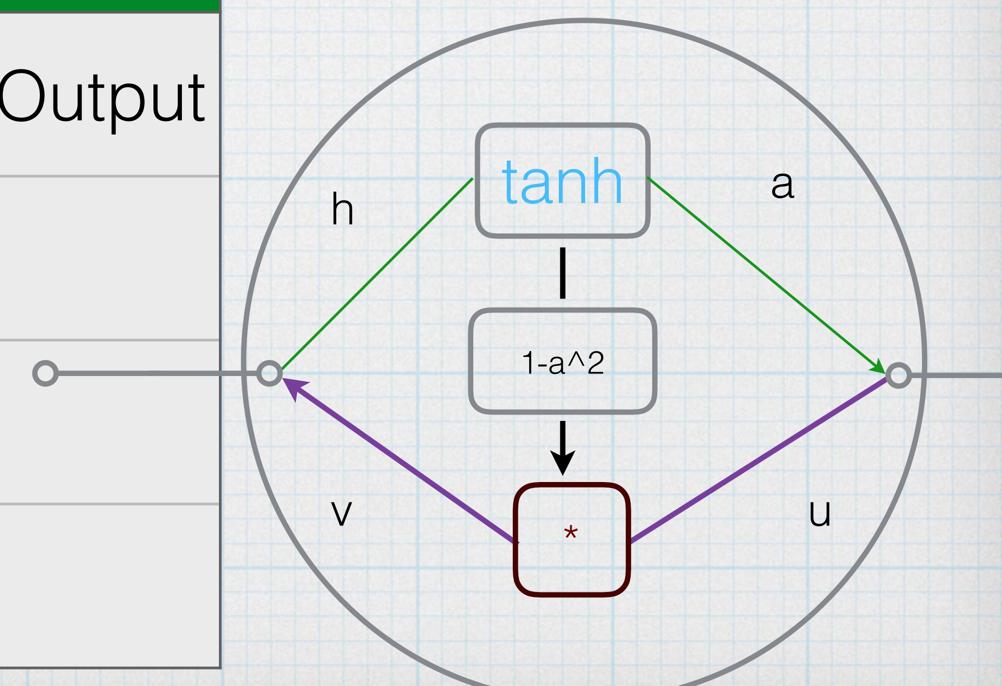
Gradients in Neural mapping

Numerical and Symbolic approaches

Deep perceptrons



Size	u	v	backward. w'	h	W	a	
4	2x2	2x2		2x1		2x1	Output
3	3x2	3x2	3x2	3x1	2x4	3x1	
2	2x2	2x2	2x3	2x1	3x3	2x1	
1			3x2		2x4	3x1	



$$W\{n\} = \begin{bmatrix} w\{n\} & b\{n\} \\ & \end{bmatrix}$$

$$h\{n+1\} = W\{n\} \begin{bmatrix} a\{n\} \\ 1 \end{bmatrix}$$

a{n}=tanh(h{n}), for n>1

Backward Linear transformation

- For n=L-1:1, where L denotes the number of layers

$$\frac{dh\{n+1\}}{da\{n\}} = w\{n\}$$

Eq 1

$$\frac{dy}{da\{n\}} = \frac{dy}{dh\{n+1\}} \frac{dh\{n+1\}}{da\{n\}}$$

Eq 2

$$\Rightarrow \quad u\{n\} = w^T\{n\} v\{n+1\}$$

Backward Nonlinear transformation

$$a\{n\} = \mathbf{tanh}(h\{n\})$$

Eq 3

$$\frac{dy}{dh\{n\}} = \frac{dy}{da\{n\}} \frac{da\{n\}}{dh\{n\}}$$

$$v\{n\} = u\{n\} \frac{da\{n\}}{dh\{n\}}$$

Sizes of matrices

$h\{n+1\}=w\{n\}*a\{n\}$

$a\{n\}=\tanh(h\{n\})$, for $n>1$

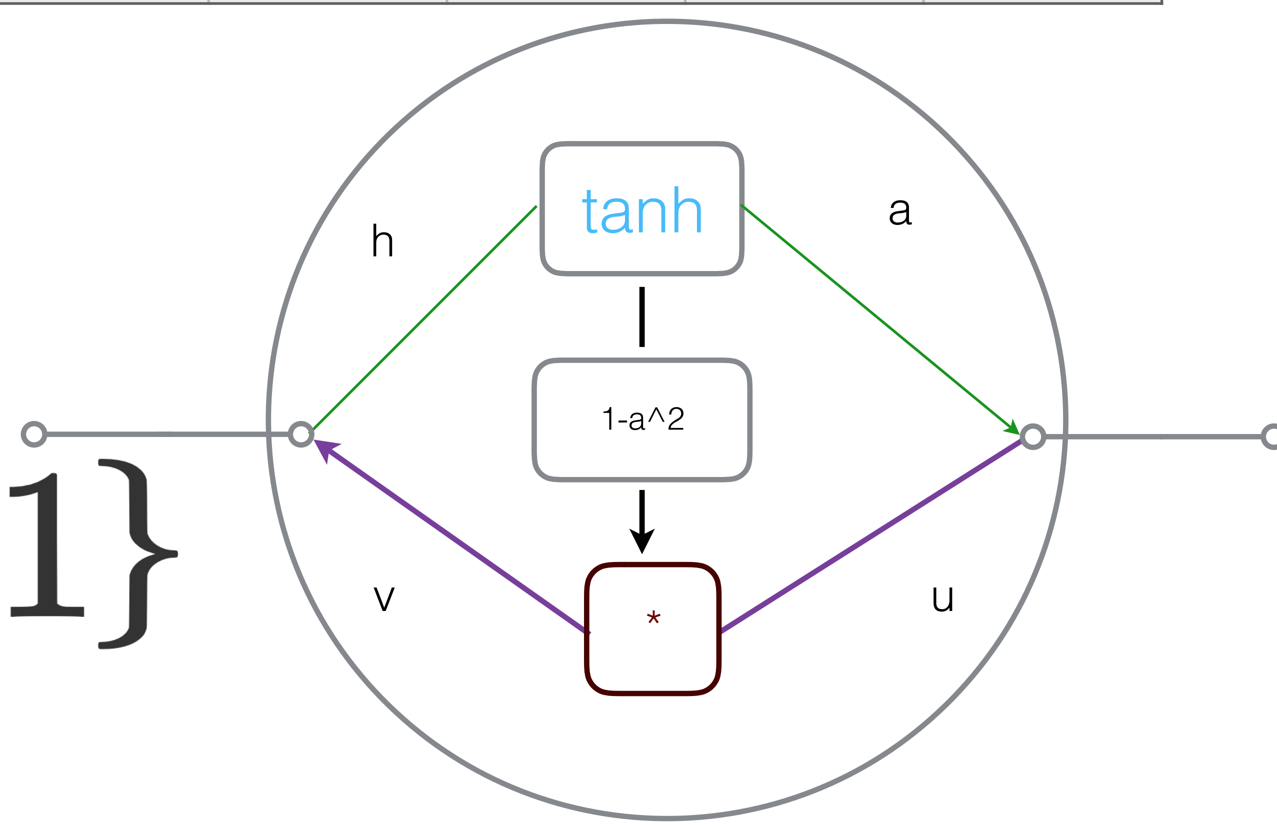
$$u\{n\}\equiv\frac{dy}{da\{n\}}$$

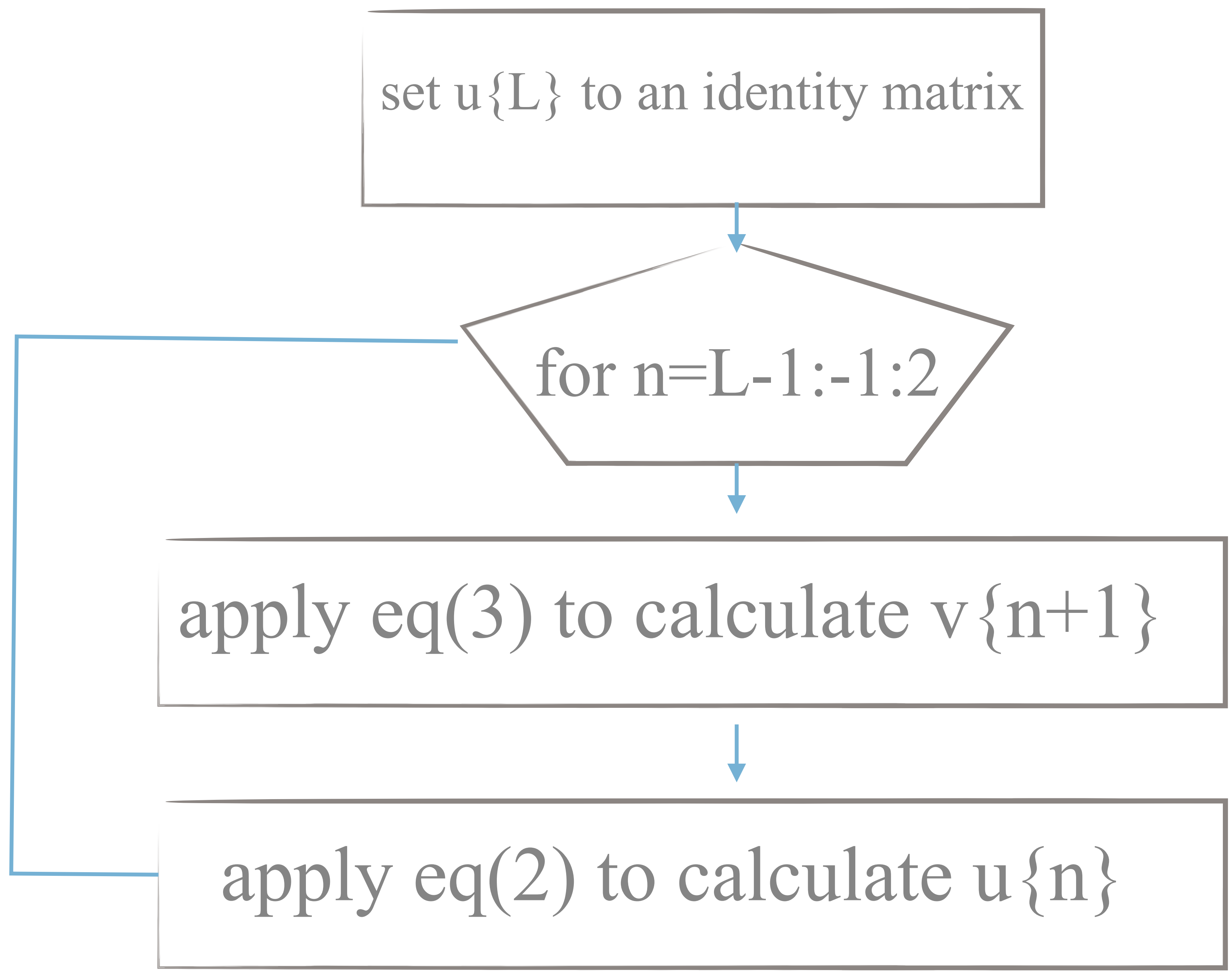
$$v\{n\}=\frac{dy}{dh\{n\}}$$

Size	u	v	backward. w'	h	W	a	
4	2x2	2x2		2x1		2x1	Output
3	3x2	3x2	3x2	3x1	2x4	3x1	
2	2x2	2x2	2x3	2x1	3x3	2x1	
1			3x2		2x4	3x1	

Eq 2

$$u\{n\}=w^T\{n\}v\{n+1\}$$




set u{L} to an identity matrix
for n=L-1:-1:2
apply eq(3) to calculate v{n+1}
apply eq(2) to calculate u{n}

$$h\{n+1\} = W\{n\} \begin{bmatrix} a\{n\} \\ 1 \end{bmatrix}$$

y{m} denotes an output component

$$\frac{dy\{m\}}{dW\{n\}} = \frac{dy\{m\}}{dh\{n+1\}} \frac{dh\{n+1\}}{dW\{n\}}$$

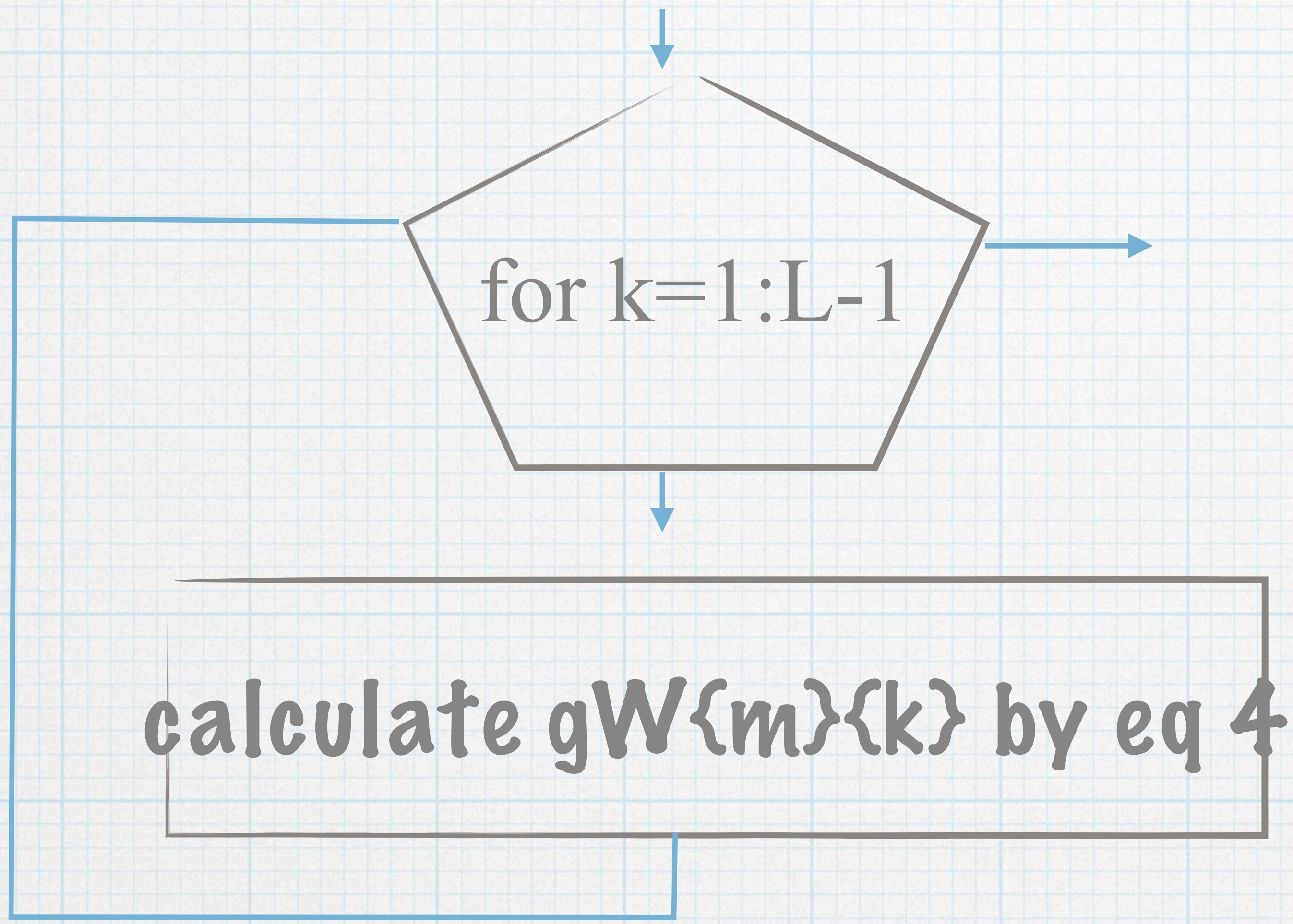
denoted by $v_m\{n+1\}$

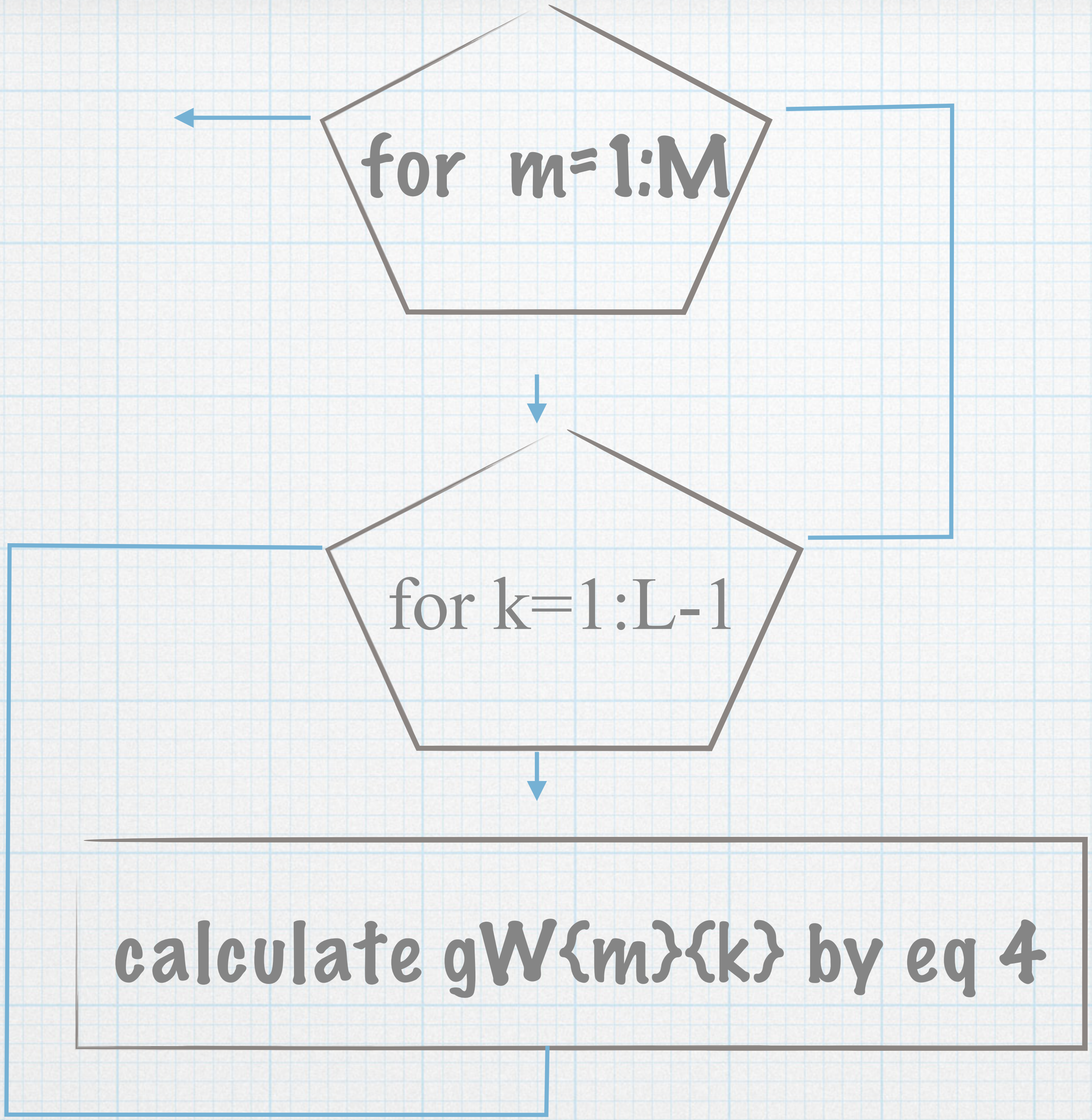
$$= v_m\{n+1\} \frac{dh\{n+1\}}{dW\{n\}}$$

$$= v_m\{n+1\} \begin{bmatrix} a\{n\} \\ 1 \end{bmatrix}^T$$

eq 4

$computing \quad the \quad gradient \quad of \quad y\{m\} \quad with \quad respect \quad to \quad w\{k\} \quad for \quad all \quad k$




for m=1:M
for k=1:L-1
calculate gW{m}{k} by eq 4

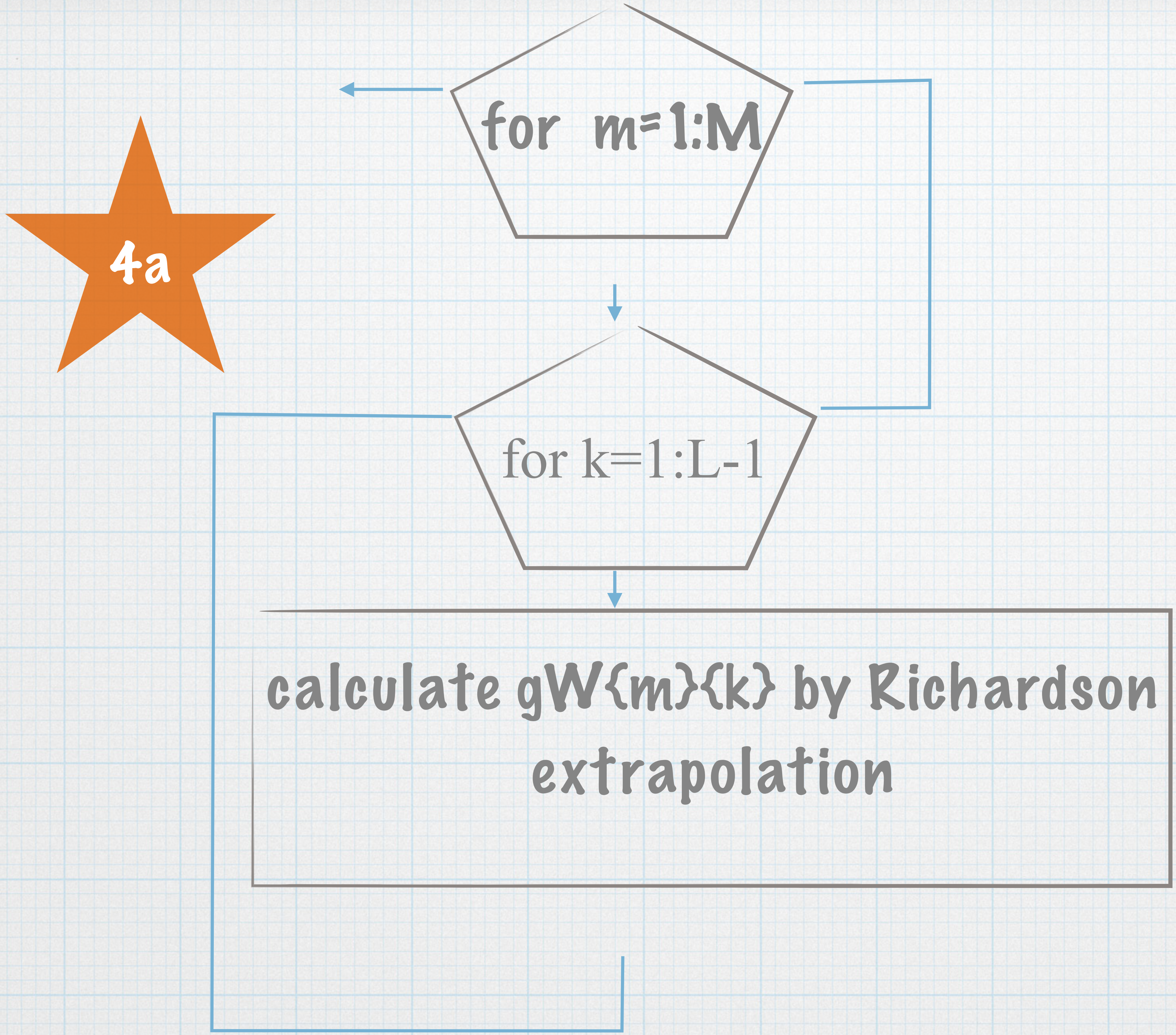

3

Gradient check

Richardson extrapolation

$$f'(x) \approx \varphi\left(\frac{h}{2}\right) + \frac{1}{3}\left[\varphi\left(\frac{h}{2}\right) - \varphi(h)\right]$$

$$\varphi(h) = \frac{f(x+h) - f(x-h)}{2h}$$


4a
for m=1:M
for k=1:L-1
calculate gW{m}{k} by Richardson extrapolation

Richardson Extra

Richardson extrapolation

$$f'(x) \approx \varphi(\frac{h}{2}) + \frac{1}{3}\left[\varphi(\frac{h}{2}) - \varphi(h)\right]$$

$$\varphi(h) = \frac{f(x+h) - f(x-h)}{2h}$$

4b

set small z

for each element in W{m}{k}

Perturb it by adding z,-z,z/2,-z/2 respectively and determine four corresponding output y{m}, denoted by f1,f2,f3,f4

set corresponding element in gW{m}{k} to extrapolation determined by f1-f4

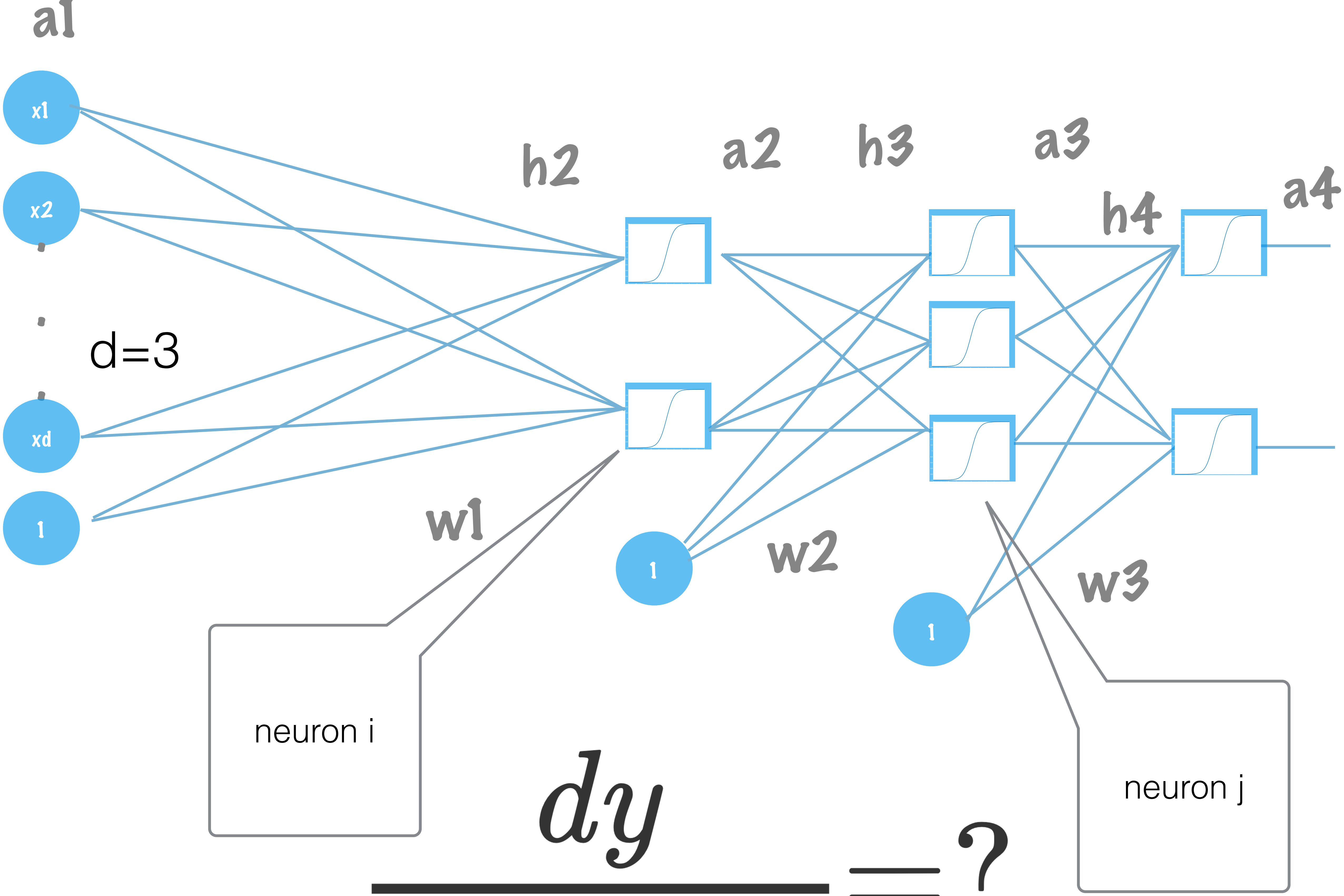
g1 =(f1-f2)/(2z)

g2=(f3-f4)/z

$$\frac{dy}{dW_3} = ? \quad \frac{dy}{dW_2} = ?$$

$$\frac{dy}{dW_1} = ?$$

dw{n}=? for all n



$$\frac{dy}{dW_{m2}[i,j]} = ?$$

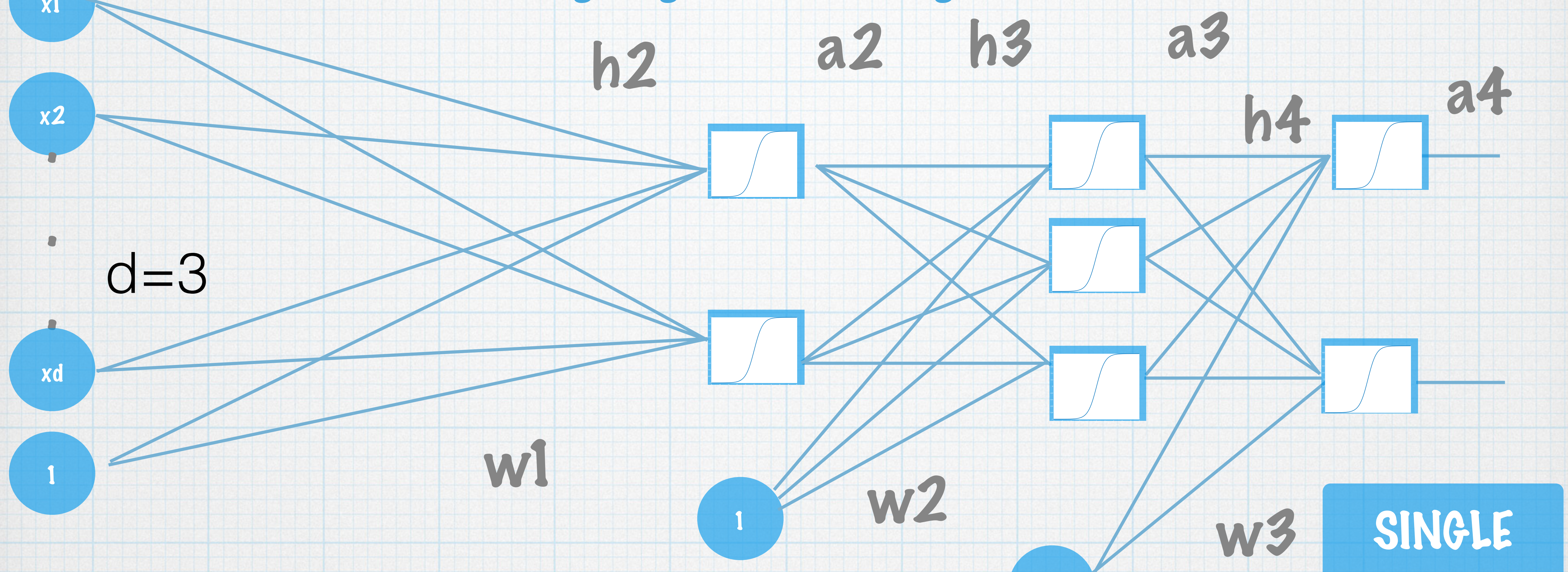
Batch

A batch contains a set of inputs.

Let u{m} and v{m} denote gradients corresponding the mth output component.

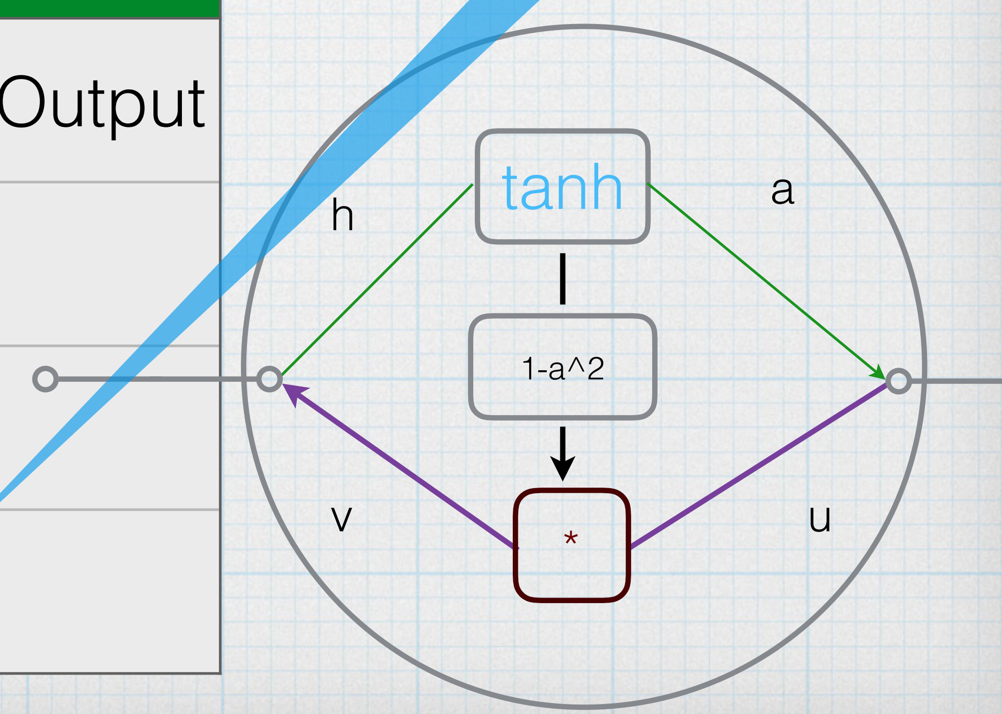
Extend representations for a set of inputs

Deep perceptrons

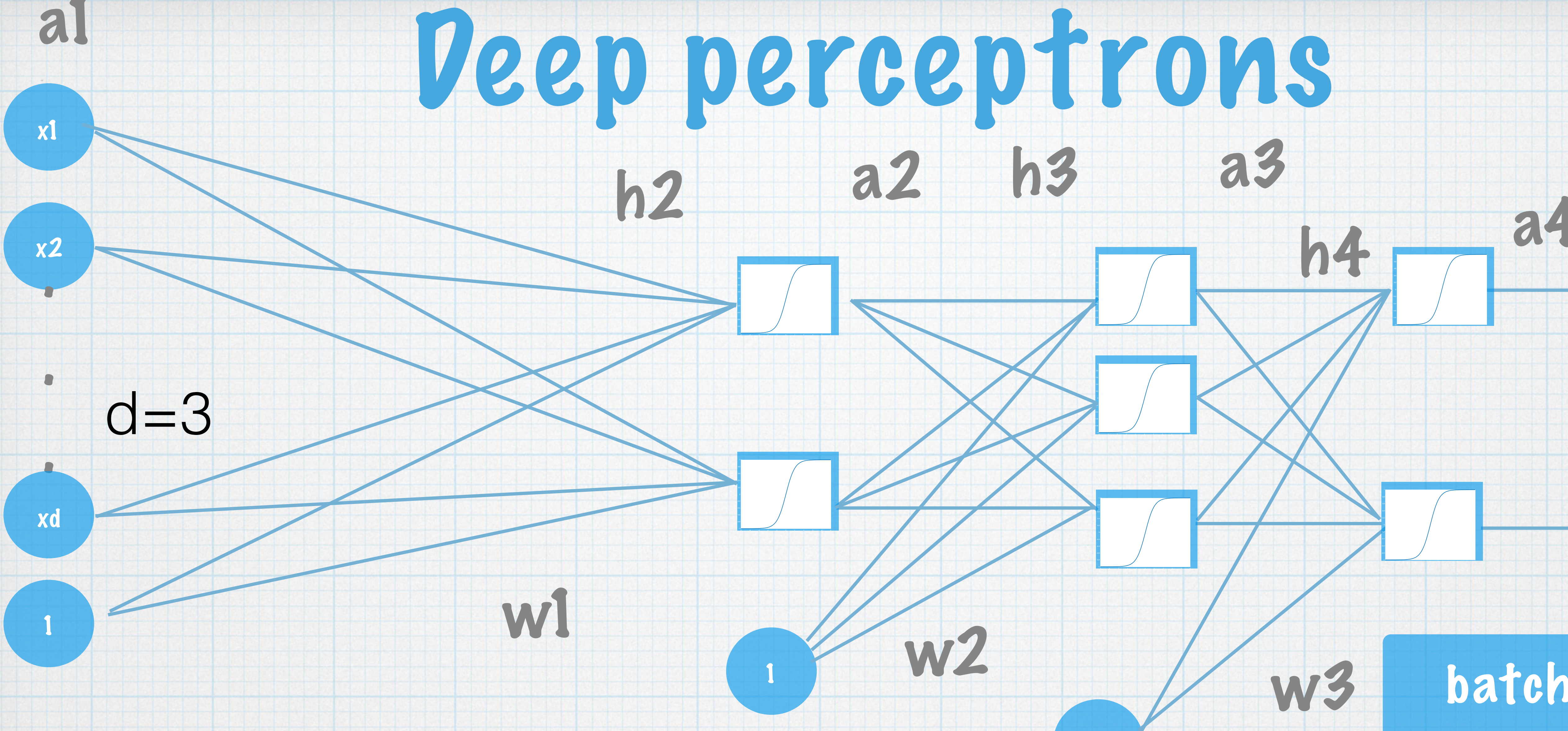


Size	u{m}	v{m}	backward. w'	h	W	a	
4	2x1	2x1		2x1		2x1	Output
3	3x1	3x1	3x2	3x1	2x4	3x1	
2	2x1	2x1	2x3	2x1	3x3	2x1	
1			3x2		2x4	3x1	

SINGLE input

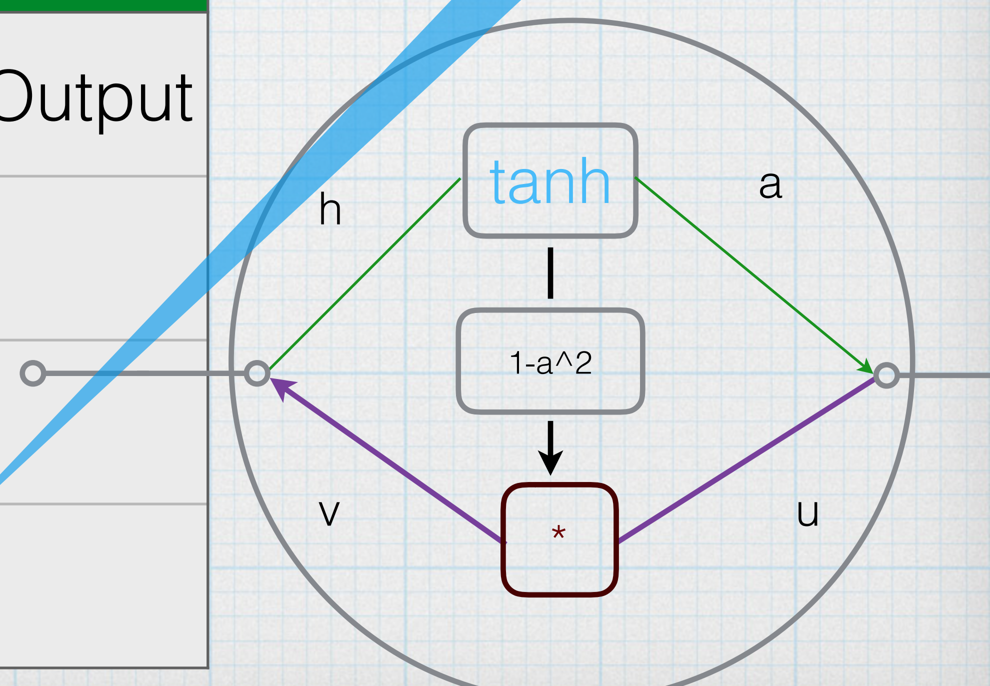


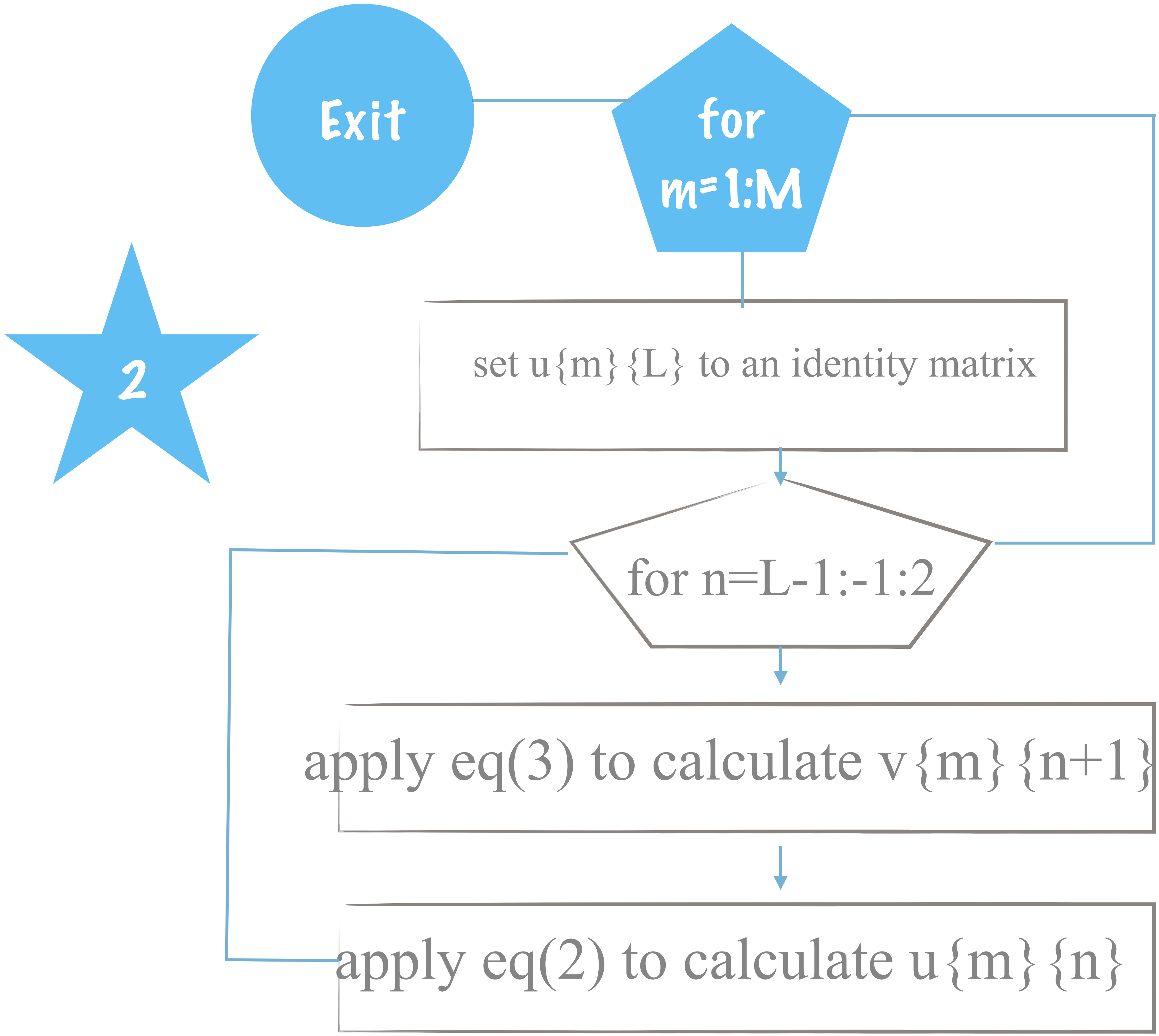
Deep perceptrons



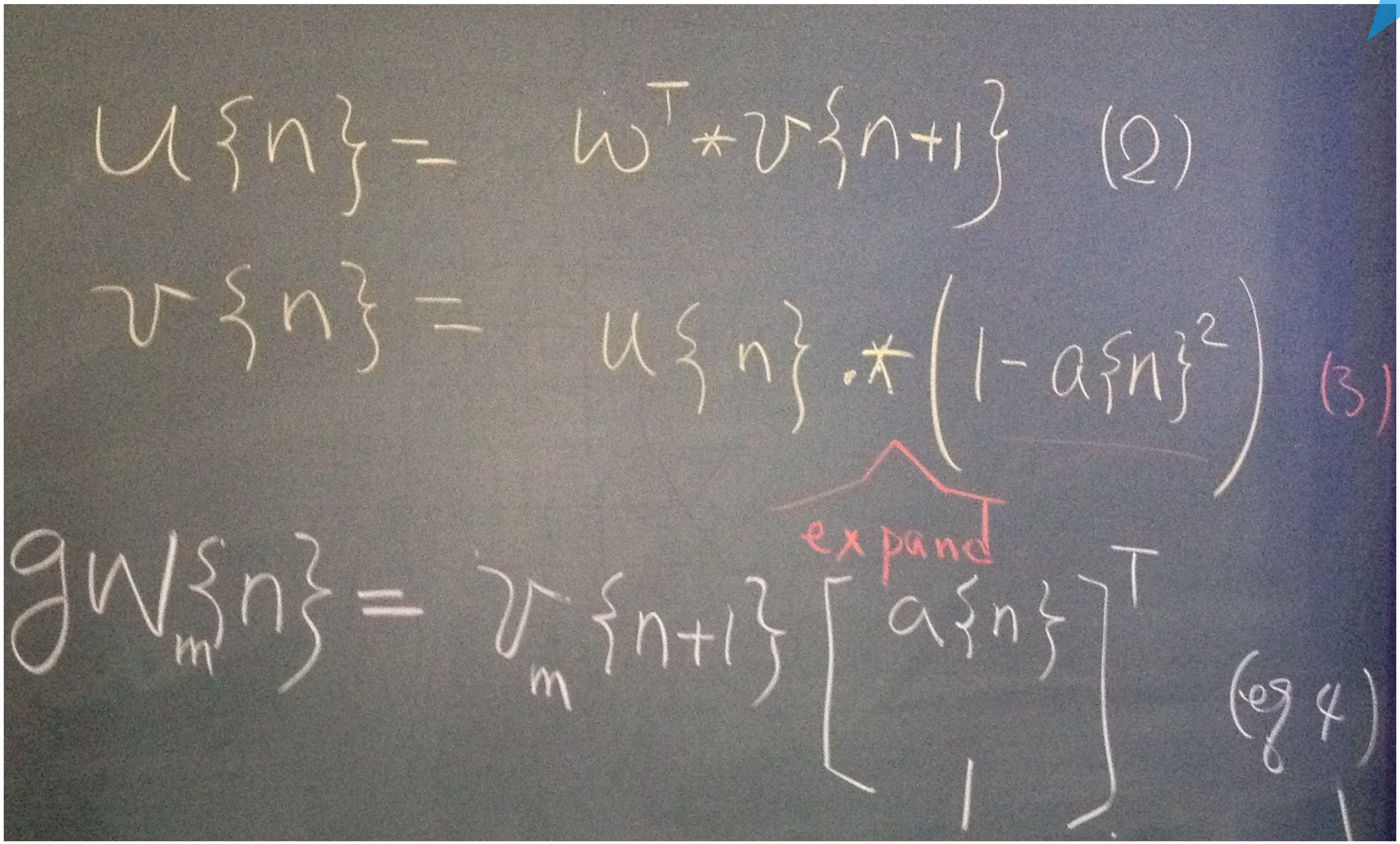
Size	u{m}	v{m}	backward. w'	h	W	a	
4	2xn	2xn		2xn		2xn	Output
3	3xn	3xn	3x2	3xn	2x4	3xn	
2	2xn	2xn	2x3	2xn	3x3	2xn	
1			3x2		2x4	3xn	

batch input




Exit
for m=1:M
2
set u{m}{L} to an identity matrix
for n=L-1:-1:2
apply eq(3) to calculate v{m}{n+1}
apply eq(2) to calculate u{m}{n}

1



Square error

$$E(W) = \sum_t \sum_m e_m^2[t]$$

$$e_m[t] = y[t] - \hat{y}[t]$$

project

Submit your codes to ett2012@gmail.com pm9 Dec. 1

Accomplish two methods of classdef of perceptrons

1. Calculate gradients of outputs with respect to weight matrices
2. Compare calculated gradients with numerical gradients for realization of Gradient-check