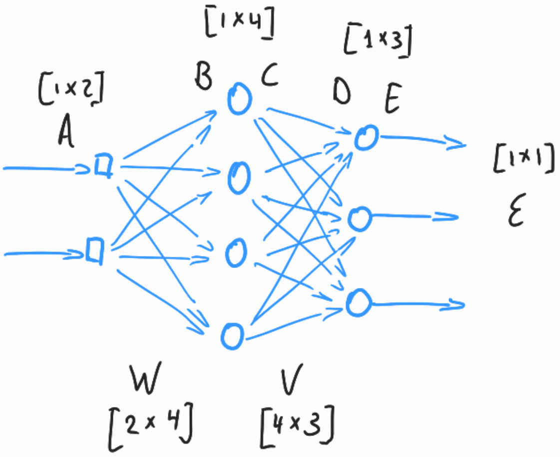
Backpropagation



$\vec{a}$ = input (row vector)

$\vec{b} = \vec{a} \cdot W$

$\vec{c} = \tanh(\vec{b})$

$\vec{d} = \vec{c} \cdot V$

$\vec{e} = \vec{d}$ (no activation)

$\varepsilon = \sum_K (e_K - y_K)^2$ target

Chain rule (multivariate!)

$f(g_1(x), g_2(x), \dots, g_N(x))$

$$\boxed{\frac{\partial f}{\partial x} = \sum_{n=1}^{N} \frac{\partial f}{\partial g_n} \frac{\partial g_n}{\partial x}}$$

$\boxed{\text{Task 1}}$ find $\frac{\partial \varepsilon}{\partial v_{jK}}$ and $\frac{\partial \varepsilon}{\partial w_{ij}}$

How?

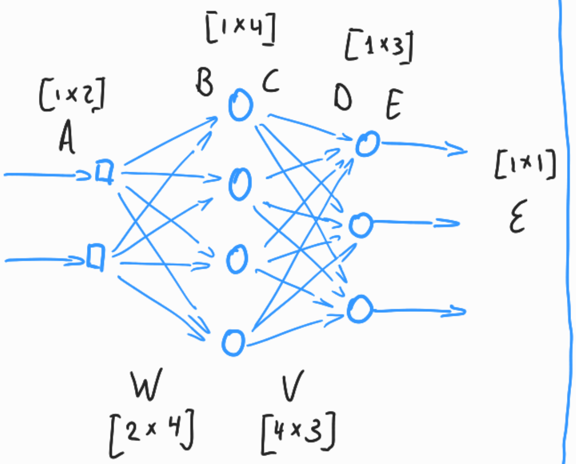
1. compute all intermediate partial derivatives:

$\frac{\partial \varepsilon}{\partial e_K}$, $\frac{\partial e_K}{\partial d_K}$, $\frac{\partial d_K}{\partial v_{jK}}$, $\frac{\partial d_K}{\partial c_j}$

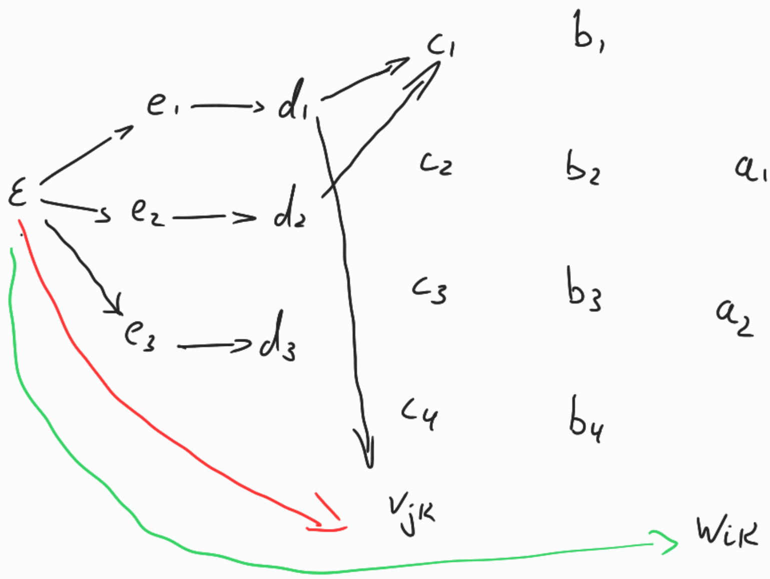
$\frac{\partial c_j}{\partial b_j}$, $\frac{\partial b_j}{\partial w_{ij}}$

2. Apply chain rule...

... but when do sums appear?



Variable dependence



Task 2 find $\frac{\partial E}{\partial V}$ and $\frac{\partial E}{\partial W}$

How?

- Examine the formulas for $\frac{\partial E}{\partial v_{j,k}}$ and $\frac{\partial E}{\partial w_{i,j}}$
- Try to combine terms with the same indices (i, j, k)
- Note that $\frac{\partial E}{\partial A}$ has the same shape as A!
- Some operations will be elementwise and some will be matrix.
- Some new terms will appear on the left and some on the right.
- Some will be transposed and some won't

Two things to look out for:

- Matrix-vector product:

$$\vec{y} = \vec{x} A$$

$$y_k = \sum_j x_j a_{jk}$$

- Outer product:

$$\begin{cases} \vec{x} = [x_1 \dots x_N] \\ \vec{y} = [y_1 \dots y_M] \end{cases} \longrightarrow \vec{x}^T \vec{y} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_M \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_M \\ \vdots & & & \vdots \\ x_N y_1 & x_N y_2 & \cdots & x_N y_M \end{bmatrix}$$

$[N \times M]$

$x_i y_j \longrightarrow$ outer

Try to identify these in your formulas!