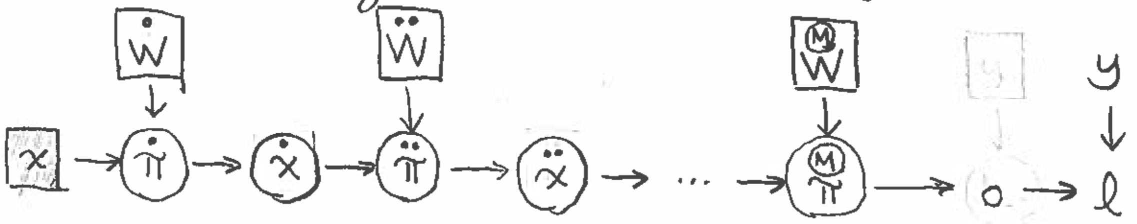
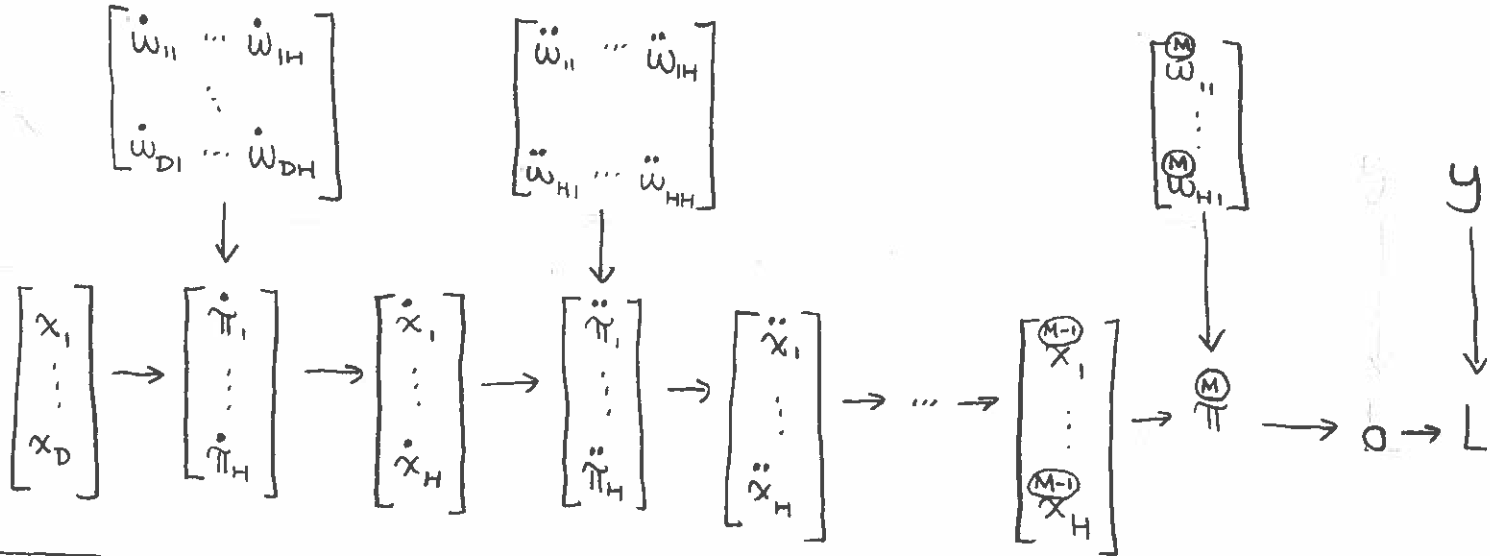
Backpropagation: A Matrix Formulation

① Recall the M-layer feedforward neural network:



which, more explicitly representing the matrices and vecters, looks like:



② The major computational step of backpropagation is to compute

$$\frac{\partial \overset{(M)}{\pi}}{\partial \overset{(m)}{\pi}_j} = a'\left(\overset{(m)}{\pi}_j\right) \sum_{h=1}^{H} \overset{(m+1)}{w}_{hj} \cdot \frac{\partial \overset{(M)}{\pi}}{\partial \overset{(m+1)}{\pi}_h}$$

for each $m \in \{1, \dots, M-1\}$ and $j \in \{1, \dots, H\}$.

(3) This is quite a lot of computation!

M layers
└ H variables of the form $\pi_j^{(m)}$ (per layer)
  └ H terms to sum over (per variable)

So it is order $O(MH^2)$.

(4) Generally, when faced with quadratic computation (or cubic in this case), it's useful to investigate whether the computation can be expressed as a matrix computation. This is because:

(a) there are libraries that are heavily optimized (like Torch) to do matrix operations quickly

(b) many computers have specialized hardware (called Graphics Processing Units, or GPUs) for matrix computation.

BACKPROPAGATION: A MATRIX FORMULATION

(5) So let's see whether we can compute all the partial derivatives $\dfrac{\partial \overset{(M)}{\pi}}{\partial \overset{(m)}{\pi}_j}$ at a given layer m via matrix computation.

$$\frac{\partial \overset{(M)}{\pi}}{\partial \overset{(m)}{\pi}} = \begin{bmatrix} \dfrac{\partial \overset{(M)}{\pi}}{\partial \overset{(m)}{\pi}_1} \\ \vdots \\ \dfrac{\partial \overset{(M)}{\pi}}{\partial \overset{(m)}{\pi}_H} \end{bmatrix} = \begin{bmatrix} a'(\overset{(m)}{\pi}_1) \displaystyle\sum_{h=1}^{H} \overset{(m+1)}{\omega}_{h1} \cdot \dfrac{\partial \overset{(M)}{\pi}}{\partial \overset{(m+1)}{\pi}_h} \\ \vdots \\ a'(\overset{(m)}{\pi}_H) \displaystyle\sum_{h=1}^{H} \overset{(m+1)}{\omega}_{hH} \cdot \dfrac{\partial \overset{(M)}{\pi}}{\partial \overset{(m+1)}{\pi}_h} \end{bmatrix}$$

remember that $x \odot y$ is the elementwise product of two equal-size vectors x and y

$$= \begin{bmatrix} a'(\overset{(m)}{\pi}_1) \\ \vdots \\ a'(\overset{(m)}{\pi}_H) \end{bmatrix} \underset{\text{"Hadamard" product}}{\odot} \begin{bmatrix} \displaystyle\sum_{h=1}^{H} \overset{(m+1)}{\omega}_{h1} \cdot \dfrac{\partial \overset{(M)}{\pi}}{\partial \overset{(m+1)}{\pi}_h} \\ \vdots \\ \displaystyle\sum_{h=1}^{H} \overset{(m+1)}{\omega}_{hH} \cdot \dfrac{\partial \overset{(M)}{\pi}}{\partial \overset{(m+1)}{\pi}_h} \end{bmatrix}$$

$$= a'(\overset{(m)}{\pi}) \odot \begin{bmatrix} \overset{(m+1)}{\omega}_{11} & \cdots & \overset{(m+1)}{\omega}_{H1} \\ & \ddots & \\ \overset{(m+1)}{\omega}_{1H} & \cdots & \overset{(m+1)}{\omega}_{HH} \end{bmatrix}^{T} \begin{bmatrix} \dfrac{\partial \overset{(M)}{\pi}}{\partial \overset{(m+1)}{\pi}_1} \\ \vdots \\ \dfrac{\partial \overset{(M)}{\pi}}{\partial \overset{(m+1)}{\pi}_H} \end{bmatrix}$$

$$= a'(\overset{(m)}{\pi}) \odot \overset{(m+1)}{W}^{T} \frac{\partial \overset{(M)}{\pi}}{\partial \overset{(m+1)}{\pi}}$$

Success!

BACKPROPAGATION: A MATRIX FORMULATION

⑥ This means we can re-express backpropagation's step (a):

BACKPROPAGATION:

(a) for m in $\{M-1, \ldots, 1\}$:

$$\text{compute } \frac{\partial \overset{(M)}{\pi}}{\partial \overset{(m)}{\pi}} = a'(\overset{(m)}{\pi}) \odot \overset{(m+1)}{W}{}^{T} \frac{\partial \overset{(M)}{\pi}}{\partial \overset{(m+1)}{\pi}}$$

(b) $$\frac{\partial L}{\partial \overset{(m)}{w}_{ij}} = \frac{\partial L}{\partial \overset{(M)}{\pi}} \cdot \overset{(m-1)}{x}_i \cdot \frac{\partial \overset{(M)}{\pi}}{\partial \overset{(m)}{\pi}_j}$$

---

⑦ While we're on a roll, let's try to matrixify step (b)

$$\frac{\partial L}{\partial \overset{(m)}{W}} = \begin{bmatrix} \frac{\partial L}{\partial \overset{(m)}{w}_{11}} & \cdots & \frac{\partial L}{\partial \overset{(m)}{w}_{1H}} \\ & \ddots & \\ \frac{\partial L}{\partial \overset{(m)}{w}_{H1}} & \cdots & \frac{\partial L}{\partial \overset{(m)}{w}_{HH}} \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial \overset{(M)}{\pi}} \overset{(m-1)}{x}_1 \frac{\partial \overset{(M)}{\pi}}{\partial \overset{(m)}{\pi}_1} & \cdots & \frac{\partial L}{\partial \overset{(M)}{\pi}} \overset{(m-1)}{x}_1 \frac{\partial \overset{(M)}{\pi}}{\partial \overset{(m)}{\pi}_H} \\ & \ddots & \\ \frac{\partial L}{\partial \overset{(M)}{\pi}} \overset{(m-1)}{x}_H \frac{\partial \overset{(M)}{\pi}}{\partial \overset{(m)}{\pi}_1} & \cdots & \frac{\partial L}{\partial \overset{(M)}{\pi}} \overset{(m-1)}{x}_H \frac{\partial \overset{(M)}{\pi}}{\partial \overset{(m)}{\pi}_H} \end{bmatrix}$$

$$= \frac{\partial L}{\partial \overset{(M)}{\pi}} \cdot \begin{bmatrix} \frac{\partial \overset{(M)}{\pi}}{\partial \overset{(m)}{\pi}_1} \\ \vdots \\ \frac{\partial \overset{(M)}{\pi}}{\partial \overset{(m)}{\pi}_H} \end{bmatrix} \begin{bmatrix} \overset{(m-1)}{x}_1 & \cdots & \overset{(m-1)}{x}_H \end{bmatrix}$$

$$= \frac{\partial L}{\partial \overset{(M)}{\pi}} \cdot \frac{\partial \overset{(M)}{\pi}}{\partial \overset{(m)}{\pi}} \cdot \overset{(m-1)}{x}$$

BACKPROPAGATION: A MATRIX FORMULATION

(8) This gives us a matrix formulation of backpropagation

BACKPROPAGATION:

for m in $\{M-1, \dots, 1\}$:

(a) Compute $\dfrac{\partial \pi^{(M)}}{\partial \pi^{(m)}} = a'(\pi^{(m)}) \odot W^{(m+1)T} \dfrac{\partial \pi^{(M)}}{\partial \pi^{(m+1)}}$

(b) Compute $\dfrac{\partial L}{\partial w^{(m)}} = \dfrac{\partial L}{\partial \pi^{(M)}} \cdot \dfrac{\partial \pi^{(M)}}{\partial \pi^{(m)}} \cdot x^{(m-1)}$