Multi-Layer NN Notes

Daniel Polites

Notation Setup

source: https://arxiv.org/abs/1801.05894

Scalars

Layers: 1-L, indexed by l

Number of Neurons in layer l: n_l

Neuron Activations: $a_{\text{neuron num}}^{(\text{layer num})} = a_j^{(l)}$. Vector of activations for a layer is $a^{(l)}$

Activation Function: $g(\cdot)$ is our generic activation function

---

X

We have our input matrix $X \in \mathbb{R}^{\text{vars} \times \text{obs}} = \mathbb{R}^{n_0 \times m}$:

$$X = n_0 \text{ outputs} \left\{ \overbrace{\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_0,1} & x_{n_0,2} & \cdots & x_{n_0,m} \end{bmatrix}}^{m \text{ obs}} \right.$$

W

our Weight matrices $W^{(l)} \in \mathbb{R}^{\text{out} \times \text{in}} = \mathbb{R}^{n_l \times n_{l-1}}$:

$$W^{(l)} = n_l \text{ outputs} \left\{ \overbrace{\begin{bmatrix} w_{1,1}^{(l)} & w_{1,2}^{(l)} & \cdots & w_{1,n_{l-1}}^{(l)} \\ w_{2,1}^{(l)} & w_{2,2}^{(l)} & \cdots & w_{2,n_{l-1}}^{(l)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n_l,1}^{(l)} & w_{n_l,2}^{(l)} & \cdots & w_{n_l,n_{l-1}}^{(l)} \end{bmatrix}}^{n_{l-1} \text{ inputs}} \right.$$

$W^{(l)}$ is the weight matrix for the lth layer

---

b

our Bias matrices $b^{(l)} \in \mathbb{R}^{\text{out}\times 1} = \mathbb{R}^{n_l \times 1}$:

$$b^{(l)} = {\scriptstyle n_l \text{ outputs}} \left\{ \begin{bmatrix} b_1^{(l)} \\ b_2^{(l)} \\ \vdots \\ b_{n_l}^{(l)} \end{bmatrix} \right.$$

$b^{(l)}$ is the bias matrix for the lth layer

Y

our target layer matrix $Y \in \mathbb{R}^{\text{cats}\times\text{obs}} = \mathbb{R}^{n_L \times m}$:

$$Y = {\scriptstyle n_L \text{ categories}} \left\{ \overbrace{\begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,m} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n_L,1} & y_{n_L,2} & \cdots & y_{n_L,m} \end{bmatrix}}^{m \text{ obs}} \right.$$

---

z

our neuron layer's activation function input $z^{(l)} \in \mathbb{R}^{\text{out}\times 1} = \mathbb{R}^{n_l \times 1}$:

$$z^{(l)} = {\scriptstyle n_l \text{ outputs}} \left\{ \begin{bmatrix} z_1^{(l)} \\ z_2^{(l)} \\ \vdots \\ z_{n_l}^{(l)} \end{bmatrix} \right.$$

$z^{(l)}$ is the neuron 'weighted input' matrix for the lth layer

We have that:

$z^{(l)} = W^{(l)} \cdot a^{(l-1)} + b^{(l)}$

$$= n_l \text{ outputs} \left\{ \overbrace{\begin{bmatrix} w_{1,1}^{(l)} & w_{1,2}^{(l)} & \cdots & w_{1,n_{l-1}}^{(l)} \\ w_{2,1}^{(l)} & w_{2,2}^{(l)} & \cdots & w_{2,n_{l-1}}^{(l)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n_l,1}^{(l)} & w_{n_l,2}^{(l)} & \cdots & w_{n_l,n_{l-1}}^{(l)} \end{bmatrix}}^{n_{l-1} \text{ inputs}} \right. \cdot \; n_{l-1} \text{ inputs} \left\{ \begin{bmatrix} a_1^{(l-1)} \\ a_2^{(l-1)} \\ \vdots \\ a_{n_l}^{(l-1)} \end{bmatrix} \right. + \; n_l \text{ outputs} \left\{ \begin{bmatrix} b_1^{(l)} \\ b_2^{(l)} \\ \vdots \\ b_{n_l}^{(l)} \end{bmatrix} \right.$$

$$= n_l \text{ outputs} \left\{ \begin{bmatrix} z_1^{(l)} \\ z_2^{(l)} \\ \vdots \\ z_{n_l}^{(l)} \end{bmatrix} \right.$$

---

a

our Neuron Activation $a^{(l)} \in \mathbb{R}^{\text{out} \times 1} = \mathbb{R}^{n_l \times 1}$:

$$a^{(l)} = n_l \text{ outputs} \left\{ \begin{bmatrix} a_1^{(l)} \\ a_2^{(l)} \\ \vdots \\ a_{n_l}^{(l)} \end{bmatrix} \right.$$

$a^{(l)}$ is the activation matrix for the lth layer

We have that:

$$
\begin{aligned}
a^{(l)} &= g\left(z^{(l)}\right) \\
&= g\left(W^{(l)} \cdot a^{(l-1)} + b^{(l)}\right) \\
&= g\left(\text{\scriptsize n_l outputs} \left\{ \overbrace{\begin{bmatrix} w_{1,1}^{(l)} & w_{1,2}^{(l)} & \cdots & w_{1,n_{l-1}}^{(l)} \\ w_{2,1}^{(l)} & w_{2,2}^{(l)} & \cdots & w_{2,n_{l-1}}^{(l)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n_l,1}^{(l)} & w_{n_l,2}^{(l)} & \cdots & w_{n_l,n_{l-1}}^{(l)} \end{bmatrix}}^{n_{l-1}\text{ inputs}} \right. \cdot \text{\scriptsize n_{l-1} inputs} \left\{ \begin{bmatrix} a_1^{(l-1)} \\ a_2^{(l-1)} \\ \vdots \\ a_{n_l}^{(l-1)} \end{bmatrix} \right. + \text{\scriptsize n_l outputs} \left\{ \begin{bmatrix} b_1^{(l)} \\ b_2^{(l)} \\ \vdots \\ b_{n_l}^{(l)} \end{bmatrix} \right. \right) \\
&= g\left(\text{\scriptsize n_l outputs} \left\{ \begin{bmatrix} z_1^{(l)} \\ z_2^{(l)} \\ \vdots \\ z_{n_l}^{(l)} \end{bmatrix} \right. \right) \\
&= \text{\scriptsize n_l outputs} \left\{ \begin{bmatrix} a_1^{(l)} \\ a_2^{(l)} \\ \vdots \\ a_{n_l}^{(l)} \end{bmatrix} \right.
\end{aligned}
$$

Forward Propagation

a