Workshop on Neural Networks

Machine Learning and Optimization Seminar

Jake Brusca

New Jersey Institute of Technology

October 6, 2022

Start Downloading!

Setup

Information on setting up a Python environment and package management can be found in the first workshop.

In Terminal:

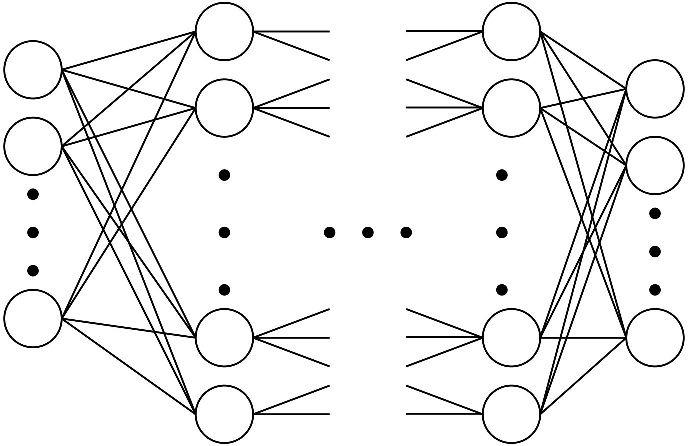
```
mamba activate workshop
mamba install numpy py-pde tensorflow matplotlib
```

You can use Google colab if unable to run local Jupyter Notebooks.

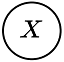
In cell:

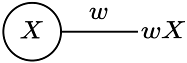
```
!pip install numpy py-pde tensorflow matplotlib
```

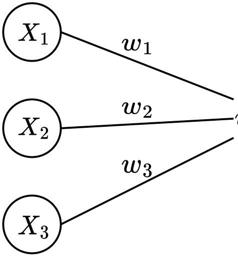
What is a Neural Network?



Components

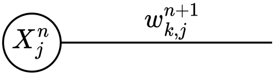




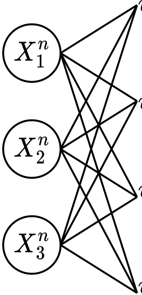


$$\vec{w} \cdot \vec{X} + b,\ \vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}, \vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

Affine Mapping



$$Z_k^{n+1}, \quad \vec{w}_k^{n+1} = \begin{bmatrix} w_{k,1}^{n+1} \\ w_{k,2}^{n+1} \\ w_{k,3}^{n+1} \end{bmatrix}, \vec{Z}^{n+1} = \begin{bmatrix} z_1^{n+1} \\ z_2^{n+1} \\ z_3^{n+1} \\ z_4^{n+1} \end{bmatrix}$$



$$\vec{w}_1^{n+1} \cdot \vec{X}^n + b_1^{n+1},$$

$$\vec{w}_2^{n+1} \cdot \vec{X}^n + b_2^{n+1},$$

$$\vec{w}_3^{n+1} \cdot \vec{X}^n + b_3^{n+1},$$

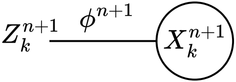
$$\vec{w}_4^{n+1} \cdot \vec{X}^n + b_4^{n+1},$$

$$\vec{b}^{n+1} = \begin{bmatrix} b_1^{n+1} \\ b_2^{n+1} \\ b_3^{n+1} \\ b_4^{n+1} \end{bmatrix}$$

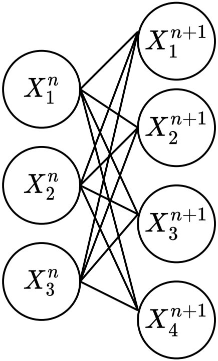
$$\underline{W}^{n+1} = \begin{bmatrix} | & | & | & | \\ \vec{w}_1^{n+1} & \vec{w}_2^{n+1} & \vec{w}_3^{n+1} & \vec{w}_4^{n+1} \\ | & | & | & | \end{bmatrix}^T$$

$$\underline{W}^{n+1}\vec{X}^n + \vec{b}^{n+1} = \vec{Z}^{n+1}$$

Nonlinear Mapping



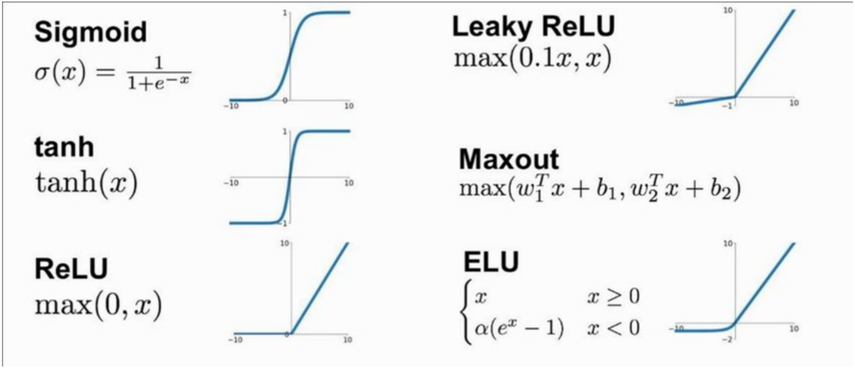
$$X_k^{n+1} := \phi^{n+1}(Z_k^{n+1})$$



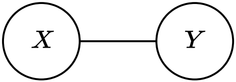
$$\vec{X}^{n+1} := \phi^{n+1}(\vec{Z}^{n+1}) = \begin{bmatrix} \phi^{n+1}(Z_1^{n+1}) \\ \phi^{n+1}(Z_2^{n+1}) \\ \phi^{n+1}(Z_3^{n+1}) \\ \phi^{n+1}(Z_4^{n+1}) \end{bmatrix}$$

$$\vec{X}^{n+1} = \phi^{n+1}(\underline{W}^{n+1}\vec{X}^n + \vec{b}^{n+1})$$

Activation Functions

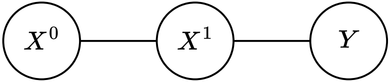


Example 1



$$Y = wX + b$$

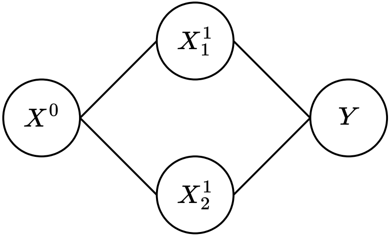
Example 2



$$Z^1 = w^1 X^0 + b^1, \quad X^1 = \phi(Z^1), \quad Y = w^2 X^1 + b^2$$

$$Y = w^2 \phi(w^1 X^0 + b^1) + b^2$$

Example 3



$$\vec{Z}^1 = \vec{w}^1 X^0 + \vec{b}^1, \quad \vec{X}^1 = \phi(\vec{Z}^1), \quad Y = \vec{w}^2 \cdot \vec{X}^1 + b^2$$

$$\begin{aligned} Y &= \vec{w}^2 \cdot \phi(\vec{w}^1 X^0 + \vec{b}^1) + b^2 \\ &= w_1^2 \phi(w_1^1 X^0 + b_1^1) + w_2^2 \phi(w_2^1 X^0 + b_2^1) + b^2 \end{aligned}$$

Remarks

- Can have arbitrary number of hidden layers
- Hidden layers can have arbitrary depth, independent of each other
- Each hidden layer can have it's own activation function
- Activation function on output layer $\vec{Y}$ is optional, depneds on context
- In general, a NN with L hidden layers can be represented as

$$
\begin{aligned}
Y &= \underline{W}^{L+1}\vec{X}^{L} + \vec{b}^{L+1} \\
&= \underline{W}^{L+1}\phi^{L}(\underline{W}^{L}\vec{X}^{L-1} + \vec{b}^{L}) + \vec{b}^{L+1} \\
&\vdots \\
&= \underline{W}^{L+1}\phi^{L}(\underline{W}^{L}\phi^{L-1}(\ldots\phi^{1}(\underline{W}^{1}\vec{X}^{0} + \vec{b}^{1})\ldots) + \vec{b}^{L-1}) + \vec{b}^{L}) + \vec{b}^{L+1}
\end{aligned}
$$

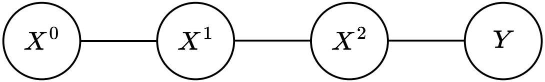
Training

$$\overline{X} = \{X_1, ..., X_N\}, \overline{Y} = \{\hat{Y}_1, ..., \hat{Y}_N\}$$

$$(X_i, \hat{Y}_i) \Rightarrow E_i(\beta) = (Y_i - \hat{Y}_i)^2 \Rightarrow E(\beta) = \sum_{i=1}^{N} E_i(\beta)$$

$$\begin{aligned} \beta_{n+1} &= \beta_n - \gamma_n \nabla E(\beta_n) \\ &= \beta_n - \gamma_n \sum_{i=1}^{N} \nabla E_i(\beta_n) \end{aligned}$$

Propagation



$$E_i = (Y_i - \hat{Y}_i)^2, \qquad \frac{\partial E_i}{\partial Y_i} = 2(Y_i - \hat{Y}_i)$$

$$Y_i = w^3 X_i^2 + b^3, \qquad \frac{\partial Y_i}{\partial w^3} = X_i^2, \ \frac{\partial Y_i}{\partial b^3} = 1$$

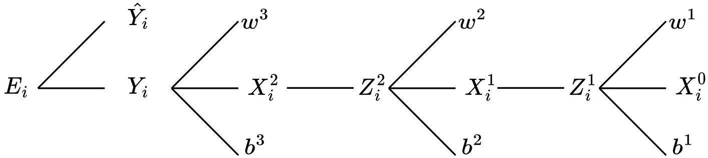
$$X_i^2 = \phi^2(Z_i^2), \qquad \frac{\partial X_i^2}{\partial Z_i^2} = \dot{\phi}^2(Z_i^2),$$

$$Z_i^2 = w^2 X_i^1 + b^2, \qquad \frac{\partial Z_i^2}{\partial w^2} = X_i^2, \ \frac{\partial Z_i^2}{\partial b^2} = 1,$$

$$X_i^1 = \phi^1(Z_i^1), \qquad \frac{\partial X_i^1}{\partial X_i^1} = \dot{\phi}^1(Z_i^1),$$

$$Z_i^1 = w^1 X^0 + b^1, \qquad \frac{\partial Z_i^1}{\partial w^1} = X_i^0, \ \frac{\partial Z_i^1}{\partial b^1} = 1$$

Chain Rule



$$\frac{\partial E_i}{\partial w^3} = \frac{\partial E_i}{\partial Y_i}\frac{\partial Y_i}{\partial w^3}$$

$$\frac{\partial E_i}{\partial b^3} = \frac{\partial E_i}{\partial Y_i}\frac{\partial Y_i}{\partial b^3}$$

$$\frac{\partial E_i}{\partial w^2} = \frac{\partial E_i}{\partial Y_i}\frac{\partial Y_i}{\partial X_i^2}\frac{\partial X_i^2}{\partial Z_i^2}\frac{\partial Z_i^2}{\partial w^2}$$

$$\frac{\partial E_i}{\partial b^2} = \frac{\partial E_i}{\partial Y_i}\frac{\partial Y_i}{\partial X_i^2}\frac{\partial X_i^2}{\partial Z_i^2}\frac{\partial Z_i^2}{\partial b^2}$$

$$\frac{\partial E_i}{\partial w^1} = \frac{\partial E_i}{\partial Y_i}\frac{\partial Y_i}{\partial X_i^2}\frac{\partial X_i^2}{\partial Z_i^2}\frac{\partial Z_i^2}{\partial X_i^1}\frac{\partial X_i^1}{\partial Z_i^1}\frac{\partial Z_i^1}{\partial w^1}$$

$$\frac{\partial E_i}{\partial b^1} = \frac{\partial E_i}{\partial Y_i}\frac{\partial Y_i}{\partial X_i^2}\frac{\partial X_i^2}{\partial Z_i^2}\frac{\partial Z_i^2}{\partial X_i^1}\frac{\partial X_i^1}{\partial Z_i^1}\frac{\partial Z_i^1}{\partial b^1}$$