AIL721: Deep Learning

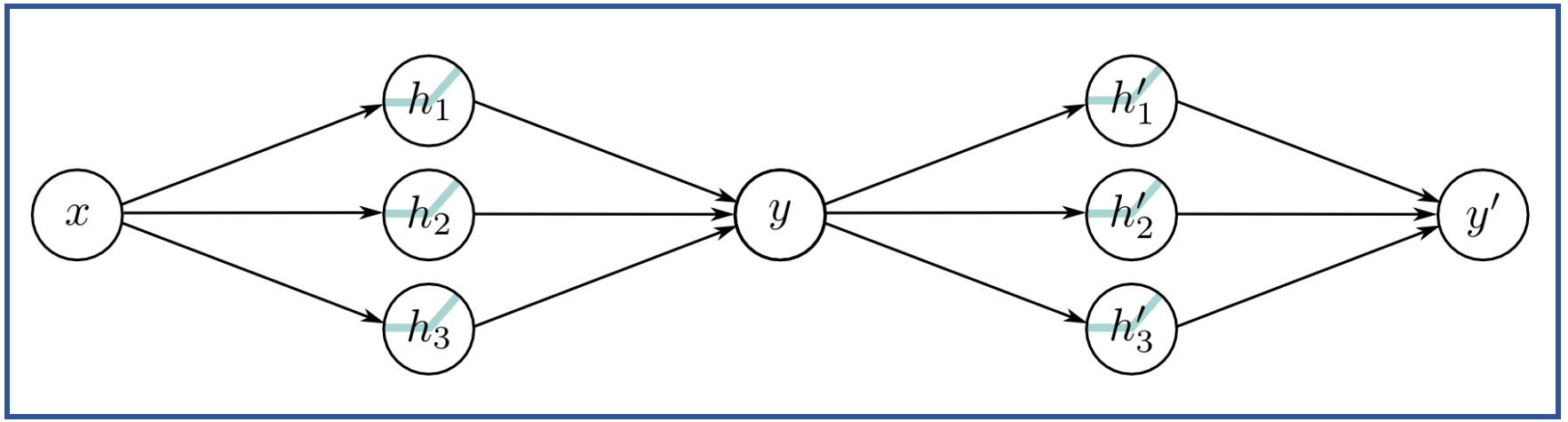
Instructor: James Arambam

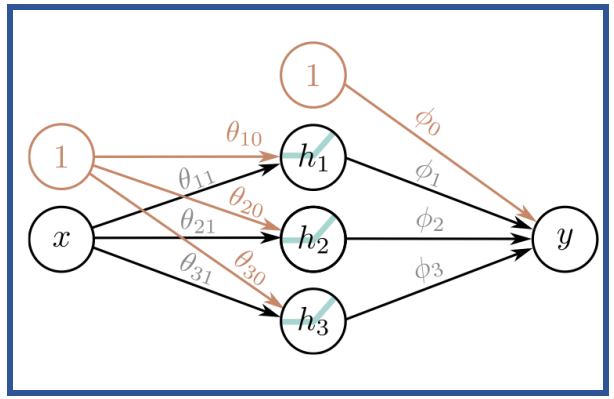
ScAI | Yardi School of Artificial Intelligence
Indian Institute of Technology Delhi

Class Announcements

- **Please read the reference textbooks.**
  - Slides are not class notes.
  - Does not contain full information.

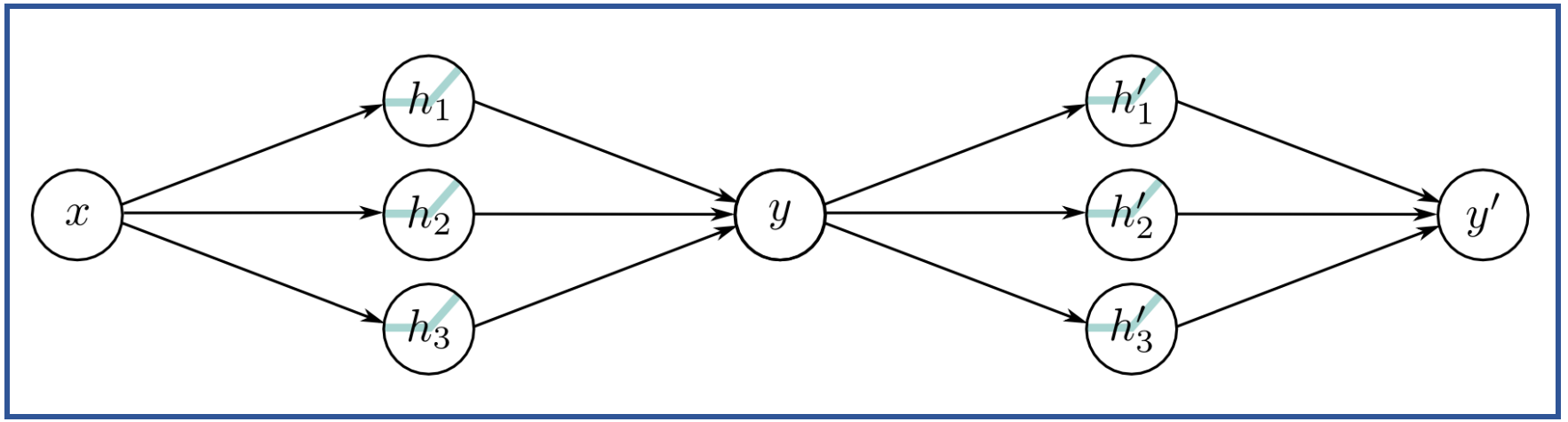
Deep Neural Network

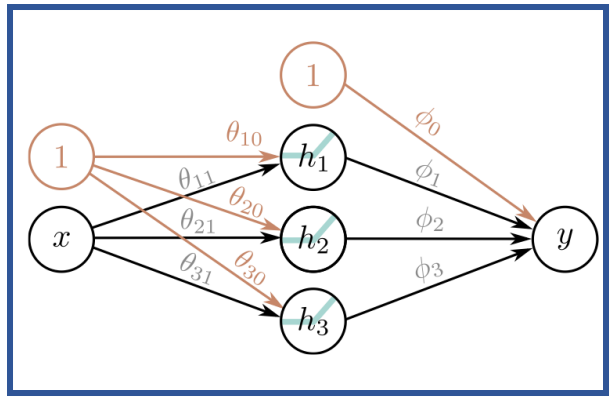




Write the neural network equation.

Deep Neural Network





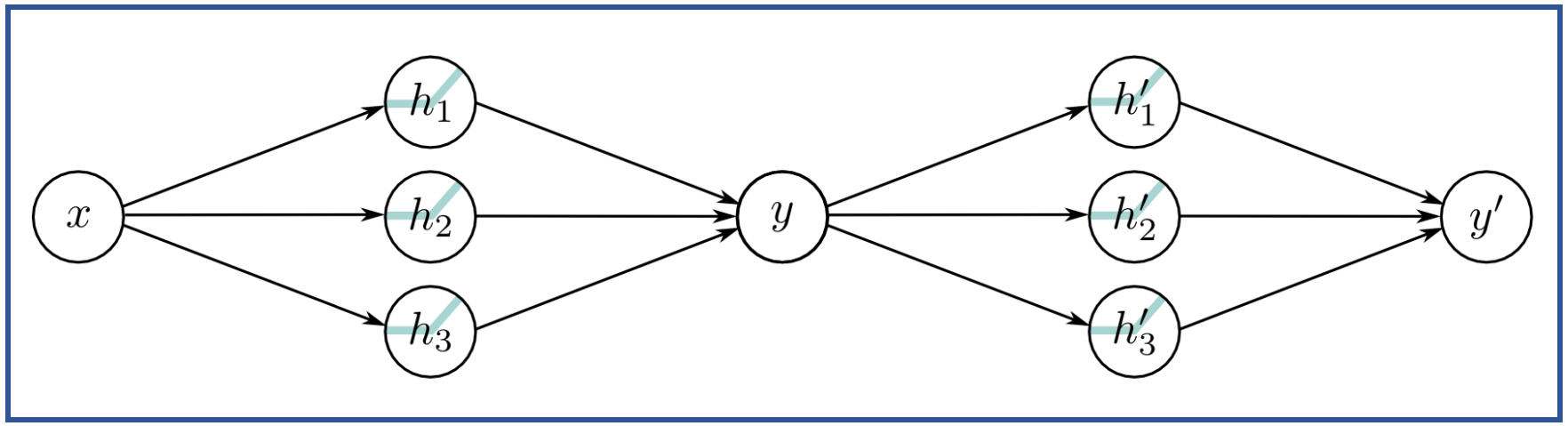
$$
\begin{aligned}
h_1 &= \mathrm{a}[\theta_{10} + \theta_{11}x] \\
h_2 &= \mathrm{a}[\theta_{20} + \theta_{21}x] \\
h_3 &= \mathrm{a}[\theta_{30} + \theta_{31}x],
\end{aligned}
$$

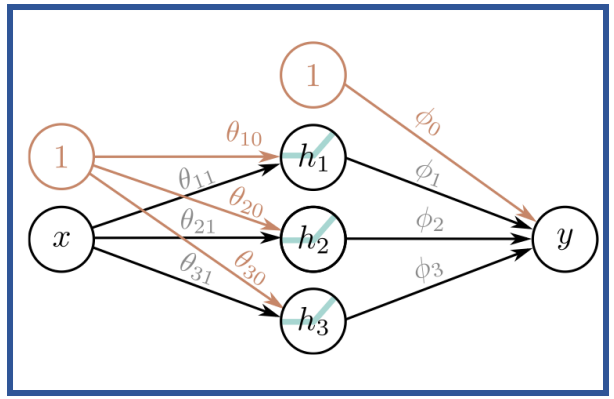
$$
\begin{aligned}
h'_1 &= \mathrm{a}[\theta'_{10} + \theta'_{11}y] \\
h'_2 &= \mathrm{a}[\theta'_{20} + \theta'_{21}y] \\
h'_3 &= \mathrm{a}[\theta'_{30} + \theta'_{31}y],
\end{aligned}
$$

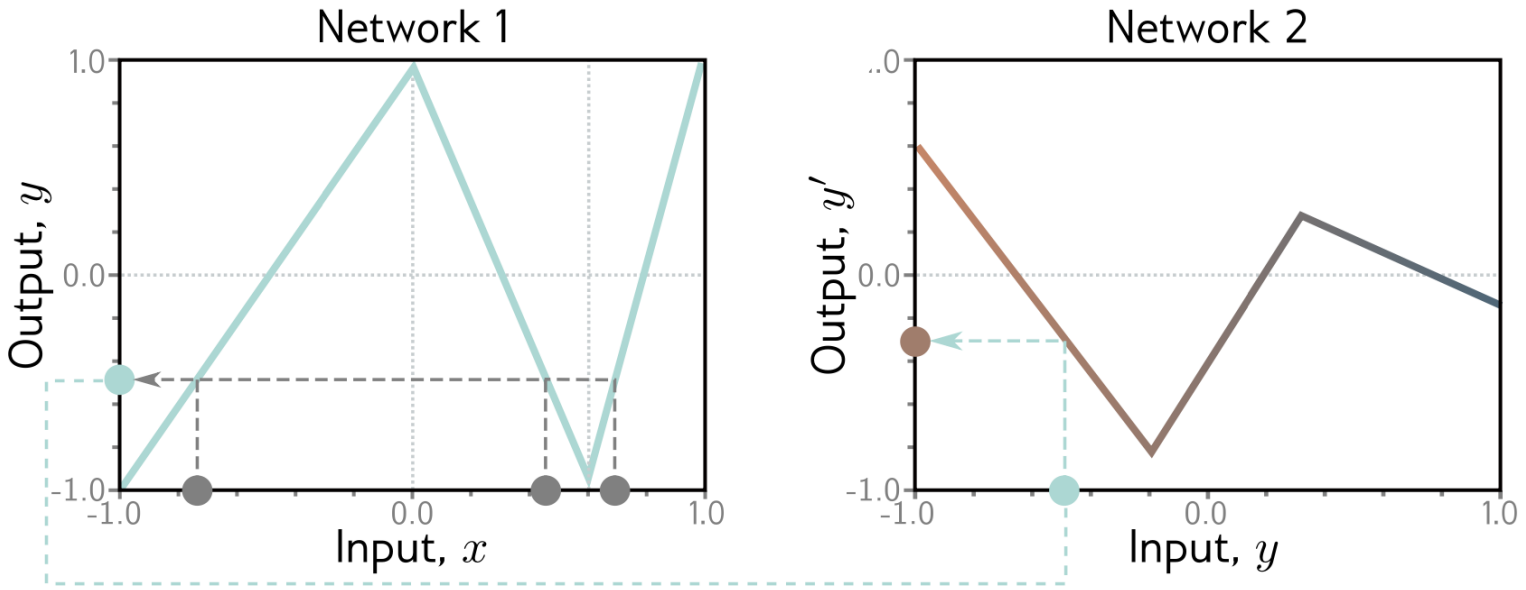
$$y = \phi_0 + \phi_1 h_1 + \phi_2 h_2 + \phi_3 h_3$$

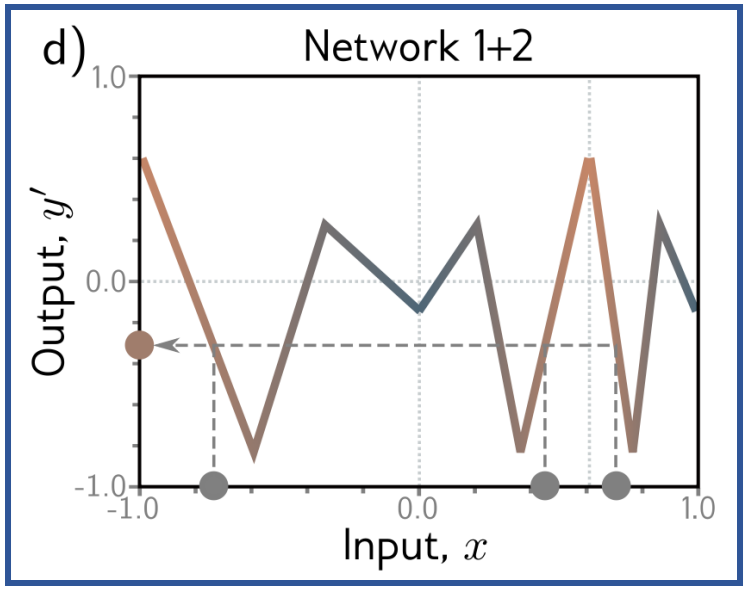
$$y' = \phi'_0 + \phi'_1 h'_1 + \phi'_2 h'_2 + \phi'_3 h'_3$$

Deep Neural Network



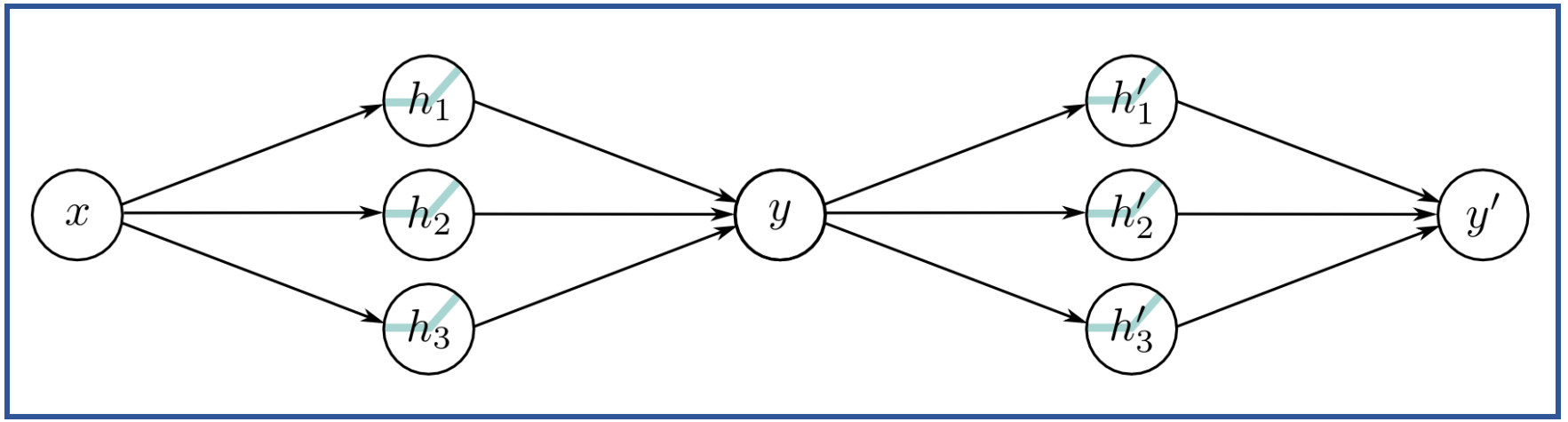


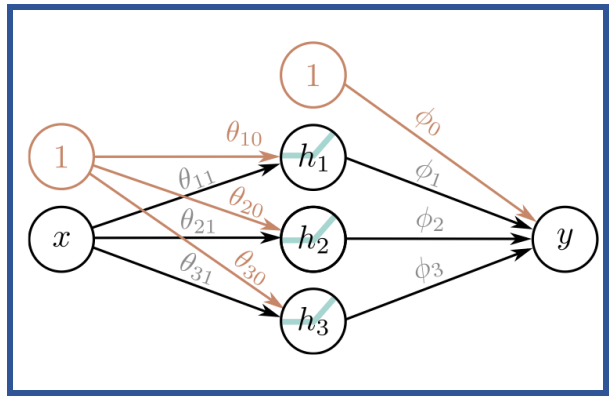




Example

Deep Neural Network





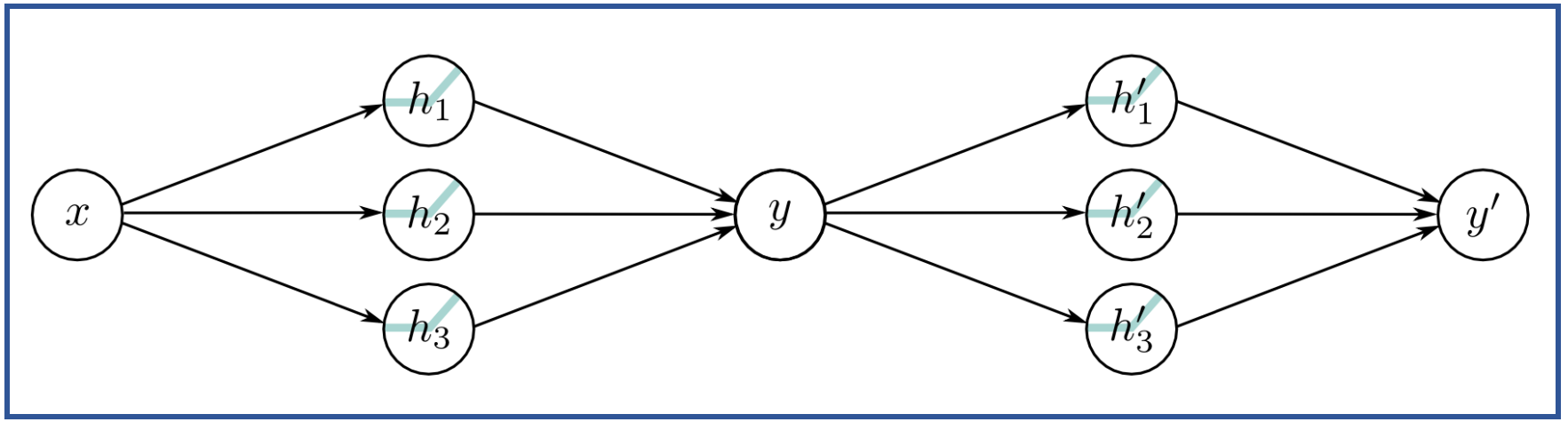
$$
\begin{aligned}
h_1 &= \mathrm{a}[\theta_{10} + \theta_{11}x] \\
h_2 &= \mathrm{a}[\theta_{20} + \theta_{21}x] \\
h_3 &= \mathrm{a}[\theta_{30} + \theta_{31}x],
\end{aligned}
$$

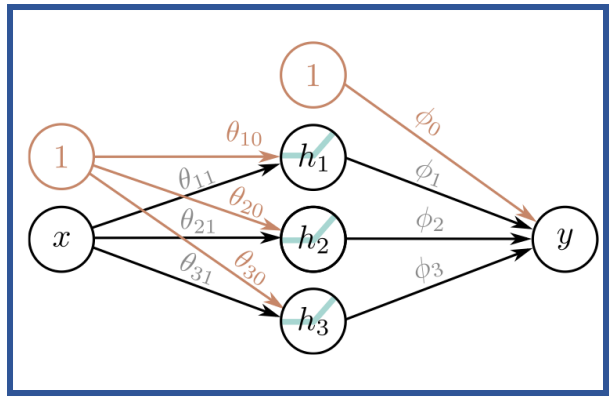
$$
\begin{aligned}
h'_1 &= \mathrm{a}[\theta'_{10} + \theta'_{11}y] \\
h'_2 &= \mathrm{a}[\theta'_{20} + \theta'_{21}y] \\
h'_3 &= \mathrm{a}[\theta'_{30} + \theta'_{31}y],
\end{aligned}
$$

$$y = \phi_0 + \phi_1 h_1 + \phi_2 h_2 + \phi_3 h_3$$

$$y' = \phi'_0 + \phi'_1 h'_1 + \phi'_2 h'_2 + \phi'_3 h'_3$$

Deep Neural Network

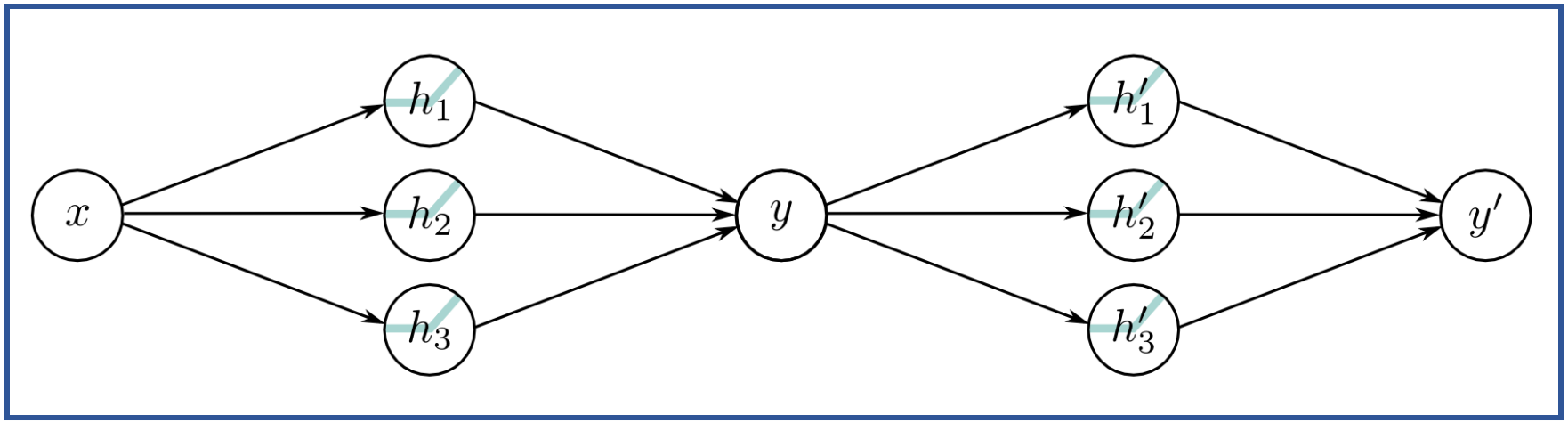


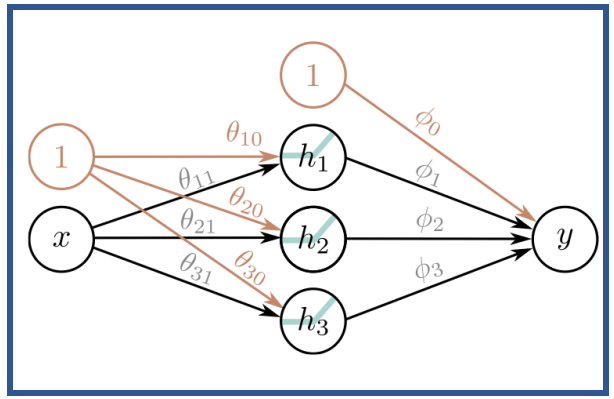


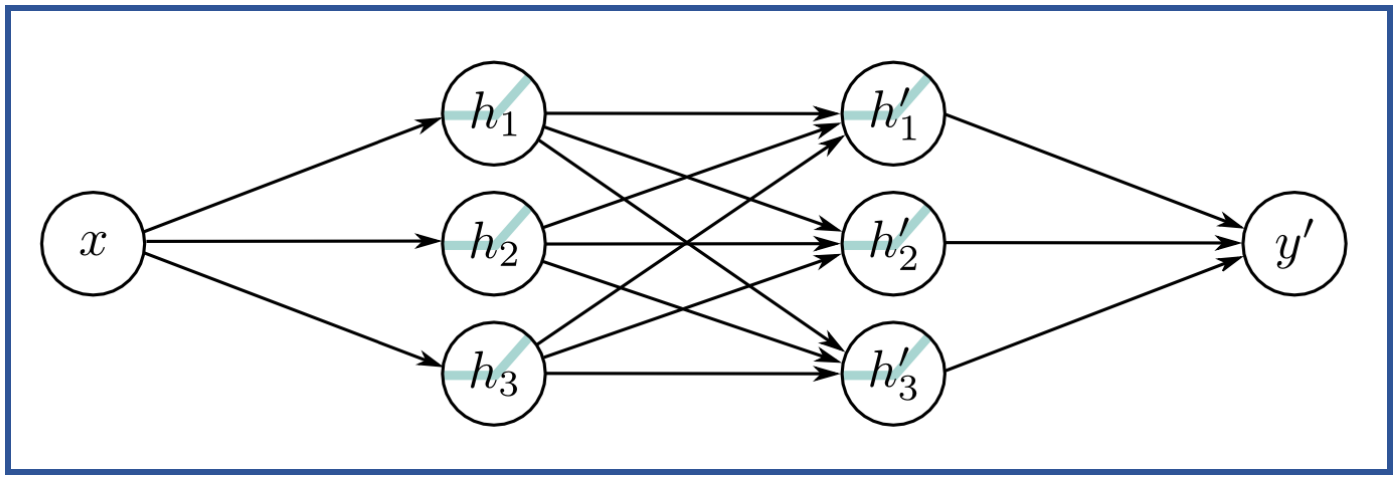
$$
\begin{aligned}
h'_1 &= \mathrm{a}[\theta'_{10} + \theta'_{11}y] = \mathrm{a}[\theta'_{10} + \theta'_{11}\phi_0 + \theta'_{11}\phi_1 h_1 + \theta'_{11}\phi_2 h_2 + \theta'_{11}\phi_3 h_3] \\
h'_2 &= \mathrm{a}[\theta'_{20} + \theta'_{21}y] = \mathrm{a}[\theta'_{20} + \theta'_{21}\phi_0 + \theta'_{21}\phi_1 h_1 + \theta'_{21}\phi_2 h_2 + \theta'_{21}\phi_3 h_3] \\
h'_3 &= \mathrm{a}[\theta'_{30} + \theta'_{31}y] = \mathrm{a}[\theta'_{30} + \theta'_{31}\phi_0 + \theta'_{31}\phi_1 h_1 + \theta'_{31}\phi_2 h_2 + \theta'_{31}\phi_3 h_3],
\end{aligned}
$$

$$
\begin{aligned}
h'_1 &= \mathrm{a}[\psi_{10} + \psi_{11}h_1 + \psi_{12}h_2 + \psi_{13}h_3] \quad \text{where } \psi_{10} = \theta'_{10} + \theta'_{11}\phi_0, \psi_{11} = \theta'_{11}\phi_1, \psi_{12} = \theta'_{11}\phi_2 \\
h'_2 &= \mathrm{a}[\psi_{20} + \psi_{21}h_1 + \psi_{22}h_2 + \psi_{23}h_3] \\
h'_3 &= \mathrm{a}[\psi_{30} + \psi_{31}h_1 + \psi_{32}h_2 + \psi_{33}h_3],
\end{aligned}
$$

Deep Neural Network



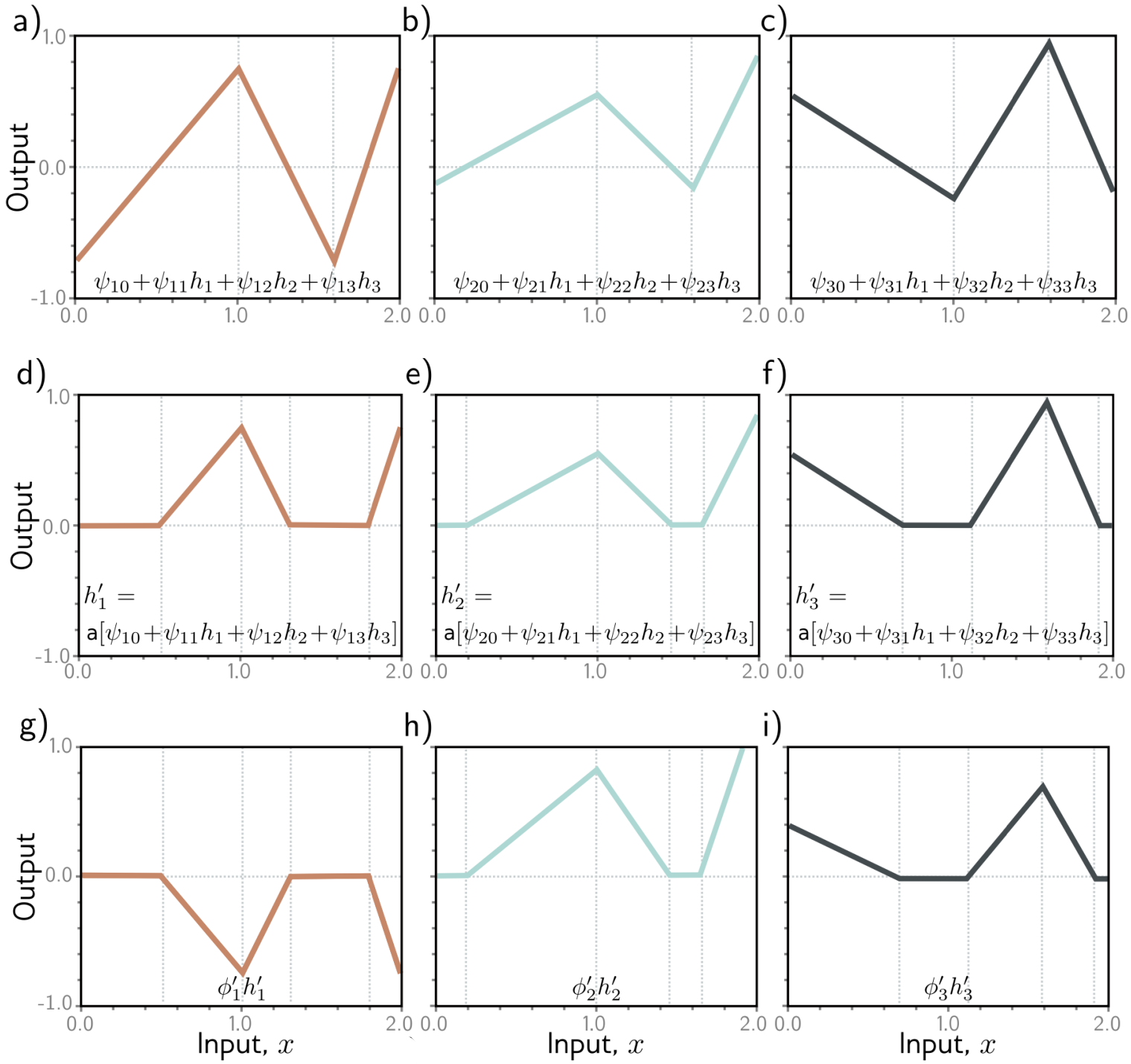




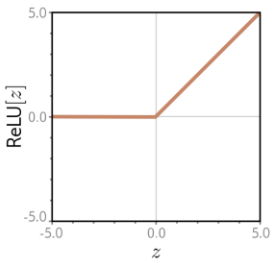
$$
\begin{aligned}
h_1 &= \mathrm{a}[\theta_{10} + \theta_{11}x] \\
h_2 &= \mathrm{a}[\theta_{20} + \theta_{21}x] \\
h_3 &= \mathrm{a}[\theta_{30} + \theta_{31}x],
\end{aligned}
$$

$$
\begin{aligned}
h'_1 &= \mathrm{a}[\psi_{10} + \psi_{11}h_1 + \psi_{12}h_2 + \psi_{13}h_3] \\
h'_2 &= \mathrm{a}[\psi_{20} + \psi_{21}h_1 + \psi_{22}h_2 + \psi_{23}h_3] \\
h'_3 &= \mathrm{a}[\psi_{30} + \psi_{31}h_1 + \psi_{32}h_2 + \psi_{33}h_3],
\end{aligned}
$$

$$y' = \phi'_0 + \phi'_1 h'_1 + \phi'_2 h'_2 + \phi'_3 h'_3$$

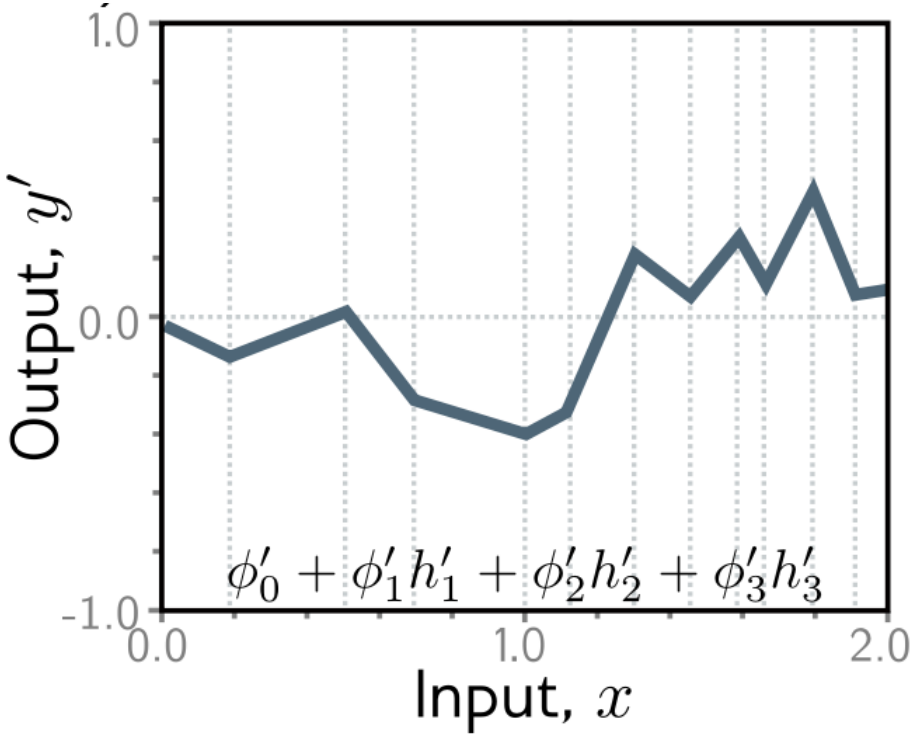


$a[\cdot]$



ReLU Activation

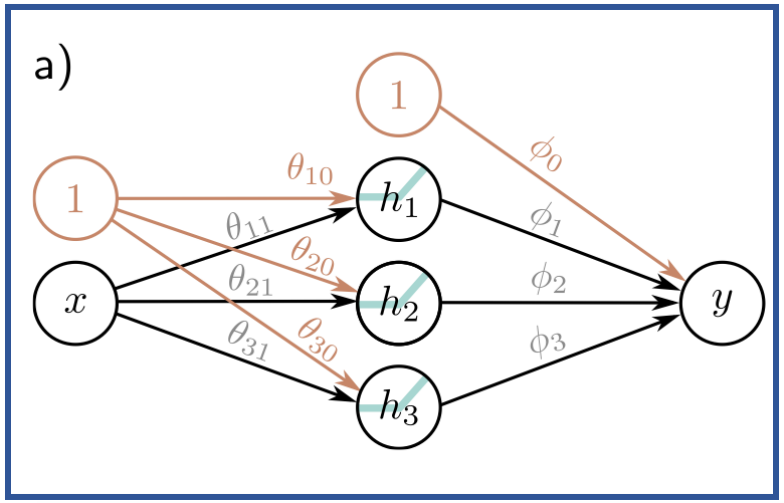
$$y' = \phi_0' + \phi_1'h_1' + \phi_2'h_2' + \phi_3'h_3'$$



Deep Neural Network

ScAI

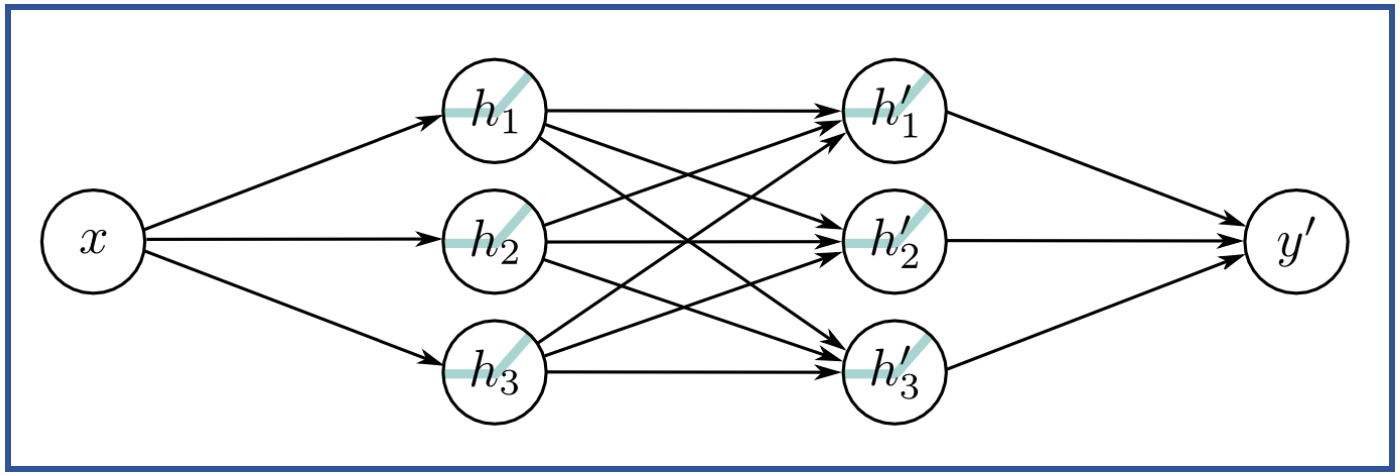
❑ Shallow Network:



How many parameters?

3D+1

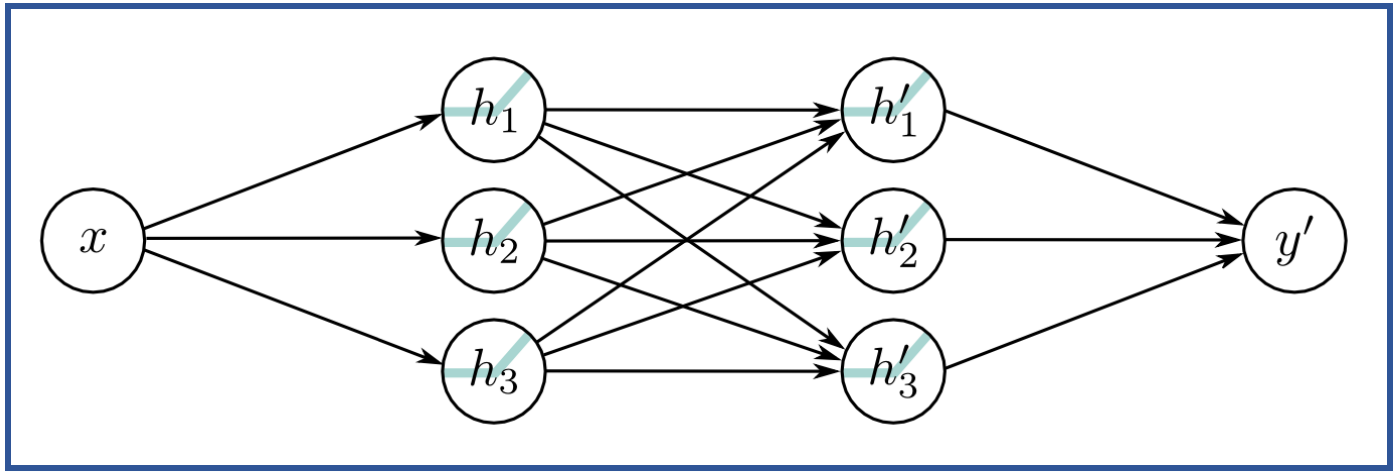
❑ Deep Network:



How many parameters?

3D+(K-1)D^2+(K-1)D+1

Deep Neural Network



$$
\begin{aligned}
h_1 &= \mathrm{a}[\theta_{10} + \theta_{11}x] \\
h_2 &= \mathrm{a}[\theta_{20} + \theta_{21}x] \\
h_3 &= \mathrm{a}[\theta_{30} + \theta_{31}x],
\end{aligned}
$$

$$
\begin{aligned}
h'_1 &= \mathrm{a}[\psi_{10} + \psi_{11}h_1 + \psi_{12}h_2 + \psi_{13}h_3] \\
h'_2 &= \mathrm{a}[\psi_{20} + \psi_{21}h_1 + \psi_{22}h_2 + \psi_{23}h_3] \\
h'_3 &= \mathrm{a}[\psi_{30} + \psi_{31}h_1 + \psi_{32}h_2 + \psi_{33}h_3],
\end{aligned}
$$

$$y' = \phi'_0 + \phi'_1 h'_1 + \phi'_2 h'_2 + \phi'_3 h'_3$$

Hyperparameters?

- ❑ Number of layers.
- ❑ Number of hidden units.
- ❑ Learning rate
- ❑ Batch size or mini-batch size.

Hyperparameters? → Family of functions

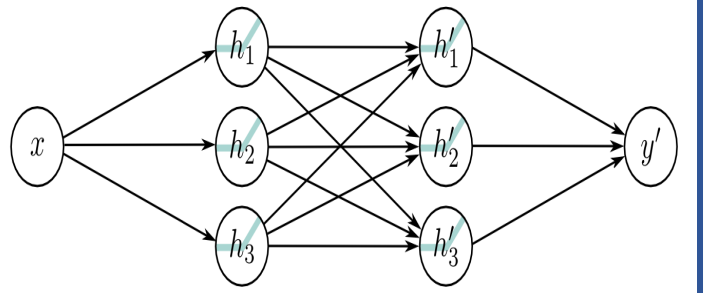
Parameters? → A particular function

Deep Neural Network

❑ Matrix notation

$$\begin{aligned} h_1 &= \mathrm{a}[\theta_{10} + \theta_{11}x] \\ h_2 &= \mathrm{a}[\theta_{20} + \theta_{21}x] \\ h_3 &= \mathrm{a}[\theta_{30} + \theta_{31}x], \end{aligned}$$

$$\begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = \mathbf{a}\left[\begin{bmatrix} \theta_{10} \\ \theta_{20} \\ \theta_{30} \end{bmatrix} + \begin{bmatrix} \theta_{11} \\ \theta_{21} \\ \theta_{31} \end{bmatrix} x\right]$$



$$\begin{aligned} h'_1 &= \mathrm{a}[\psi_{10} + \psi_{11}h_1 + \psi_{12}h_2 + \psi_{13}h_3] \\ h'_2 &= \mathrm{a}[\psi_{20} + \psi_{21}h_1 + \psi_{22}h_2 + \psi_{23}h_3] \\ h'_3 &= \mathrm{a}[\psi_{30} + \psi_{31}h_1 + \psi_{32}h_2 + \psi_{33}h_3], \end{aligned}$$

$$\begin{bmatrix} h'_1 \\ h'_2 \\ h'_3 \end{bmatrix} = \mathbf{a}\left[\begin{bmatrix} \psi_{10} \\ \psi_{20} \\ \psi_{30} \end{bmatrix} + \begin{bmatrix} \psi_{11} & \psi_{12} & \psi_{13} \\ \psi_{21} & \psi_{22} & \psi_{23} \\ \psi_{31} & \psi_{32} & \psi_{33} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}\right]$$

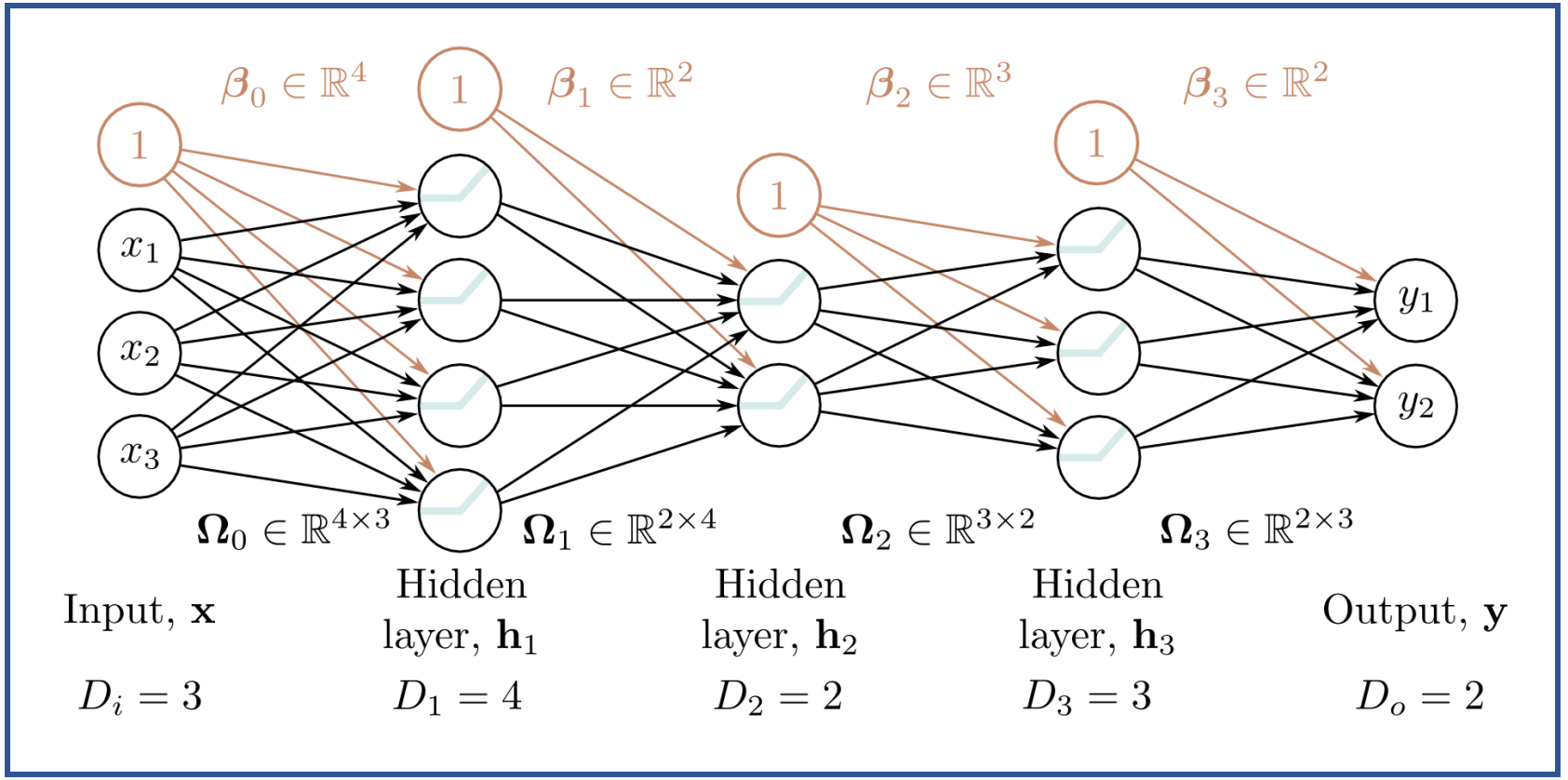
$$y' = \phi'_0 + \phi'_1 h'_1 + \phi'_2 h'_2 + \phi'_3 h'_3$$

$$y' = \phi'_0 + \begin{bmatrix} \phi'_1 & \phi'_2 & \phi'_3 \end{bmatrix} \begin{bmatrix} h'_1 \\ h'_2 \\ h'_3 \end{bmatrix}$$

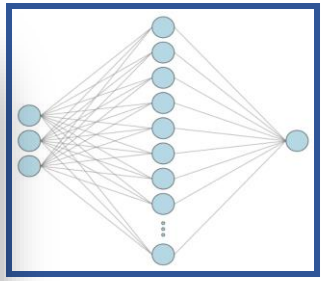
$$\begin{aligned} \mathbf{h} &= \mathbf{a}\left[\boldsymbol{\theta}_0 + \boldsymbol{\theta}x\right] \\ \mathbf{h}' &= \mathbf{a}\left[\boldsymbol{\psi}_0 + \boldsymbol{\Psi}\mathbf{h}\right] \\ y' &= \phi'_0 + \boldsymbol{\phi}'\mathbf{h}', \end{aligned}$$

Deep Neural Network

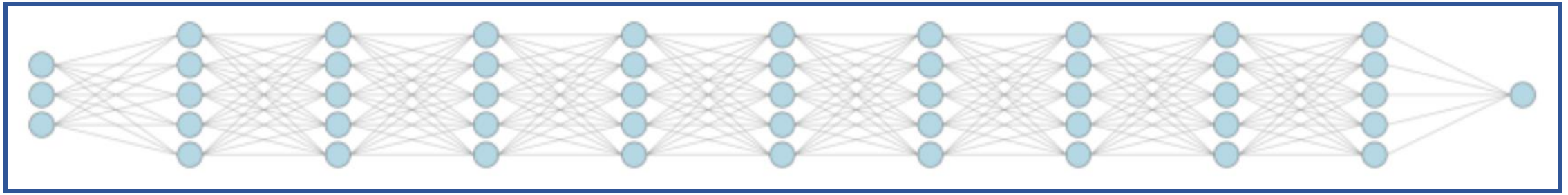
❑ Matrix notation



Deep Neural Network

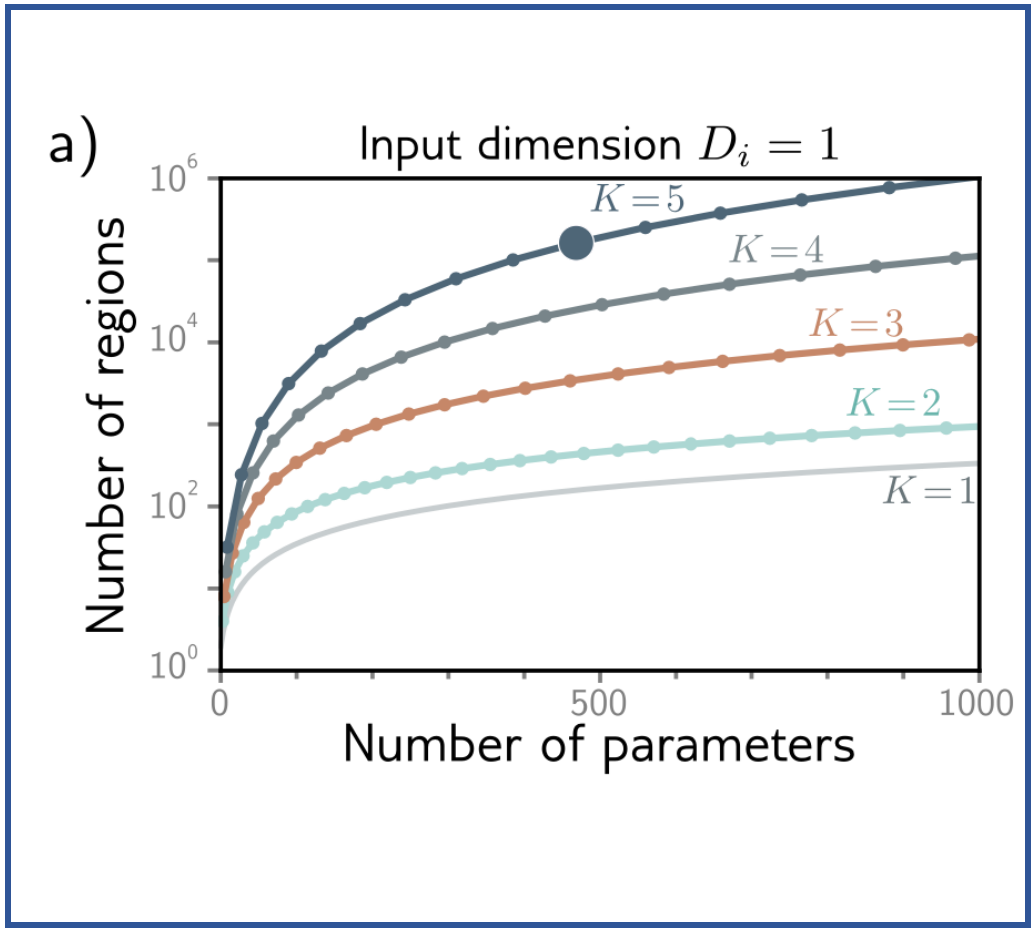


- A network with **two layers of learnable parameters - universal approximation** capabilities.
- A network with **more than two layers** – can represent a given function with **far fewer parameters**.



Paper: On the Number of Linear Regions of Deep Neural Networks. Mont´ufar et al. NeurIPS-2014

Deep Neural Network



Deep Neural Network

Key Benefits of Deep Neural Network

- Hierarchical Representation.
- Representation Learning.

Deep Neural Network

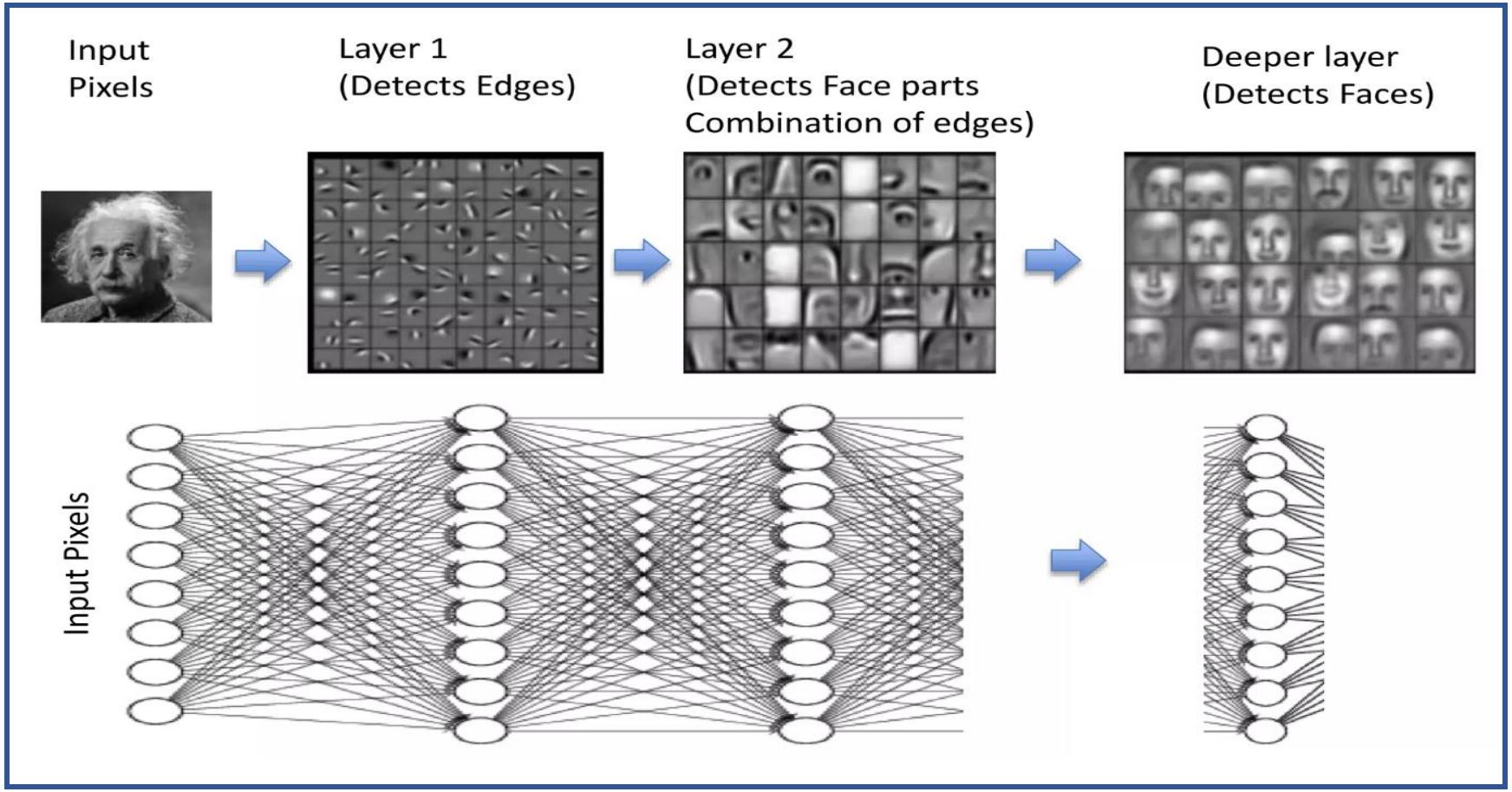
❑ Hierarchical Representation

- **Network architecture** encodes a particular form of **inductive bias**.
- **Outputs** are related to **input space** through **hierarchical representation**.
- For example, in the image classification task.

[Image from Prince 2023]

Deep Neural Network

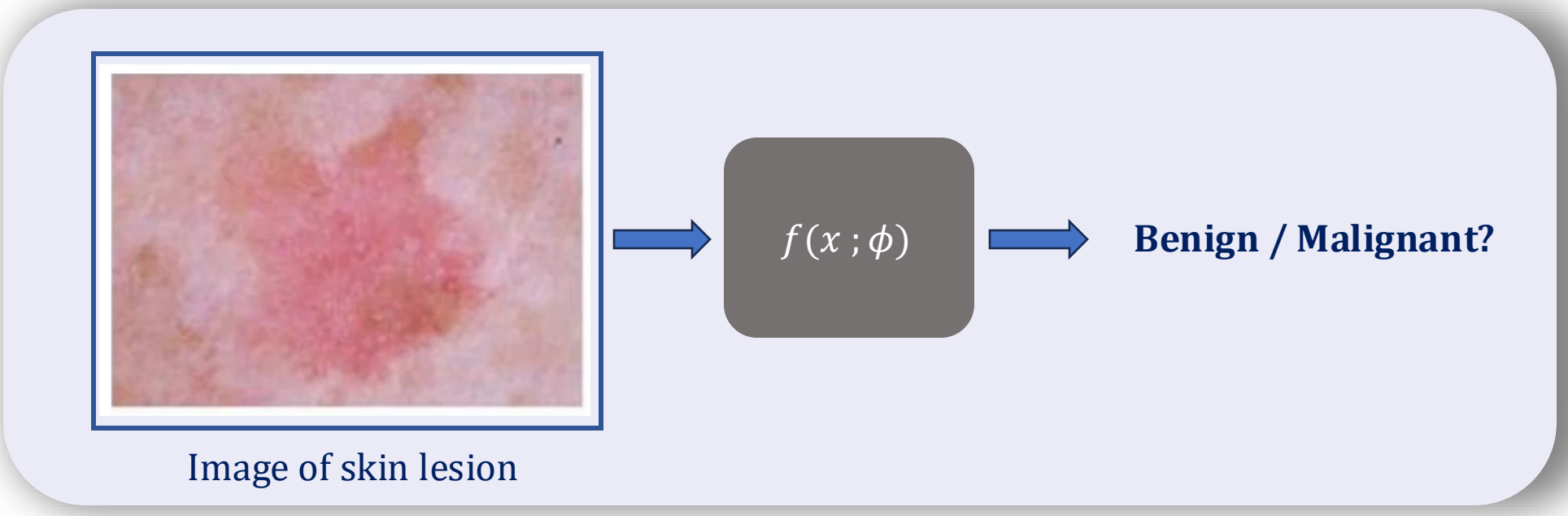
❑ Hierarchical Representation



[Image from: https://www.slideshare.net/slideshow/deep-learning-towards-general-artificial-intelligence/61772281#20]

Deep Neural Network

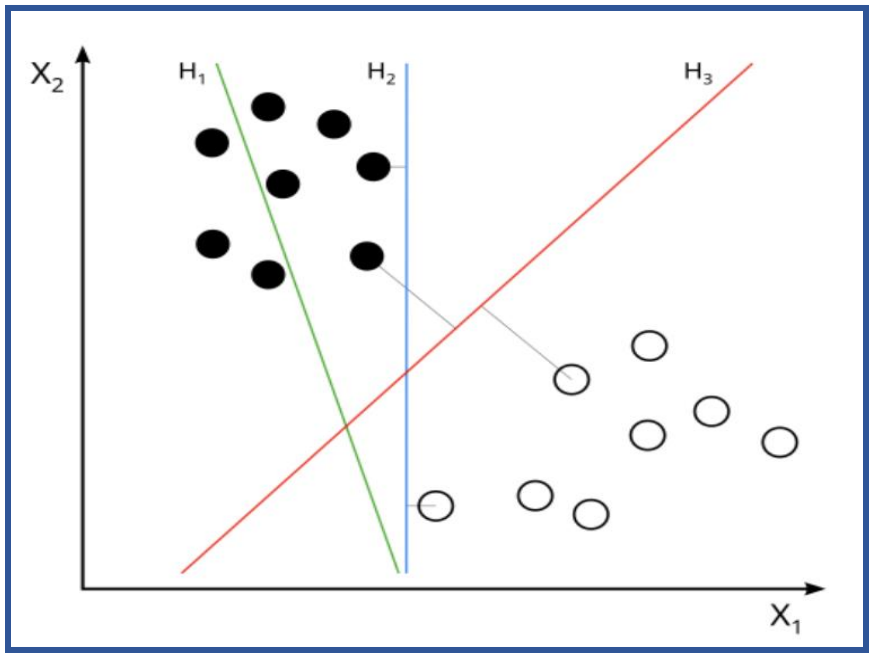
❑ Representation Learning

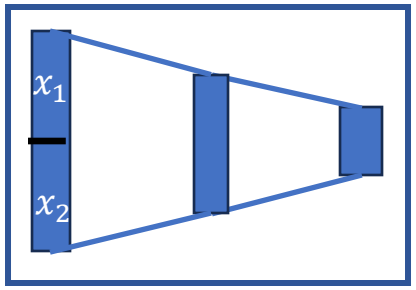


[Image from Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. Shetty et al. Nature 2022]

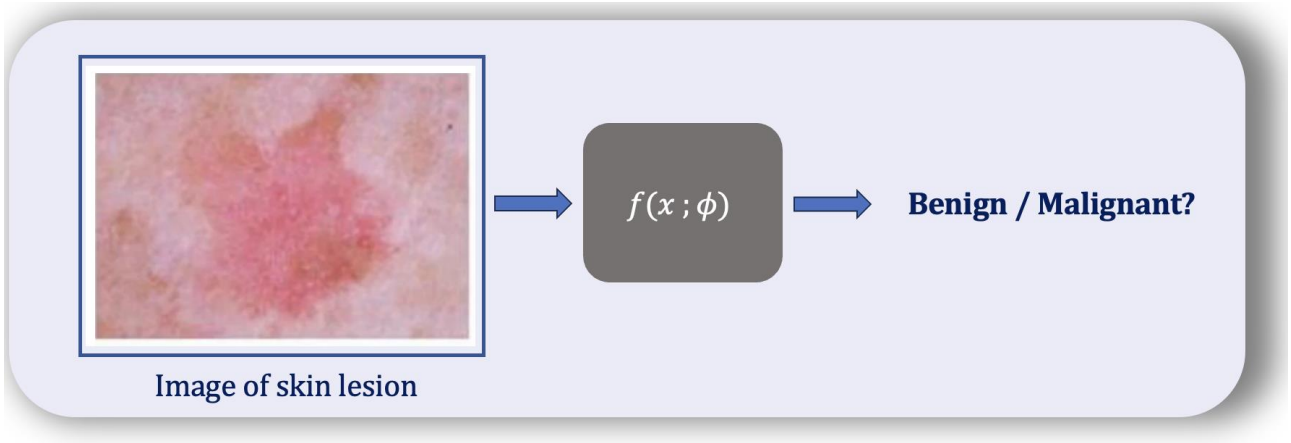
Deep Neural Network

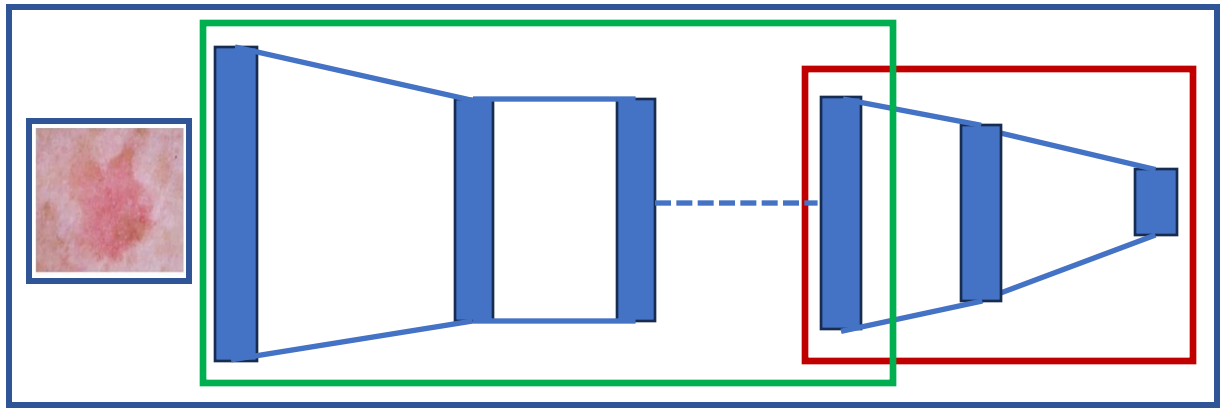
❑ Representation Learning





Shallow Neural Network



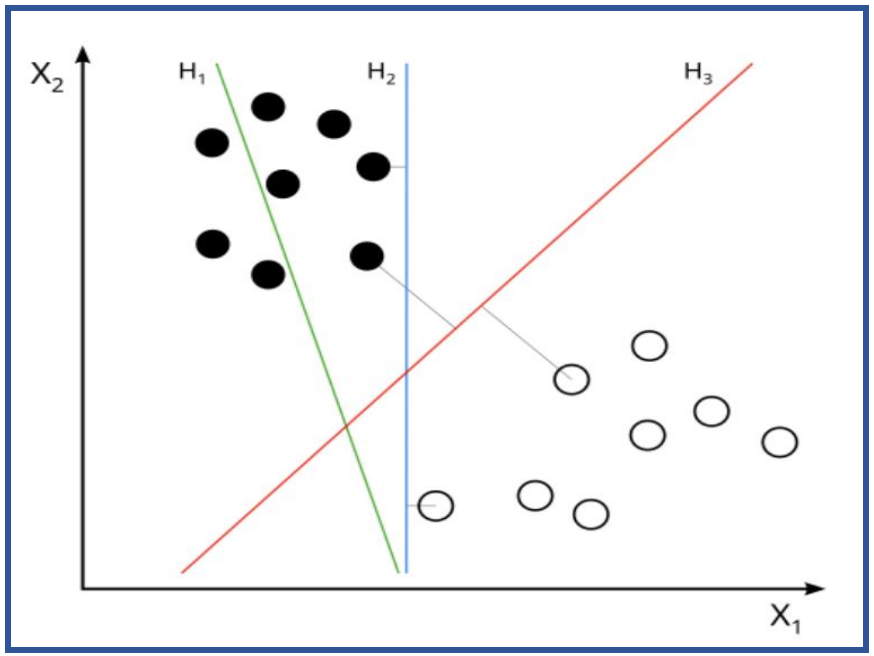


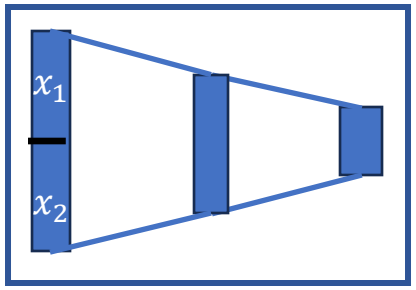
Deep Neural Network

[Image from Wikipedia]

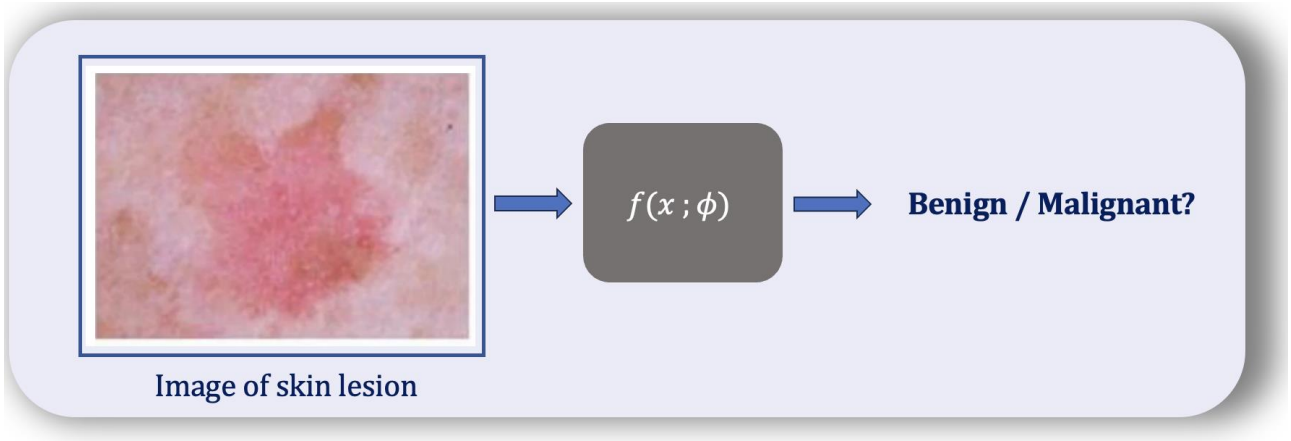
Deep Neural Network

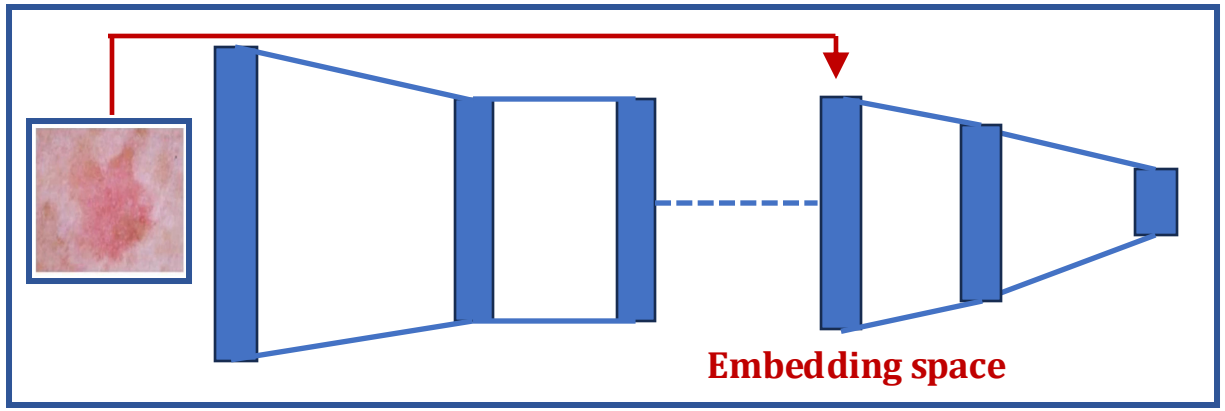
❑ Representation Learning





Shallow Neural Network





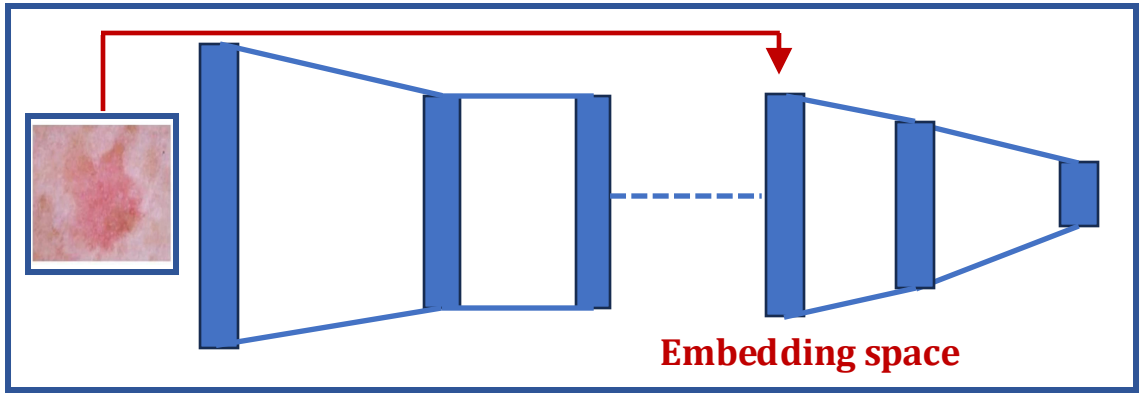
Deep Neural Network

[Image from Wikipedia]

Deep Neural Network

Representation Learning:

- Machine learning technique that helps systems automatically learn to represent raw data for subsequent tasks.



Deep Neural Network

[Image from Wikipedia]

End