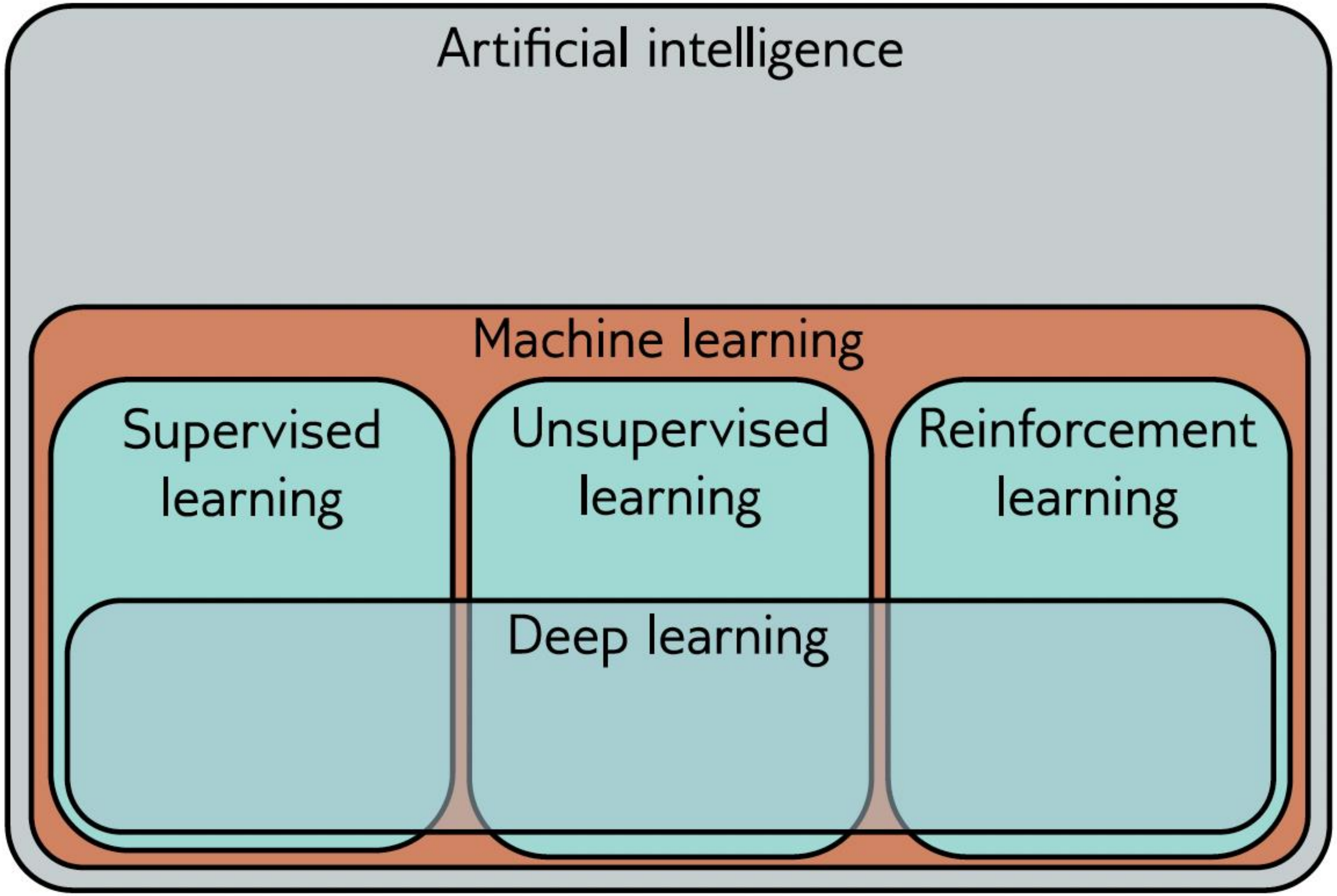
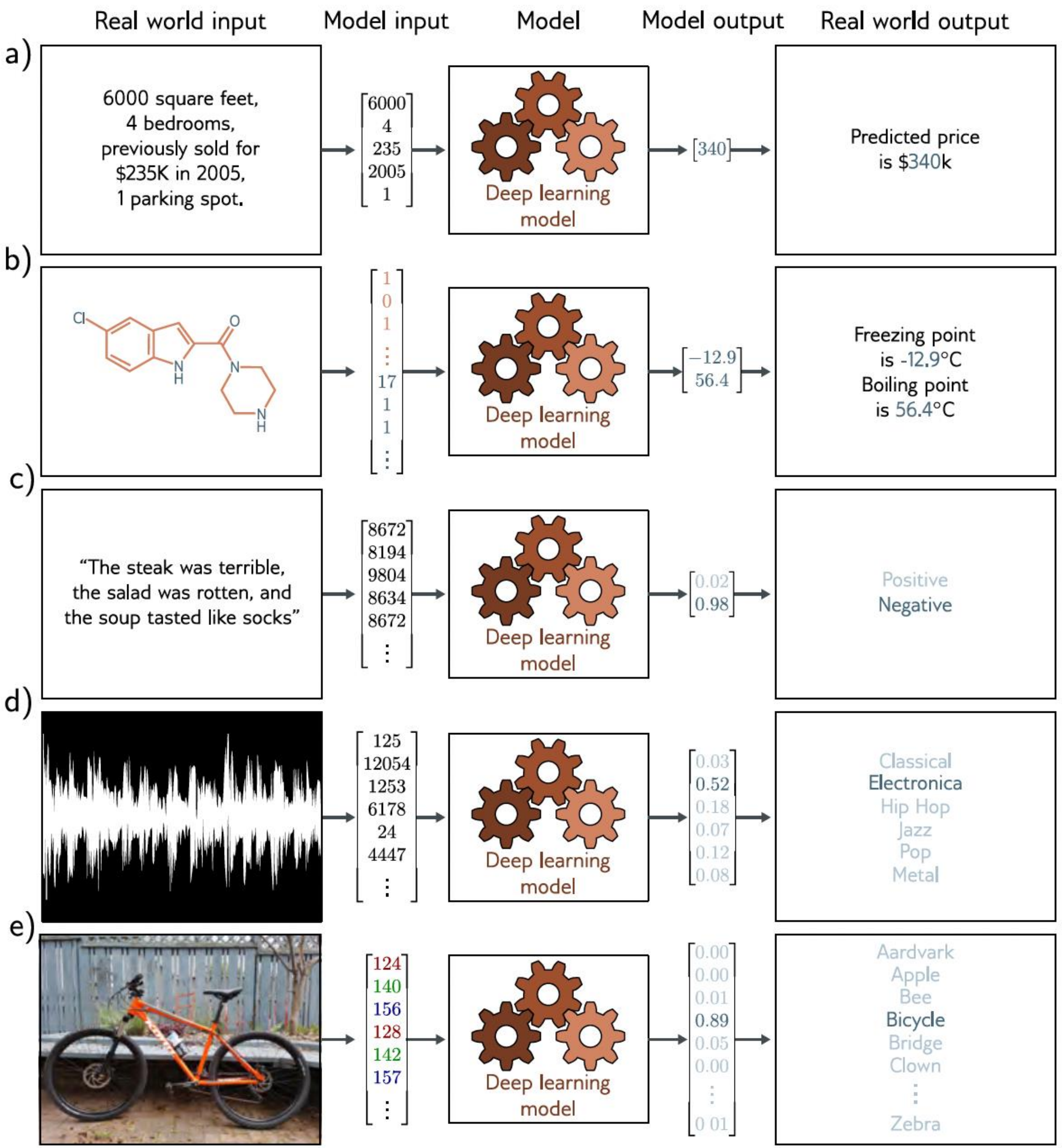
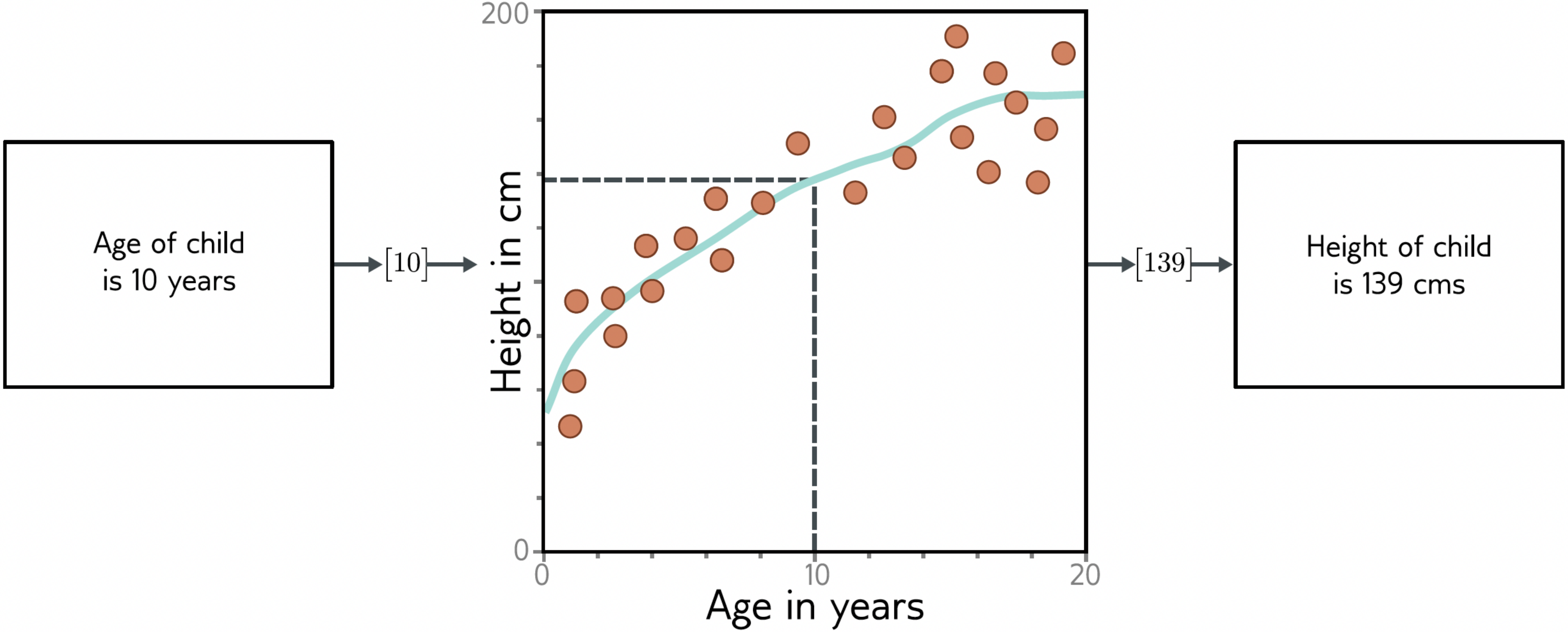
Recap and beyond



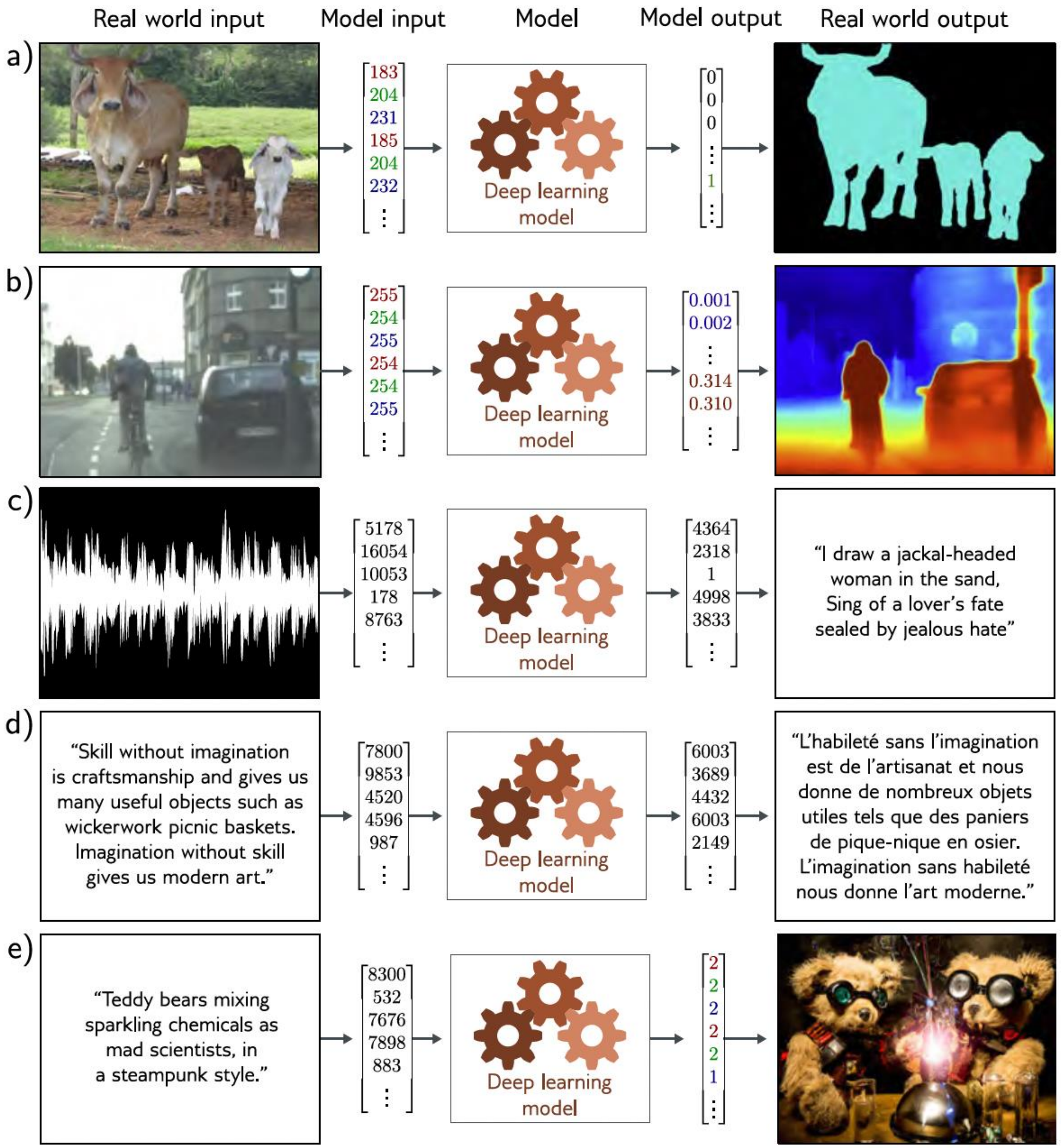
Regression vs classification problems




Age of child
is 10 years
[10]
200
0
Height in cm
0
10
20
Age in years
[139]
Height of child
is 139 cms

More complex cases

let that sink in



Supervised Learning – General Formulation

In supervised learning, we aim to build a model that takes an input x and outputs a prediction y

$$\mathbf{y} = \mathbf{f}[\mathbf{x}]$$

Supervised Learning – General Formulation

Regardless of the approach: neural networks, random forest, support-vector machines, etc. Our model have parameters

$$\mathbf{y} = \mathbf{f}[\mathbf{x}, \phi]$$

Supervised Learning – General Formulation

When we talk about learning or training, we are saying we are fitting the model to find parameters that make good predictions. In other words, we are minimizing the model prediction agains the data. Which implies in minimizing the *loss function.*

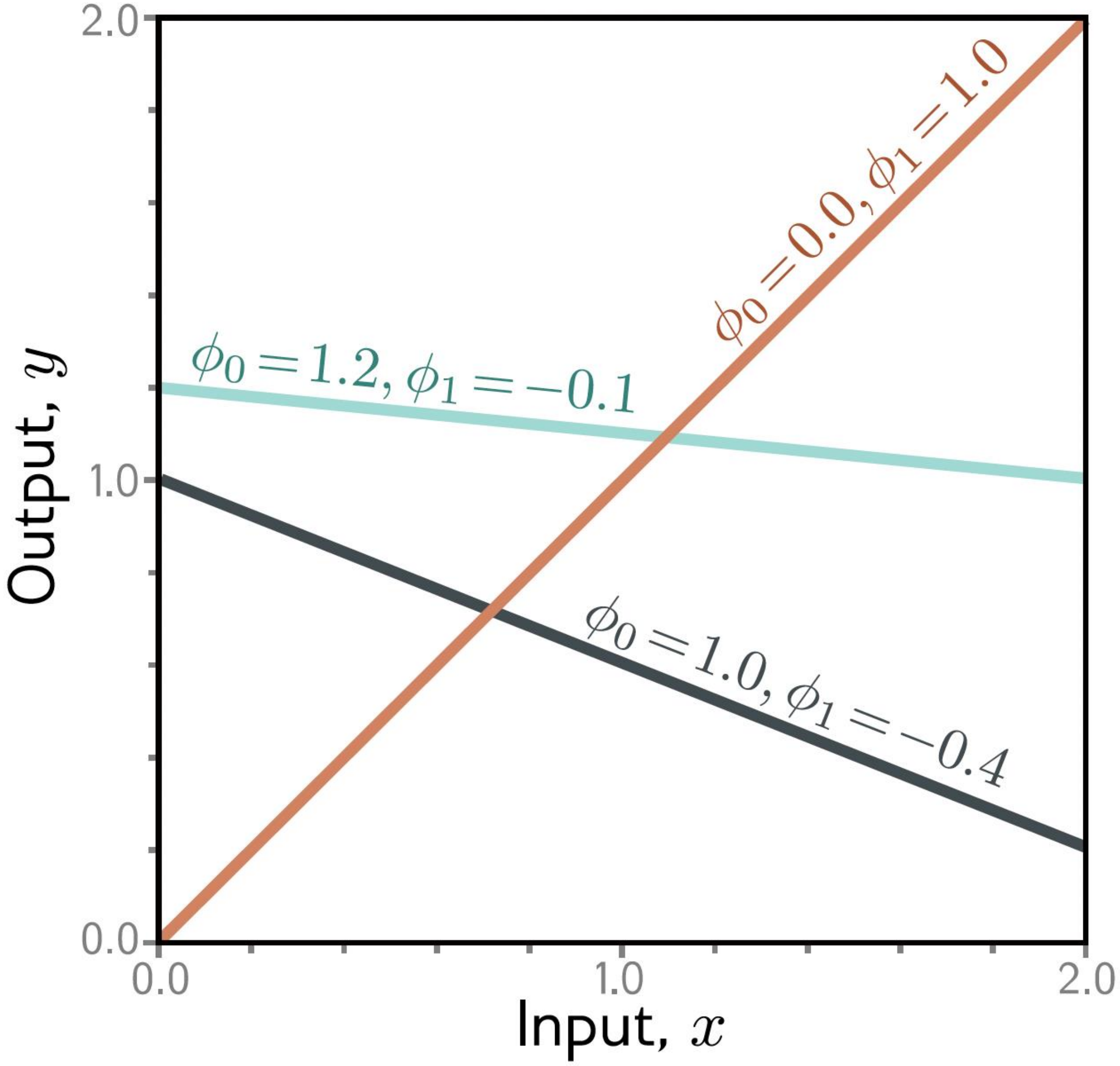
$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}} \Big[\mathrm{L}\left[\phi\right] \Big]$$

The case of Linear Regression

$$\begin{aligned} y &= \mathrm{f}[x, \boldsymbol{\phi}] \\ &= \phi_0 + \phi_1 x \end{aligned}$$

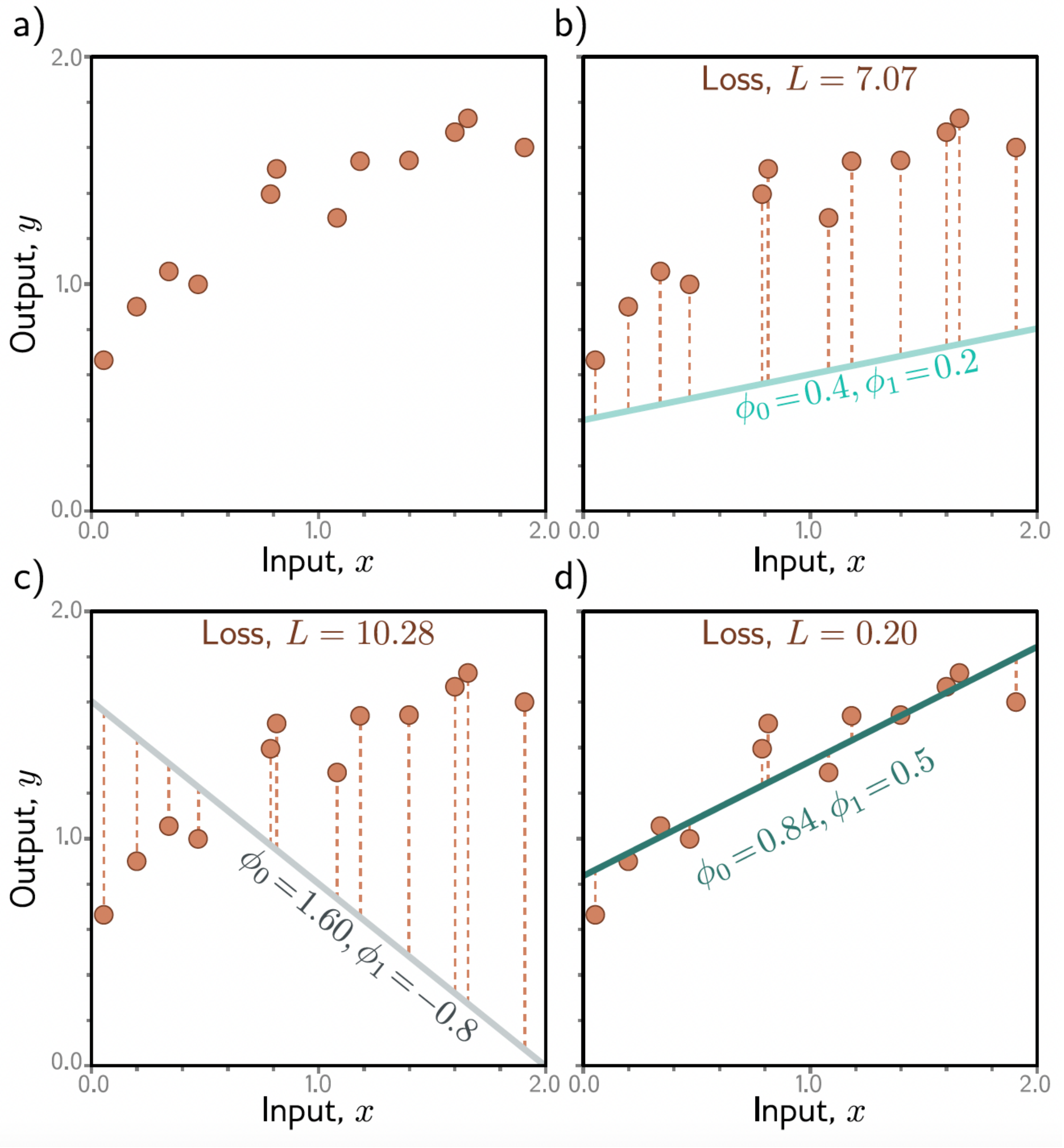
On this simple case, the function f is known, but not the parameters.

The case of Linear Regression



Linear regression model. For a given choice of parameters $\boldsymbol{\phi} = [\phi_0, \phi_1]^T$, the model makes a prediction for the output (y-axis) based on the input (x-axis). Different choices for the y-intercept ϕ_0 and the slope ϕ_1 change these predictions (cyan, orange, and gray lines). The linear regression model (equation 2.4) defines a family of input/output relations (lines) and the parameters determine the member of the family (the particular line).

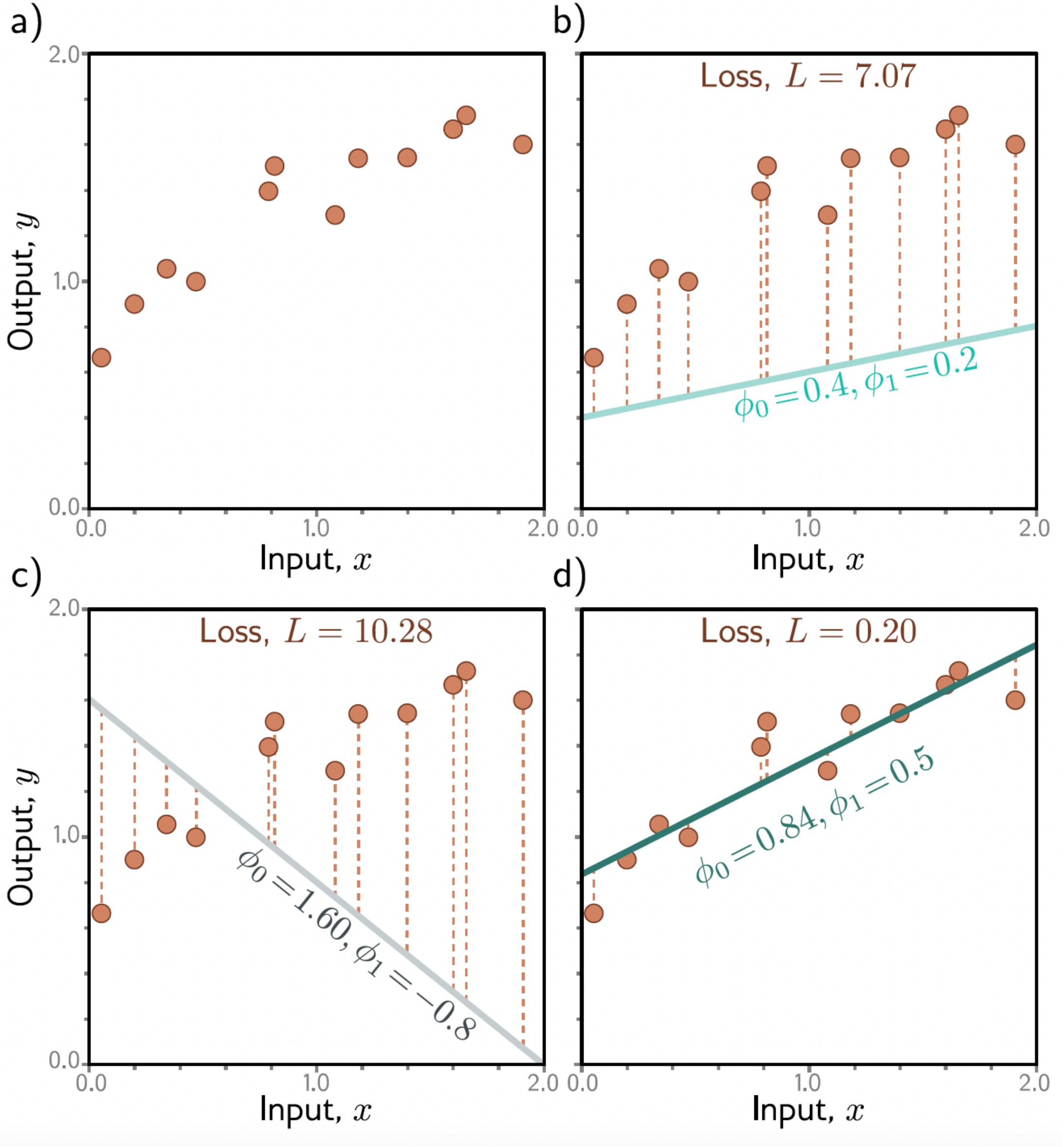
Loss Function



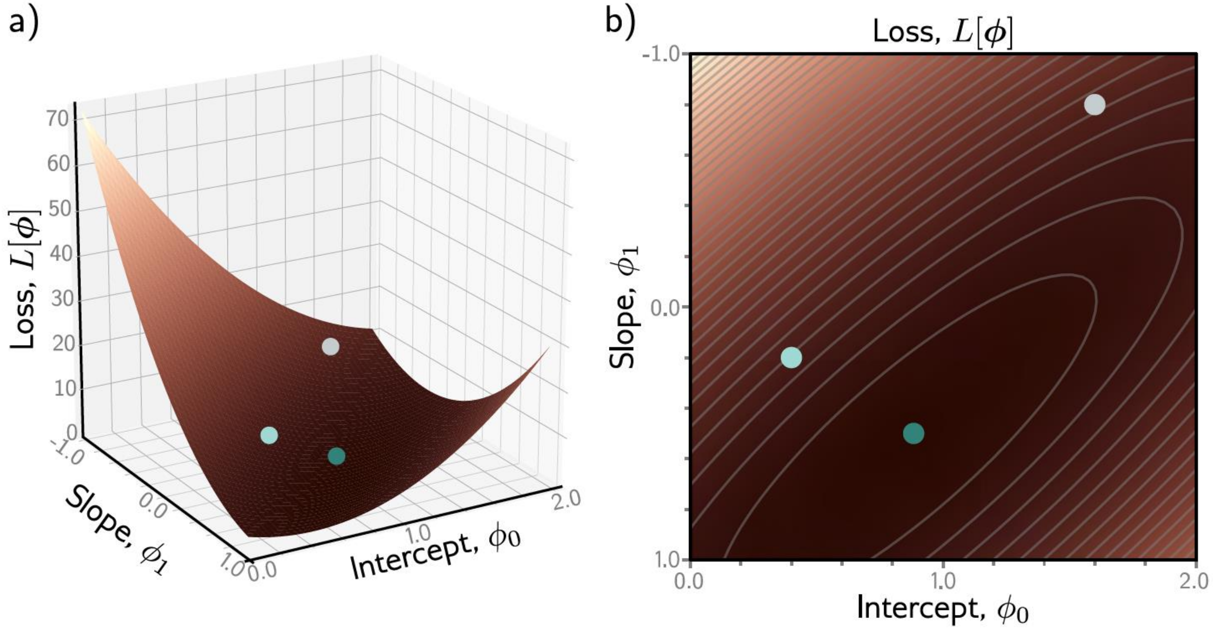
$$
\begin{aligned}
L[\phi] &= \sum_{i=1}^{I} \left(f[x_i, \phi] - y_i\right)^2 \\
&= \sum_{i=1}^{I} \left(\phi_0 + \phi_1 x_i - y_i\right)^2 .
\end{aligned}
$$

Let's appreciate this for a moment.

Now it's a good time for questions.

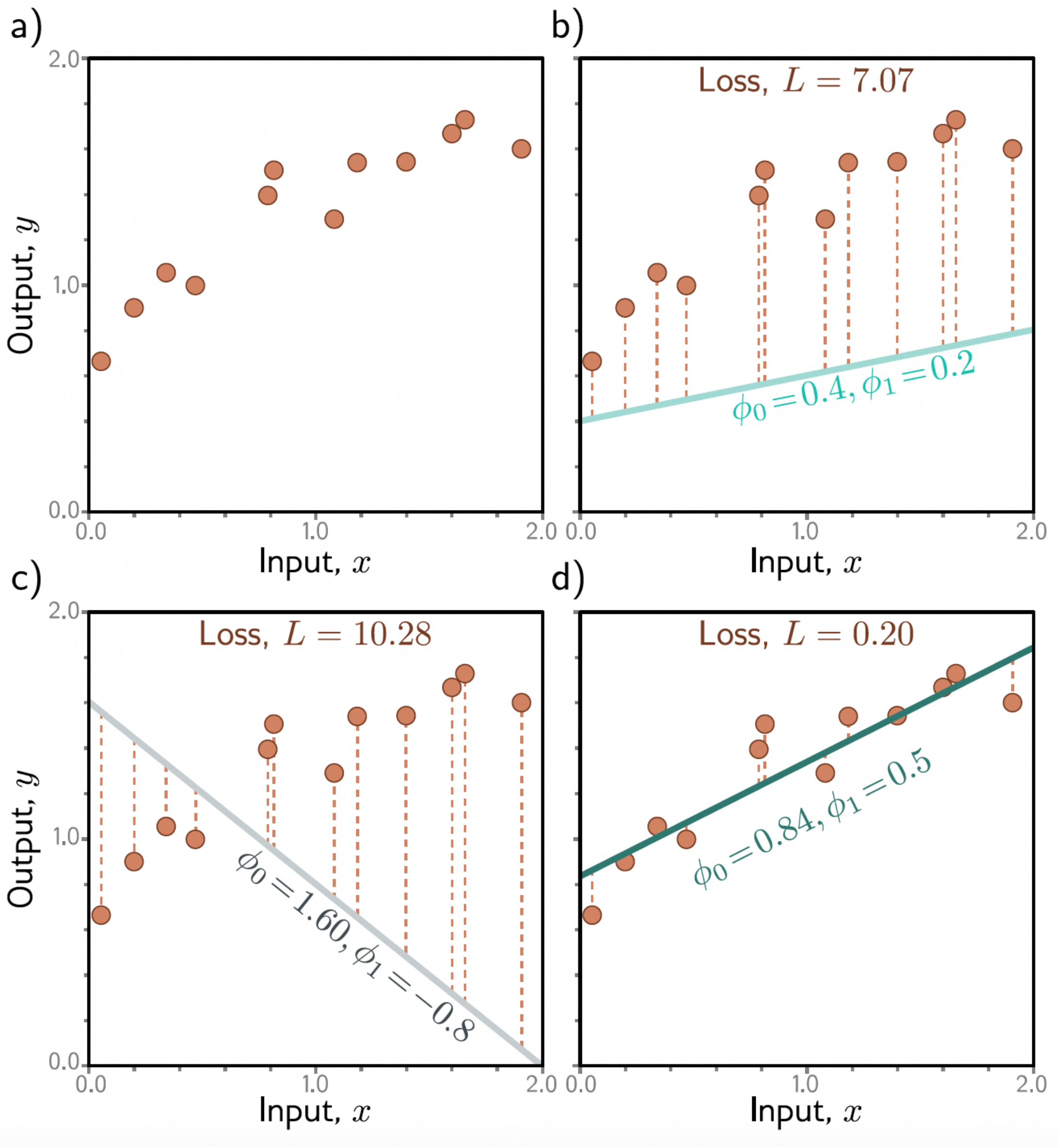


$$
\begin{aligned}
\mathrm{L}[\phi] &= \sum_{i=1}^{I} \left(\mathrm{f}[x_i, \boldsymbol{\phi}] - y_i\right)^2 \\
&= \sum_{i=1}^{I} \left(\phi_0 + \phi_1 x_i - y_i\right)^2 .
\end{aligned}
$$

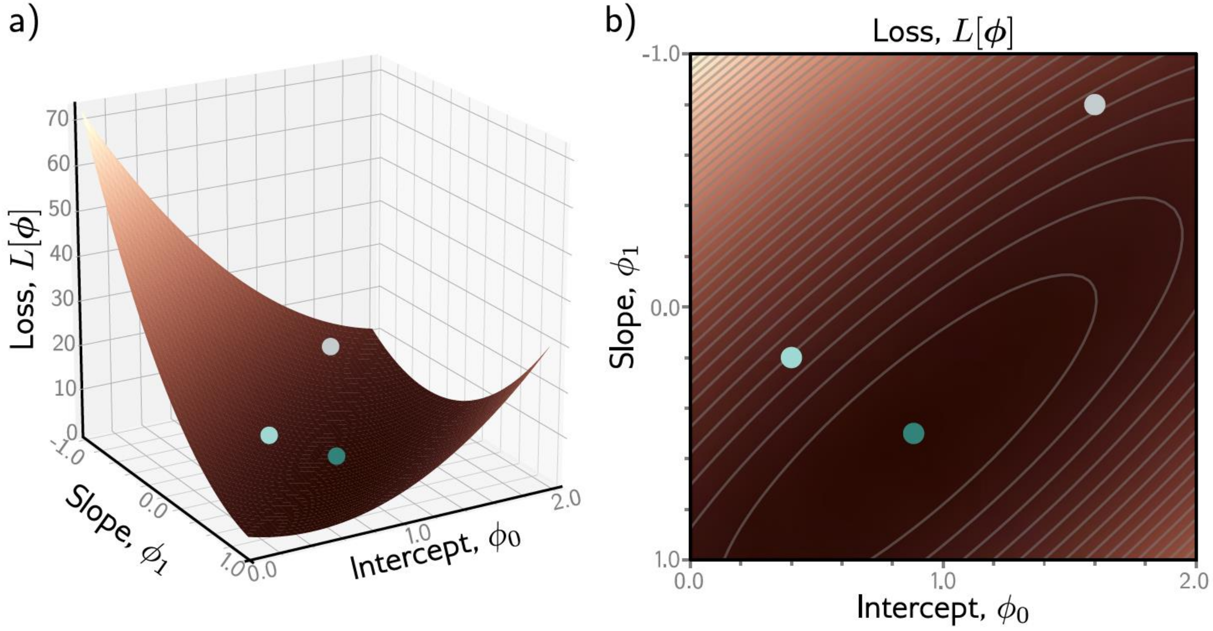


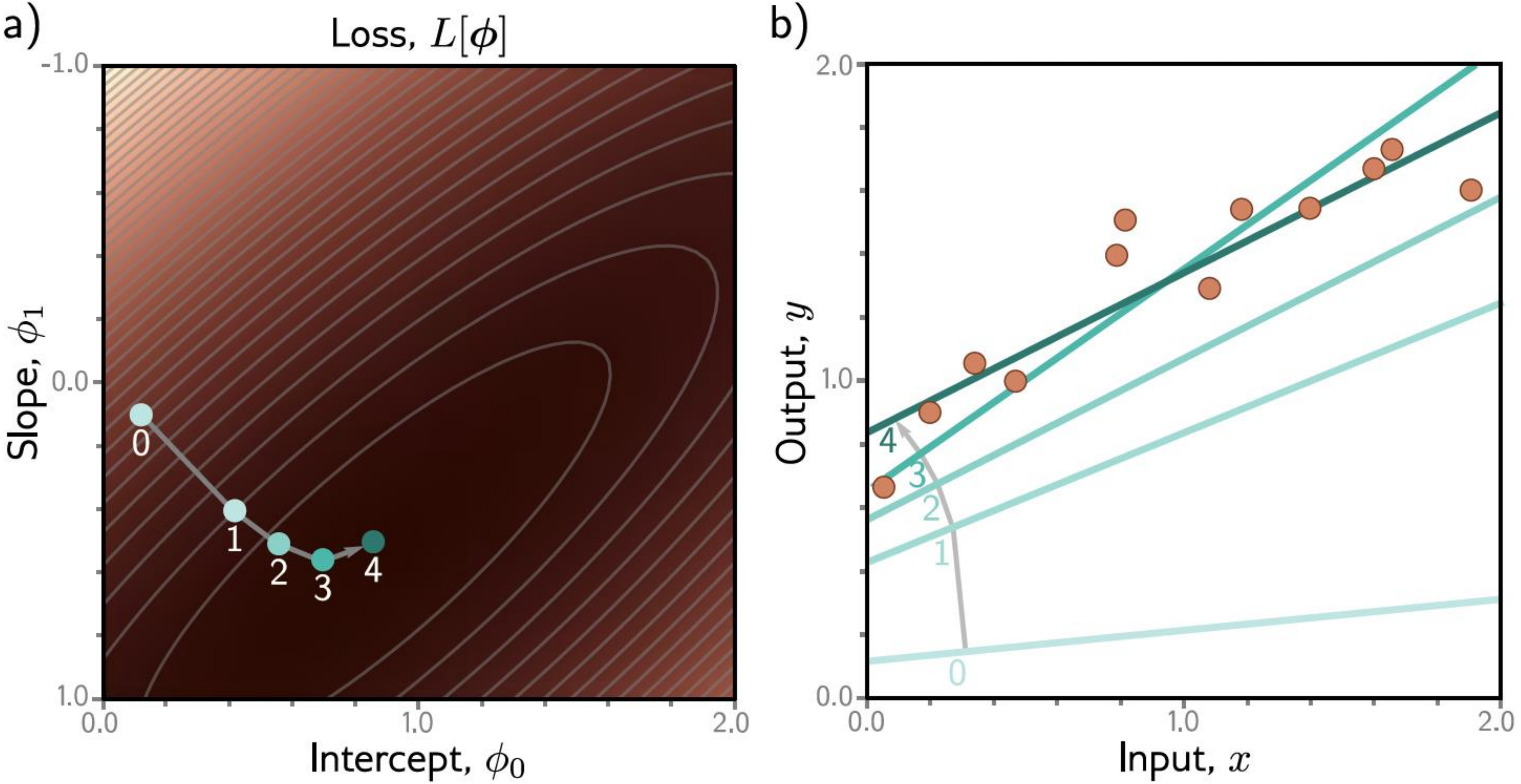
Let's appreciate this for a moment.

Now it's a good time for questions.



$$
\begin{aligned}
\hat{\boldsymbol{\phi}} &= \underset{\boldsymbol{\phi}}{\operatorname{argmin}}\Big[\mathrm{L}[\boldsymbol{\phi}]\Big] \\
&= \underset{\boldsymbol{\phi}}{\operatorname{argmin}}\left[\sum_{i=1}^{I}\left(\mathrm{f}[x_i, \boldsymbol{\phi}] - y_i\right)^2\right] \\
&= \underset{\boldsymbol{\phi}}{\operatorname{argmin}}\left[\sum_{i=1}^{I}\left(\phi_0 + \phi_1 x_i - y_i\right)^2\right].
\end{aligned}
$$





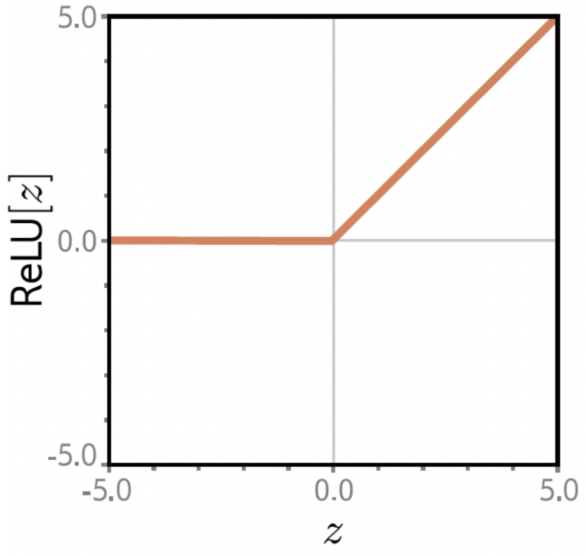
a)
Loss, $L[\phi]$
-1.0
0.0
1.0
Slope, ϕ_1
0.0
1.0
2.0
Intercept, ϕ_0
0
1
2
3
4
b)
2.0
1.0
0.0
Output, y
0.0
1.0
2.0
Input, x
4
3
2
1
0

The case of neural networks

$$\begin{aligned} y &= \mathrm{f}[x, \boldsymbol{\phi}] \\ &= \phi_0 + \phi_1 \mathrm{a}[\theta_{10} + \theta_{11} x] + \phi_2 \mathrm{a}[\theta_{20} + \theta_{21} x] + \phi_3 \mathrm{a}[\theta_{30} + \theta_{31} x]. \end{aligned}$$

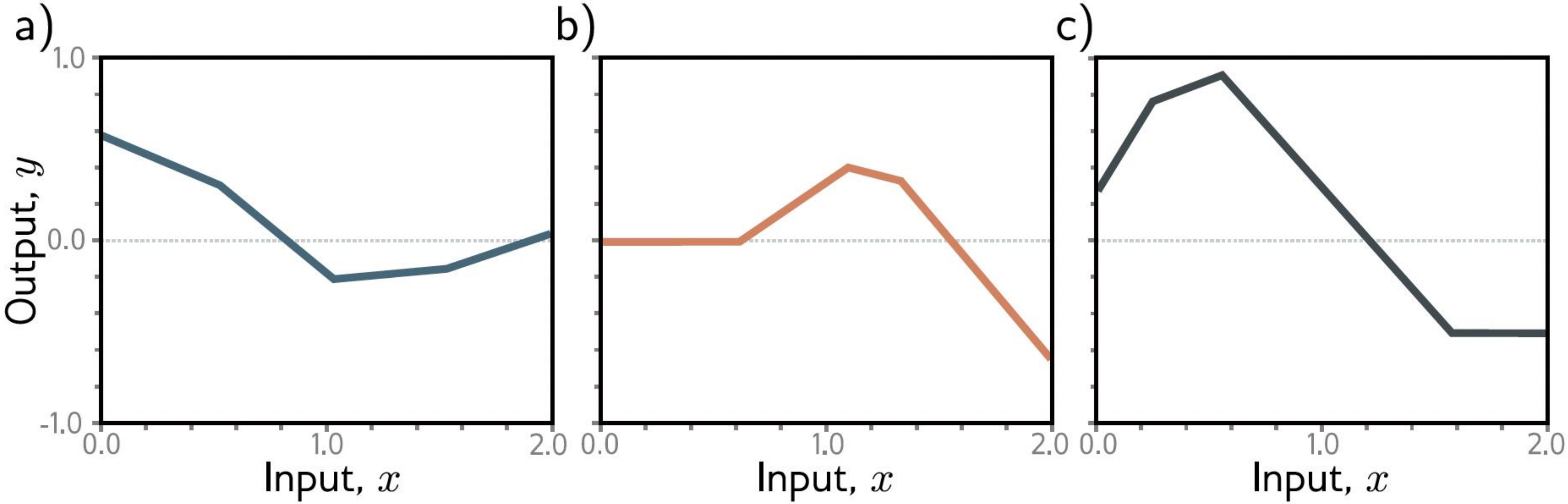
$$\boldsymbol{\phi} = \{\phi_0, \phi_1, \phi_2, \phi_3, \theta_{10}, \theta_{11}, \theta_{20}, \theta_{21}, \theta_{30}, \theta_{31}\}$$

$$\mathrm{a}[z] = \mathrm{ReLU}[z] = \begin{cases} 0 & z < 0 \\ z & z \geq 0 \end{cases}$$



The case of neural networks

$$
\begin{aligned}
y &= \mathrm{f}[x, \boldsymbol{\phi}] \\
&= \phi_0 + \phi_1 \mathrm{a}[\theta_{10} + \theta_{11} x] + \phi_2 \mathrm{a}[\theta_{20} + \theta_{21} x] + \phi_3 \mathrm{a}[\theta_{30} + \theta_{31} x].
\end{aligned}
$$



Different values of the parameters

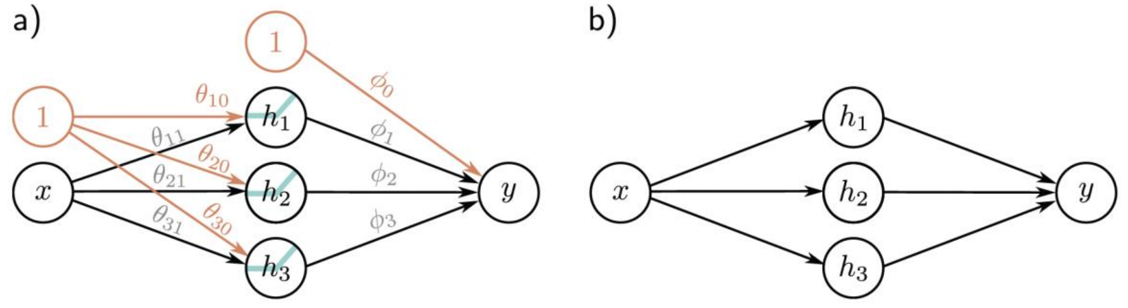
Neural Network Intuition

$$\begin{aligned} y &= \mathrm{f}[x, \phi] \\ &= \phi_0 + \phi_1 \boxed{\mathrm{a}[\theta_{10} + \theta_{11}x]} + \phi_2 \boxed{\mathrm{a}[\theta_{20} + \theta_{21}x]} + \phi_3 \boxed{\mathrm{a}[\theta_{30} + \theta_{31}x]}. \end{aligned}$$

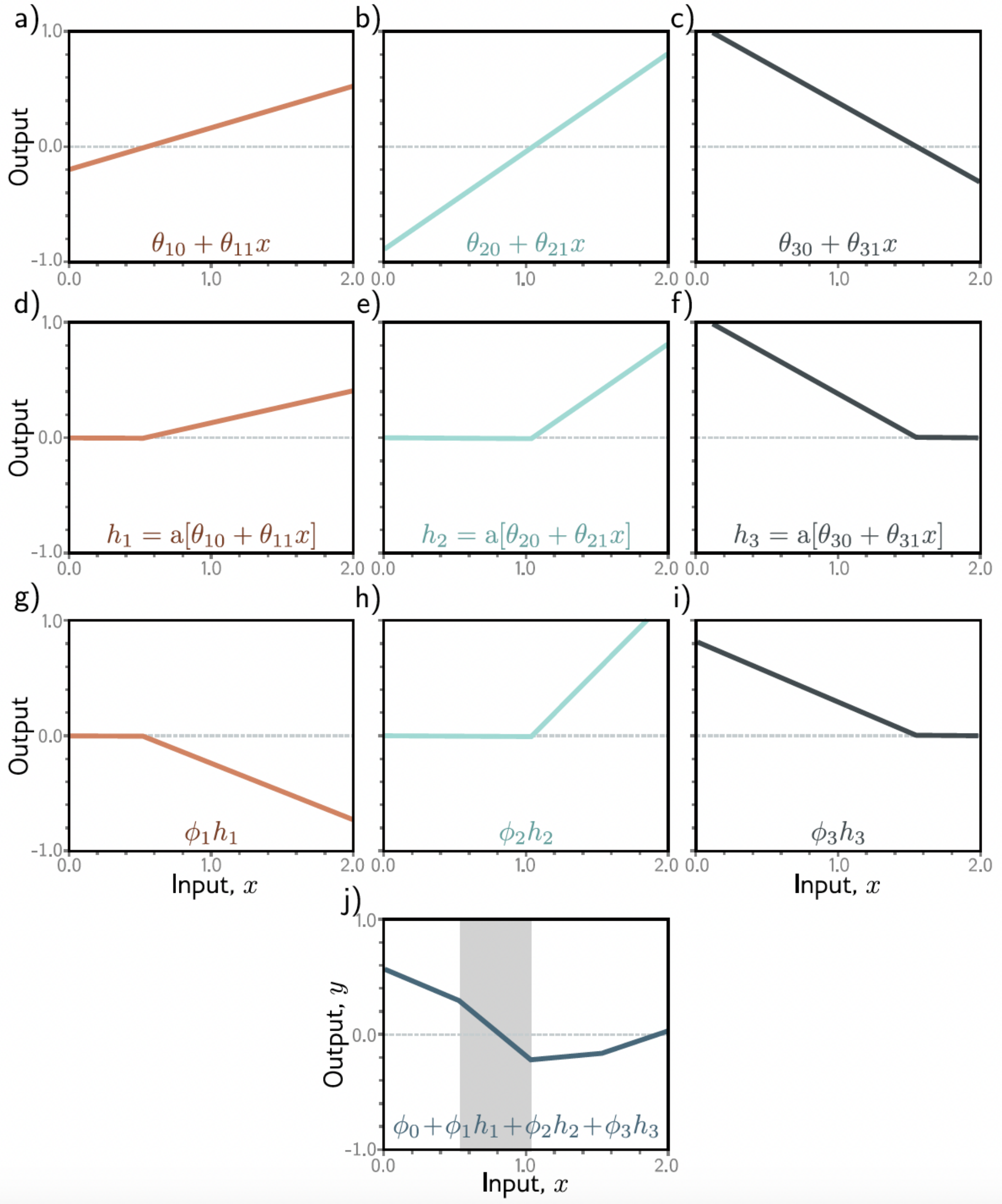
$$\begin{aligned} h_1 &= \boxed{\mathrm{a}[\theta_{10} + \theta_{11}x]} \\ h_2 &= \boxed{\mathrm{a}[\theta_{20} + \theta_{21}x]} \\ h_3 &= \boxed{\mathrm{a}[\theta_{30} + \theta_{31}x]}, \end{aligned}$$

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

Hidden Units

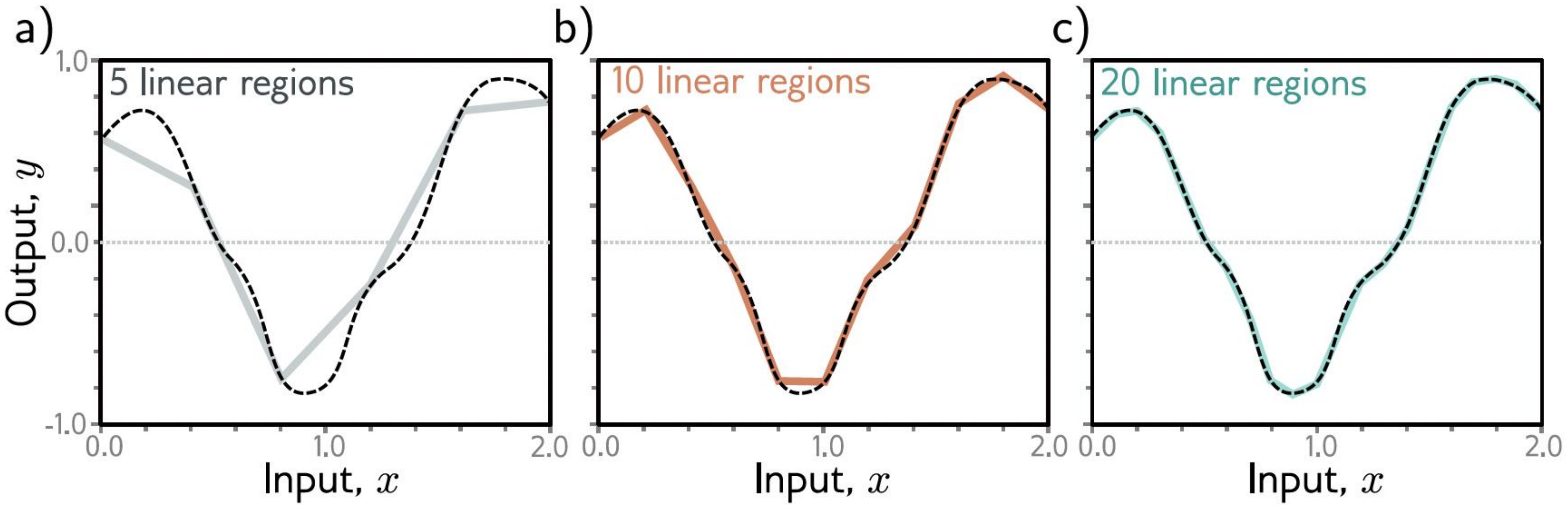


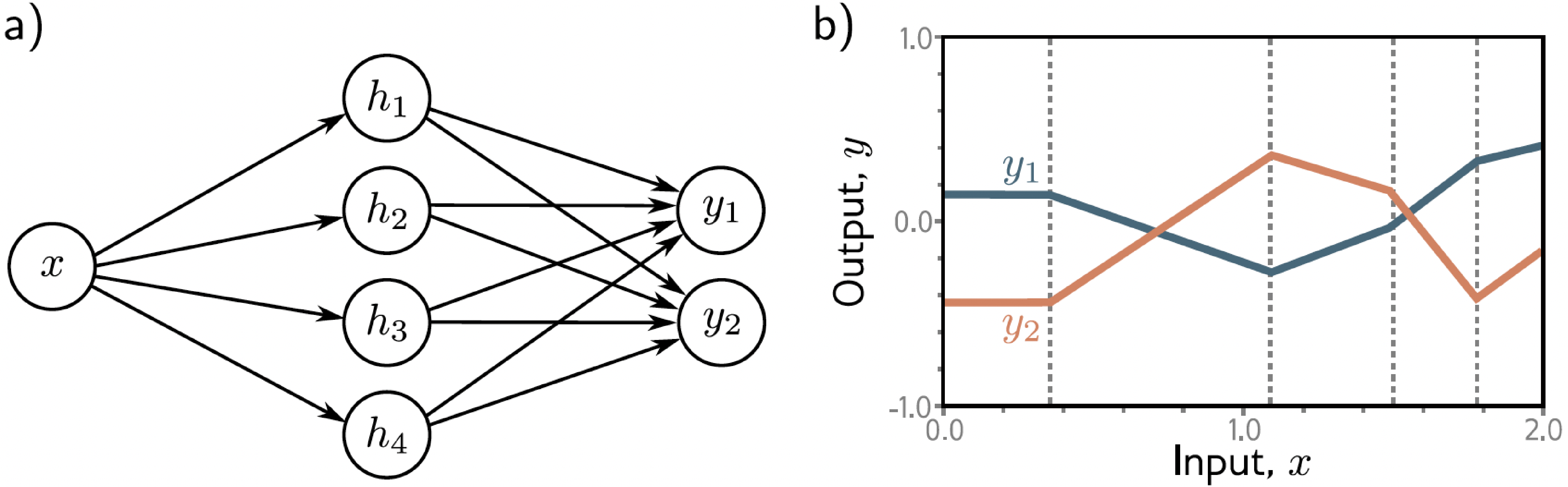
$$y = \phi_0 + \phi_1 h_1 + \phi_2 h_2 + \phi_3 h_3.$$



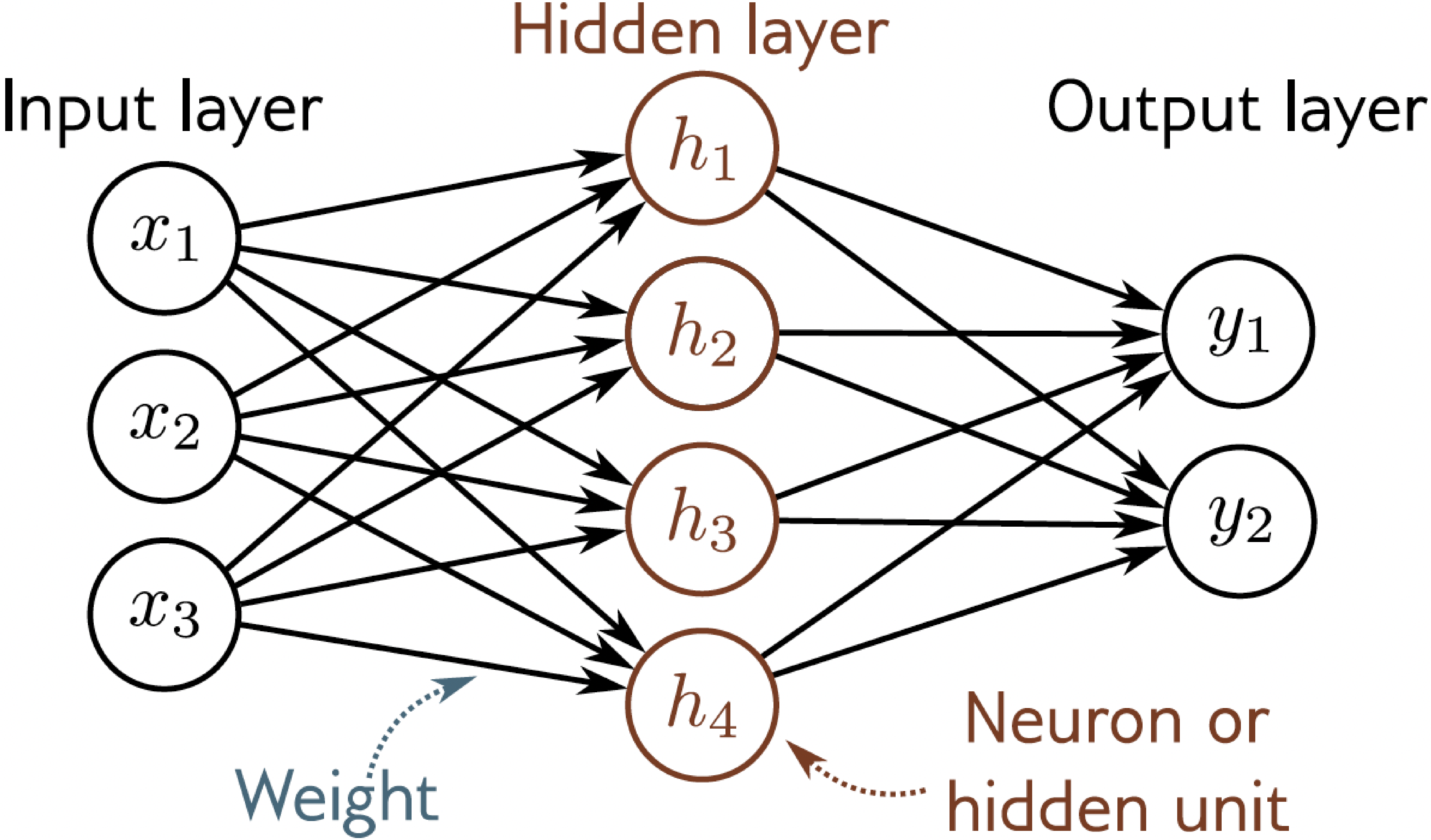
$$h_d = \mathrm{a}[\theta_{d0} + \theta_{d1} x], \qquad y = \phi_0 + \sum_{d=1}^{D} \phi_d h_d.$$

The number of hidden layers is a measure of the network flexibility

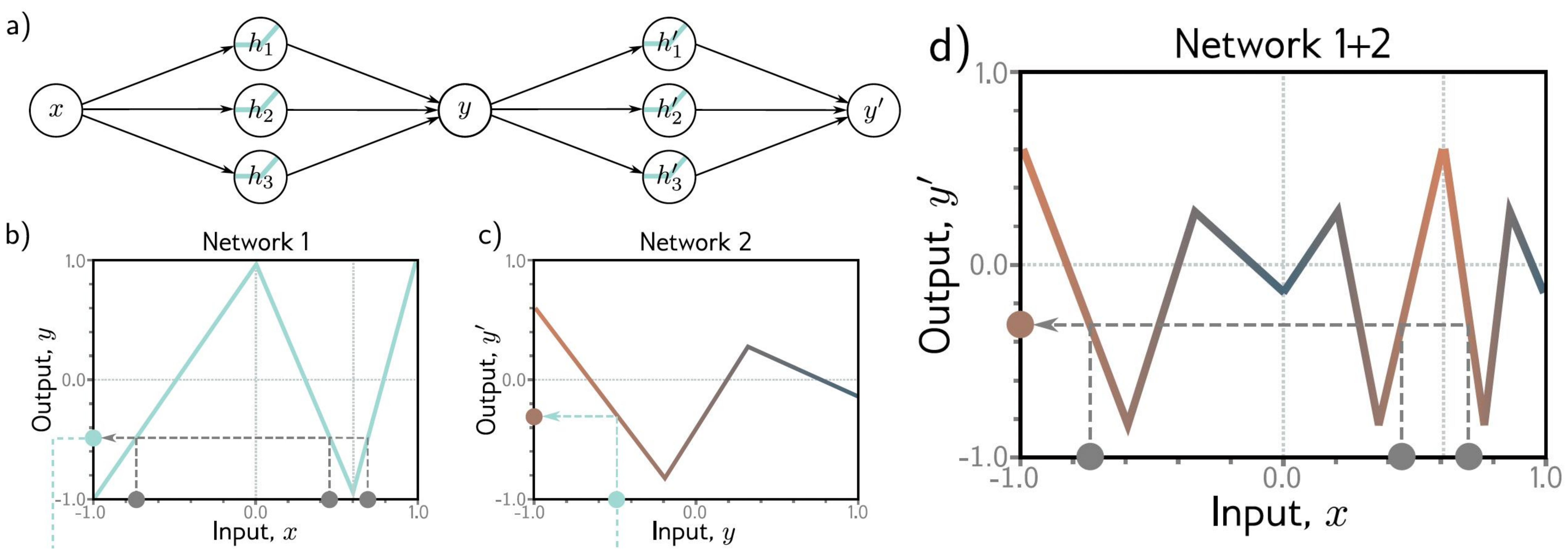


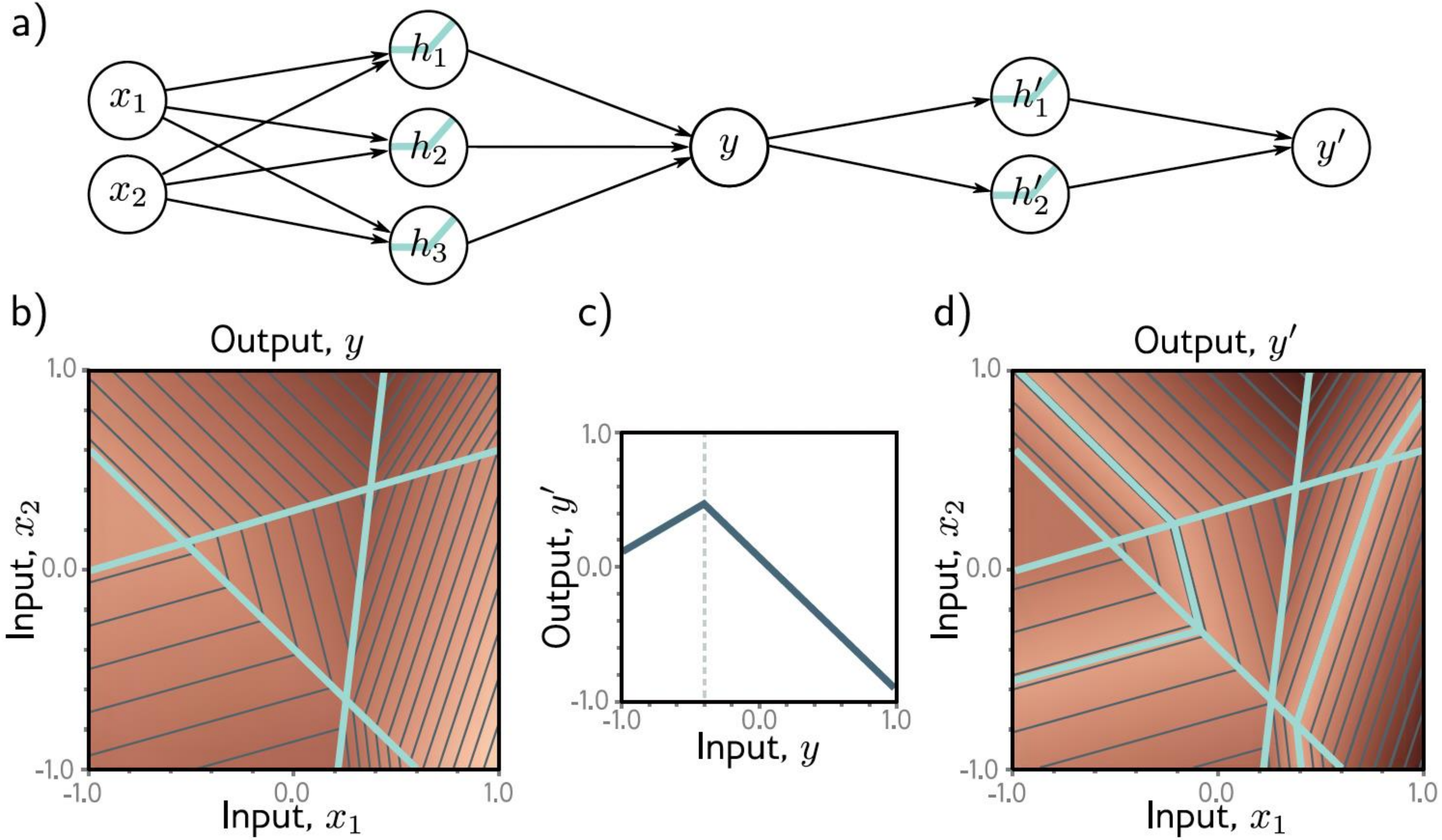


a)
x
h_1
h_2
h_3
h_4
y_1
y_2
b)
1.0
0.0
-1.0
0.0
1.0
2.0
Output, y
Input, x
y_1
y_2

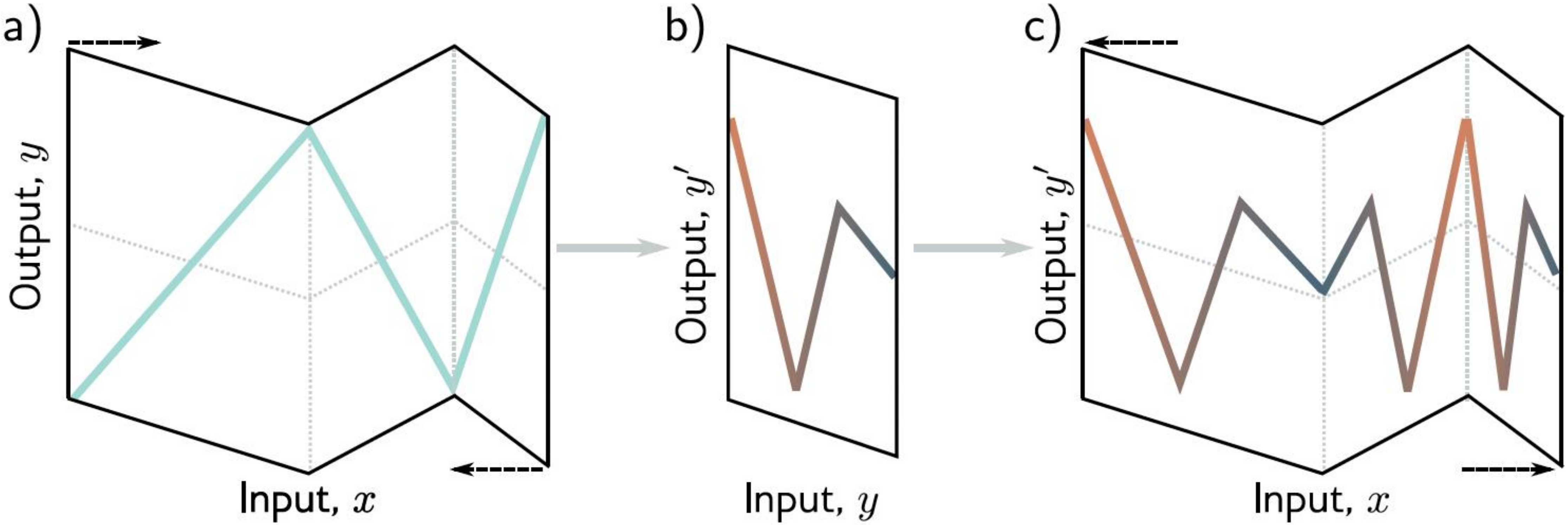

Hidden layer
Input layer
Output layer
x_1
x_2
x_3
h_1
h_2
h_3
h_4
y_1
y_2
Weight
Neuron or hidden unit

Deep neural networks





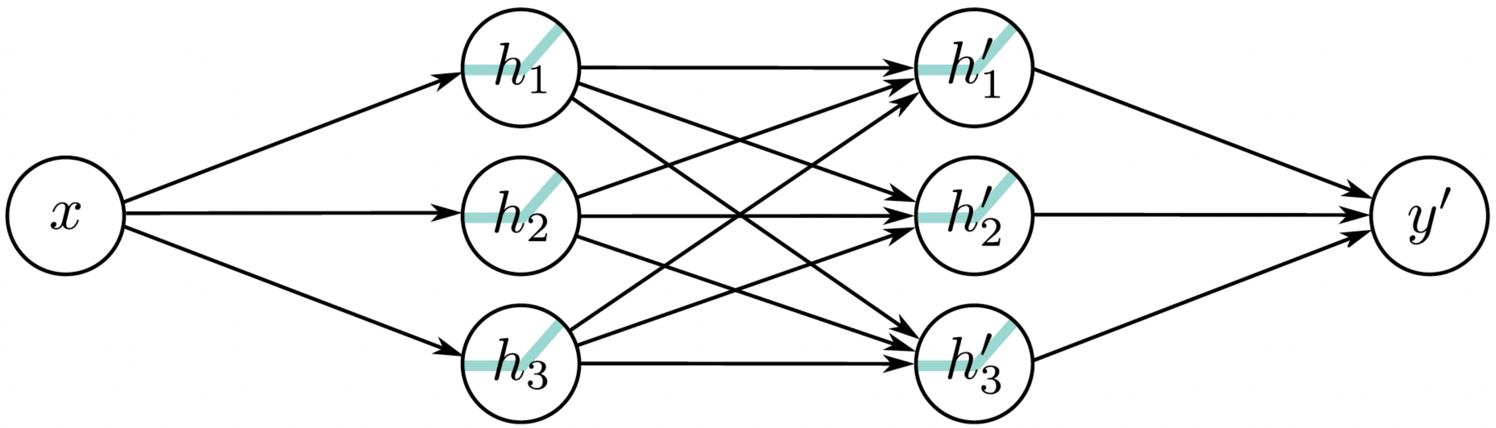
a)
x_1
x_2
h_1
h_2
h_3
y
h'_1
h'_2
y'
b)
Output, y
Input, x_2
Input, x_1
1.0
0.0
-1.0
c)
Output, y'
Input, y
d)
Output, y'



a)
Output, y
Input, x
b)
Output, y'
Input, y
c)
Output, y'
Input, x

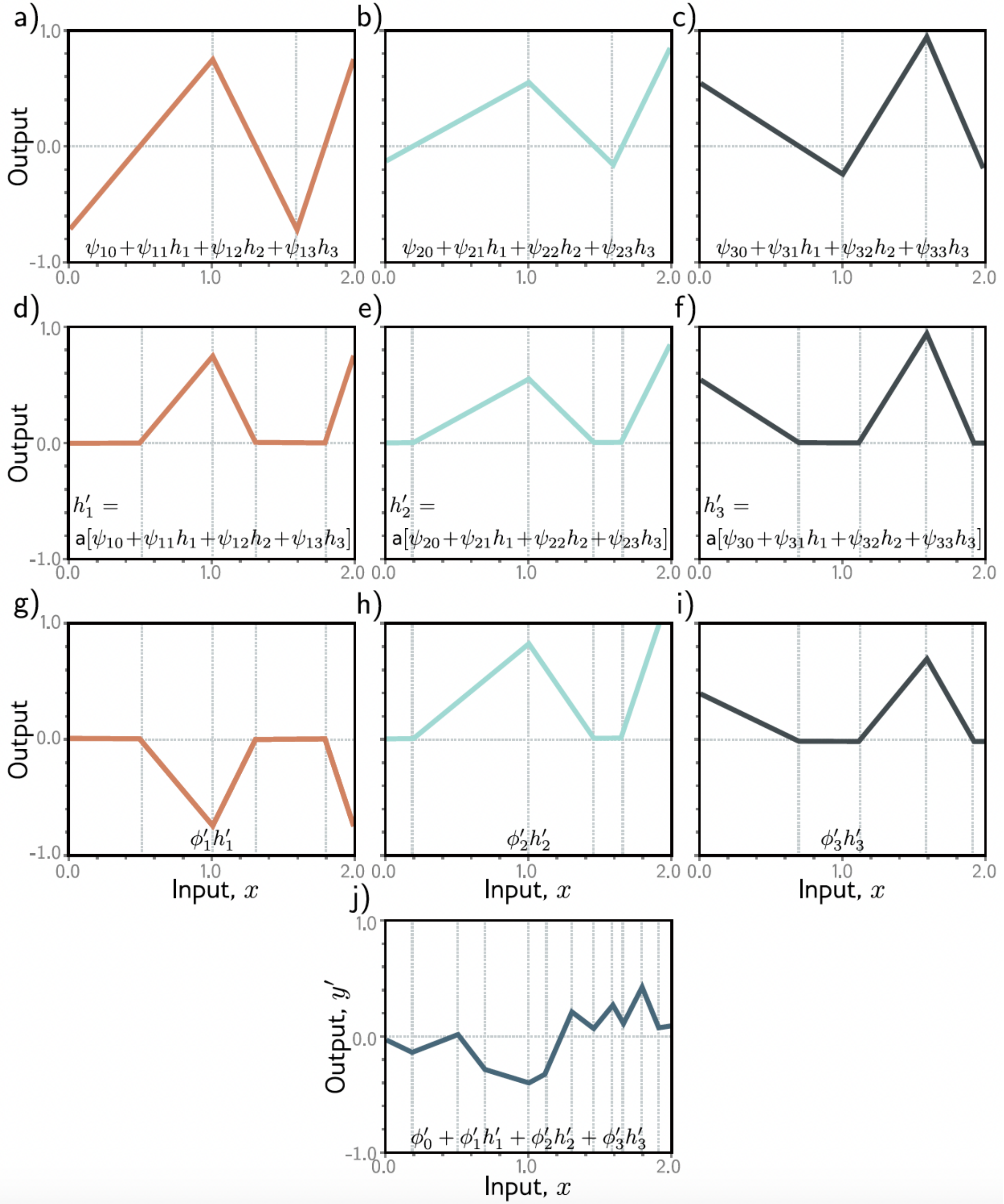
Deep neural Networks

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



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

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



$$
\begin{aligned}
y' \quad = \quad & \phi'_0 + \phi'_1 \mathrm{a}\left[\psi_{10} + \psi_{11}\mathrm{a}[\theta_{10} + \theta_{11}x] + \psi_{12}\mathrm{a}[\theta_{20} + \theta_{21}x] + \psi_{13}\mathrm{a}[\theta_{30} + \theta_{31}x]\right] \\
& + \phi'_2 \mathrm{a}[\psi_{20} + \psi_{21}\mathrm{a}[\theta_{10} + \theta_{11}x] + \psi_{22}\mathrm{a}[\theta_{20} + \theta_{21}x] + \psi_{23}\mathrm{a}[\theta_{30} + \theta_{31}x]] \\
& + \phi'_3 \mathrm{a}[\psi_{30} + \psi_{31}\mathrm{a}[\theta_{10} + \theta_{11}x] + \psi_{32}\mathrm{a}[\theta_{20} + \theta_{21}x] + \psi_{33}\mathrm{a}[\theta_{30} + \theta_{31}x]],
\end{aligned}
$$