Multi layer Perceptron

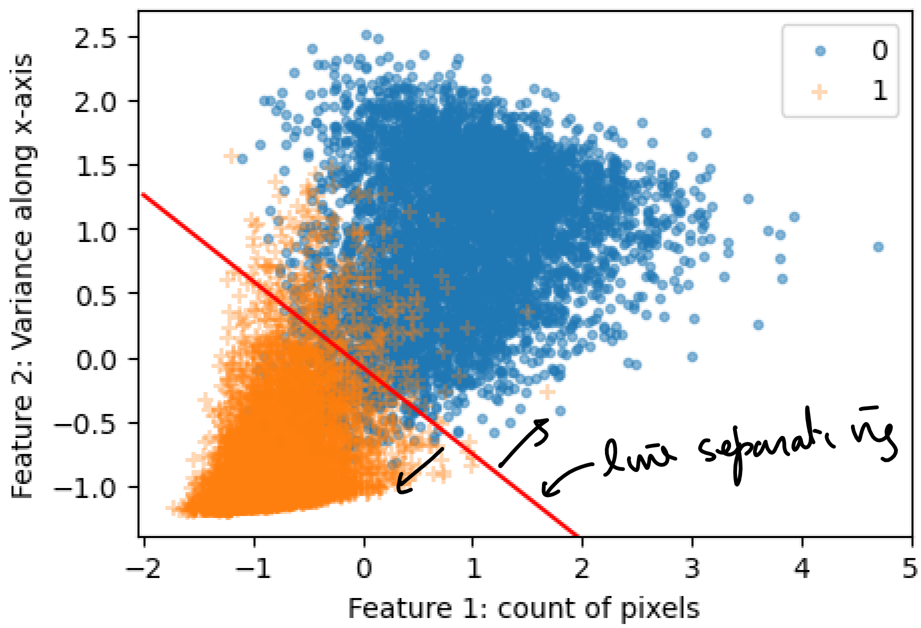
All figures are from Chapter 3 of UDLBook. https://github.com/udlbook/udlbook

Recall the single layer perceptron

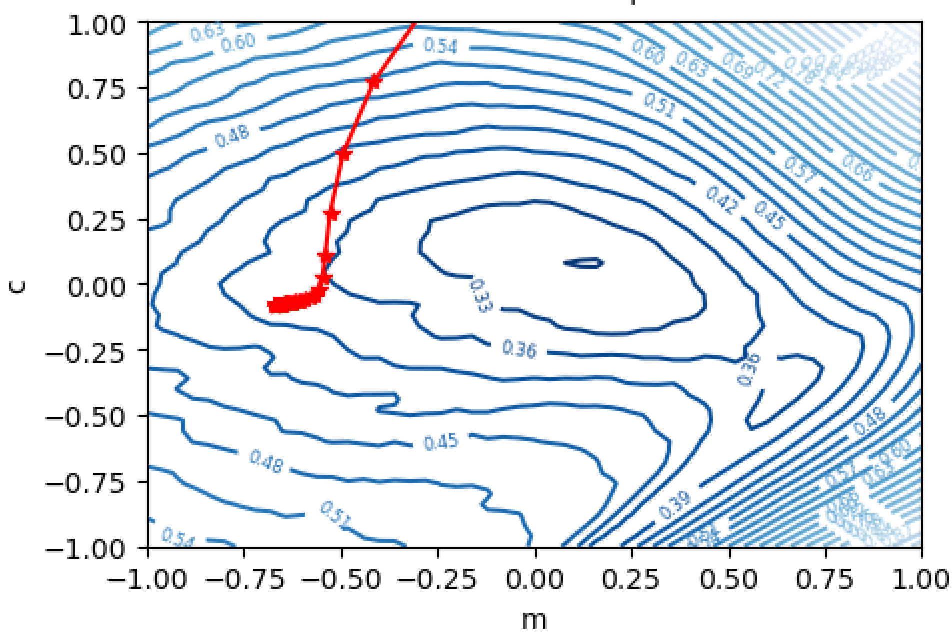
$\hat{y}_i = \underline{w}^T x_i + w_0$

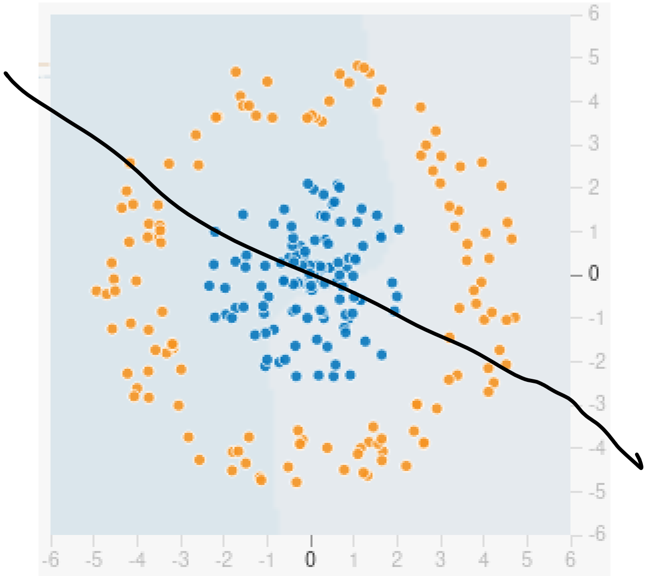
$\hat{y}_i > 0$

$y_i < 0$

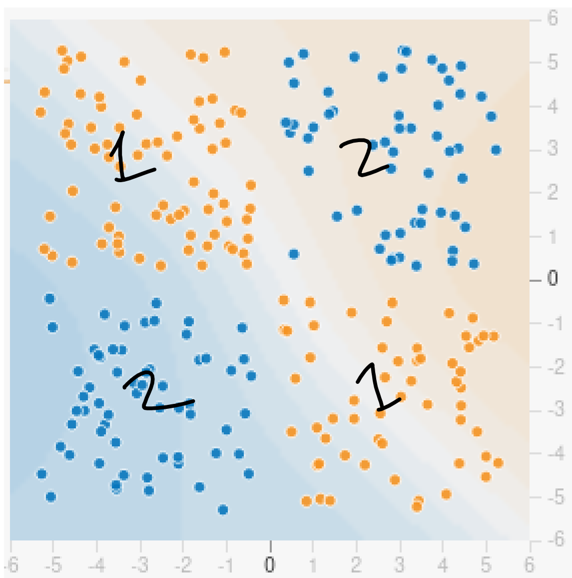


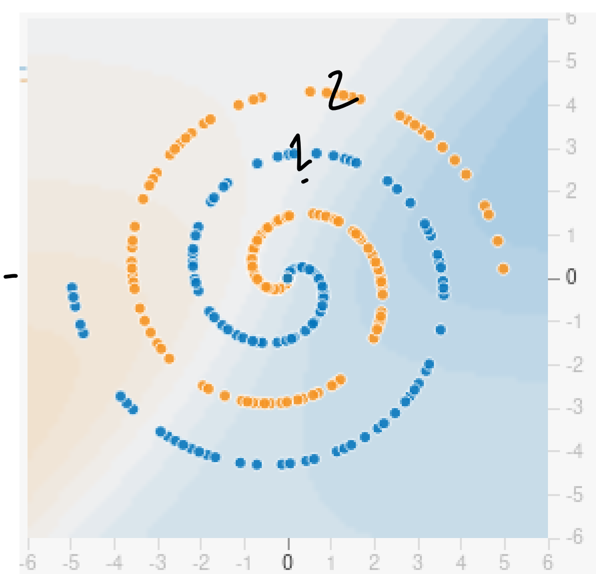
line separating the classes





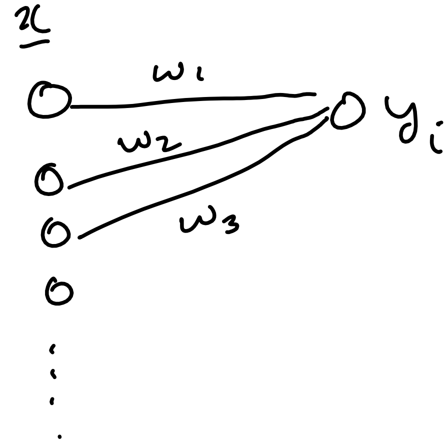
x Perceptron





Binary class classification problems that can be separated by a line (a HYPERPLANE in general) are called Linearly separable problems

$$\hat{y}_i = \underline{w}^T \underline{x} + w_0$$

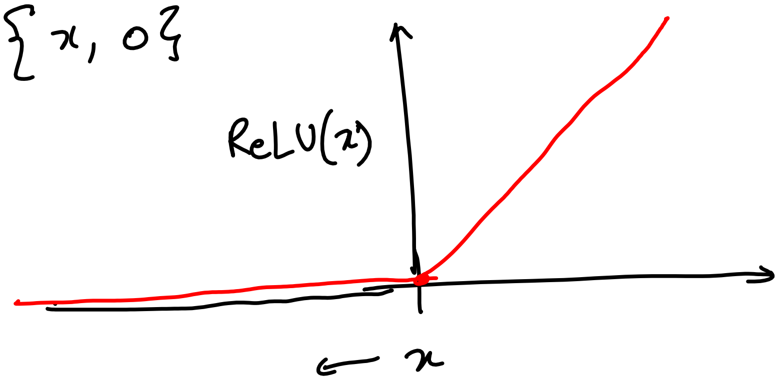


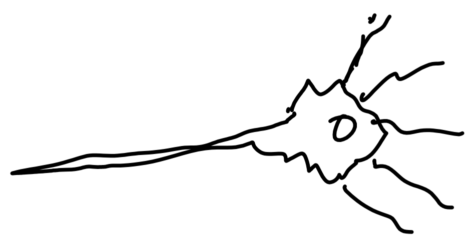
Multi Layer Perceptron model

next neuron is getting "activated" → Activation function / non-Linearities

$$a(\cdot) = \text{ReLU}, \text{ sigmoid}, \text{ Tanh}, \ldots\ldots$$

$$\text{ReLU}(x) = \max\{x, 0\}$$





$$
\begin{aligned}
y_1 &= \underline{w}_1^T \underline{x} + w_{01} \\
y_2 &= \underline{w}_2^T \underline{x} + w_{02} \\
y_3 & \\
&\vdots \\
y_m &= \underline{w}_m^T \underline{x} + w_{0m}
\end{aligned}
$$

$$\underline{x} \in \mathbb{R}^n$$

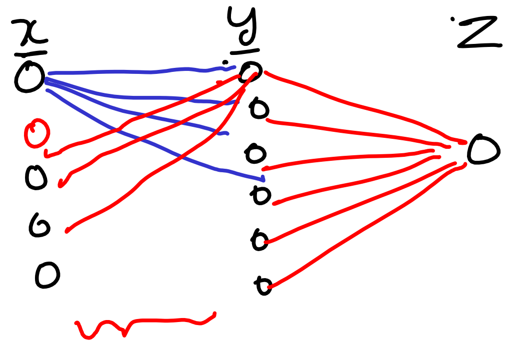
$$\underline{y}_m = \underbrace{W_{m \times n}}_{\text{weight}} \underline{x} + \underbrace{\underline{w}_0}_{\text{bias}}$$

$$z = {}_2\underline{w}^T \underline{y}_{m\times 1} + {}_2w_0$$

$$z = \underbrace{{}_2\underline{w}^T (W_{m\times n}}_{\downarrow}\, \underline{x} + \underline{w}_0) + {}_2w_0$$

$$= \underline{u}^T \underline{x} + \underbrace{{}_2\underline{w}^T w_0 + {}_2w_0}_{v}$$

2-Layer MLP



Dense

$$z = {}_2\underline{w}^T a(\underline{y}) + w_0$$

activation function

non Linearities

$z > 0$ = class 1

$z < 0$ = class 2

MLP : $\underline{x} \longrightarrow \underbrace{\text{Linear} \rightarrow \text{act.}}_{\text{1st Layer}} \rightarrow \overbrace{\text{Linear} \rightarrow \text{act}}^{2^{nd}\text{ Layer}}$

$\downarrow$

3rd $\begin{cases} \text{Linear} \\ \downarrow \\ \text{act} \end{cases}$

$\downarrow$

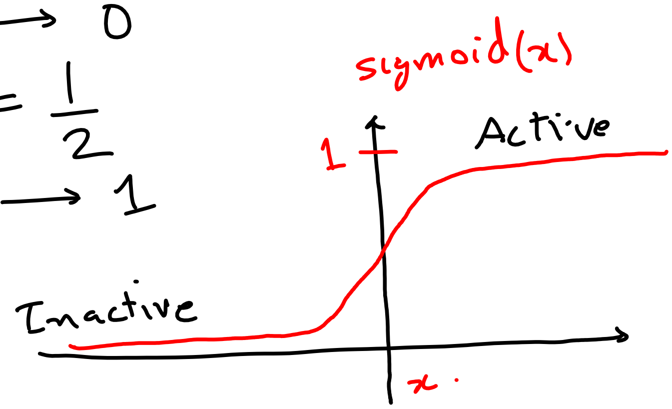
output

$$\text{Sigmoid}(x) = \frac{1}{1+e^{-x}} = \frac{1}{1+\exp(-x)}$$

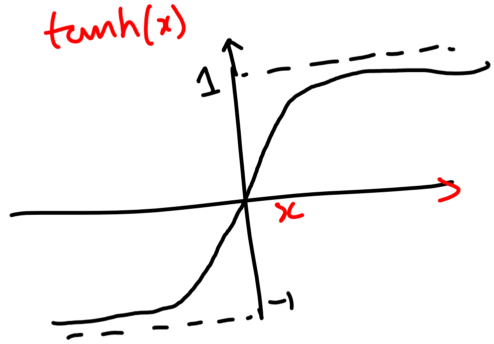
$x \to -\infty$ $\quad$ $\text{sigmoid}(x) \longrightarrow 0$

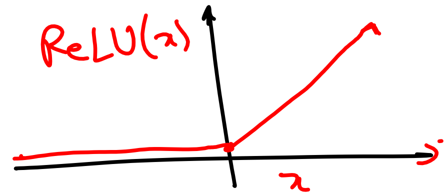
$x = 0$ $\quad$ $\text{sigmoid}(0) = \frac{1}{2}$

$x \to +\infty$ $\quad$ $\text{sigmoid}(x) \longrightarrow 1$



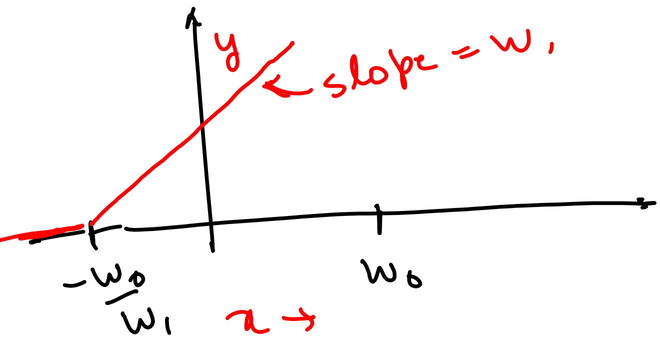
$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$$





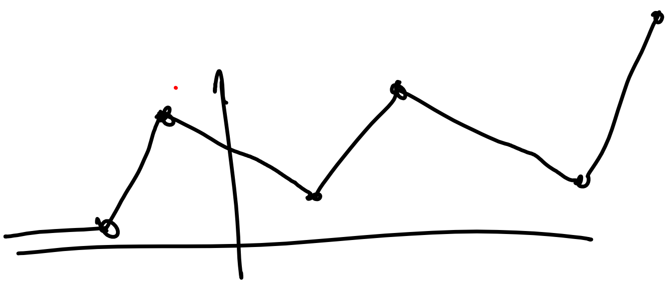
Piecewise Linear

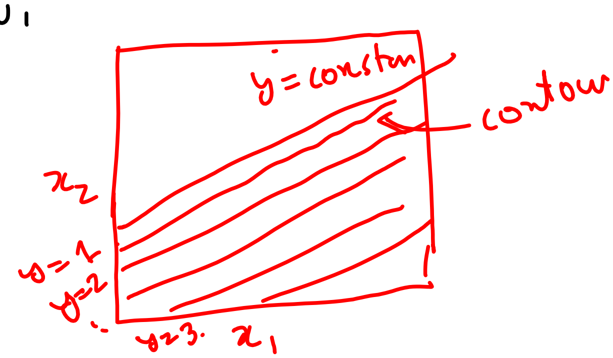
$$y = \mathrm{ReLU}(w_1 x + w_0)$$

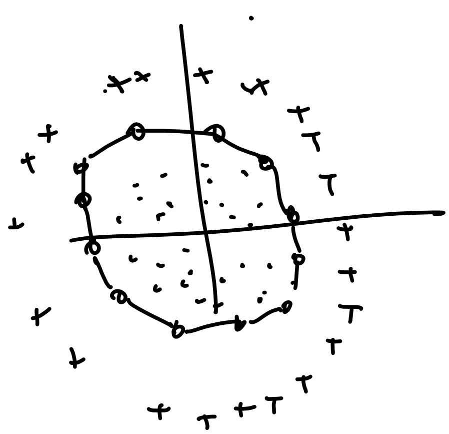


$$w_1 x + w_0 > 0$$

$$x > \frac{-w_0}{w_1}$$





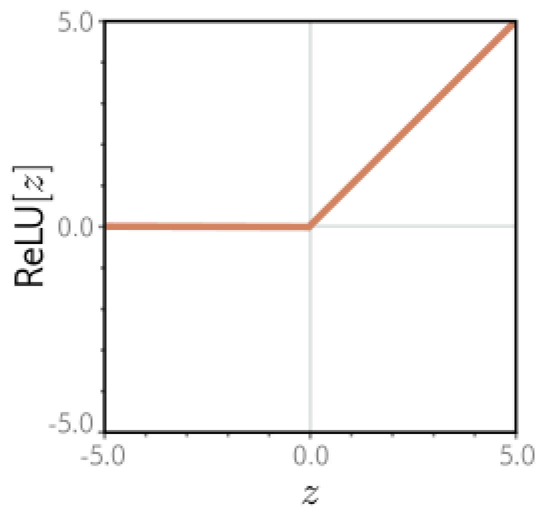


UDL Book

$$l = f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + w_0$$

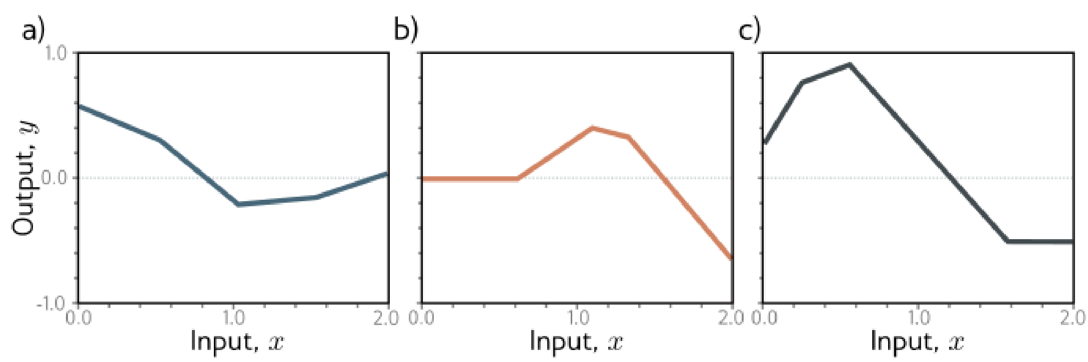
Multi Layer Perceptrons

ReLU activation function $\mathrm{ReLU}(z) = \max\{0, z\}$



Two layer Perceptron

$$y = f(\mathbf{x}) = \mathrm{Linear}(\ \mathrm{ActivationFunction}(\ \mathrm{Linear}(x)))$$



Example

$$a(x) = \mathrm{ReLU}(x)$$

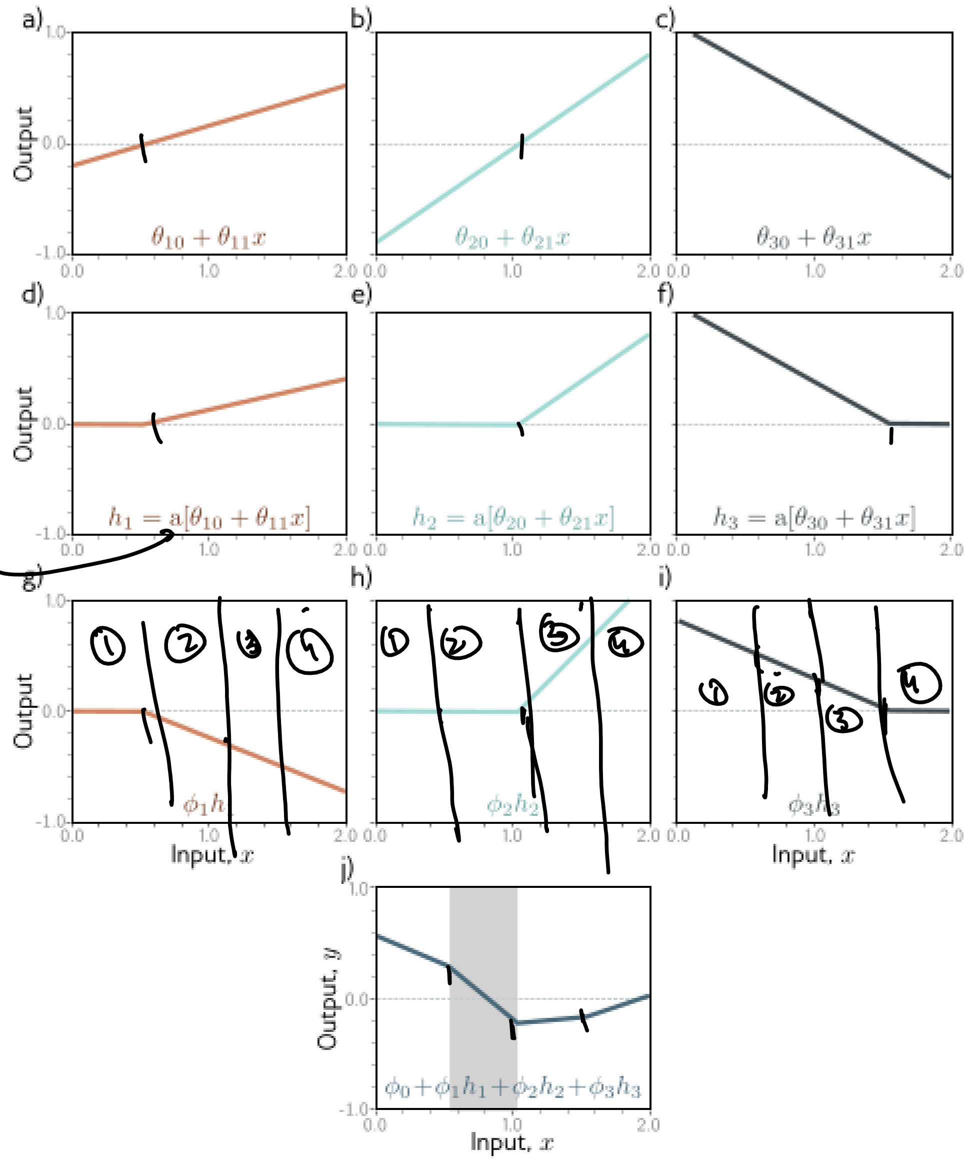
$$h_1 = a(\theta_{10} + \theta_{11} x)$$

$$h_2 = a(\theta_{20} + \theta_{21} x)$$

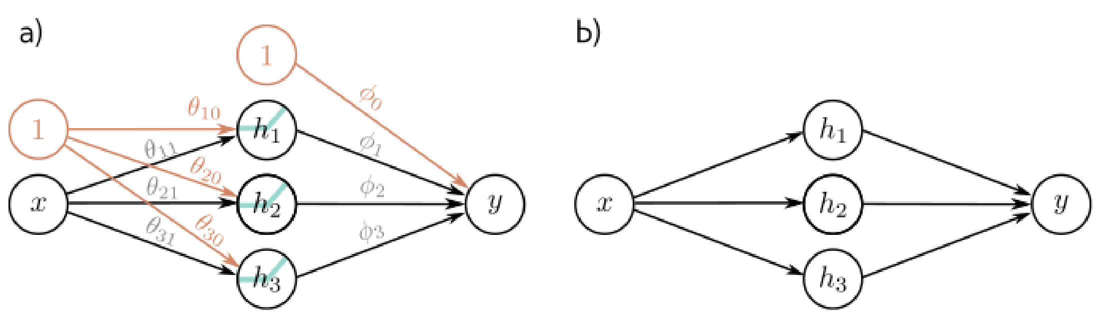
$$h_3 = a(\theta_{30} + \theta_{31}x)$$

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

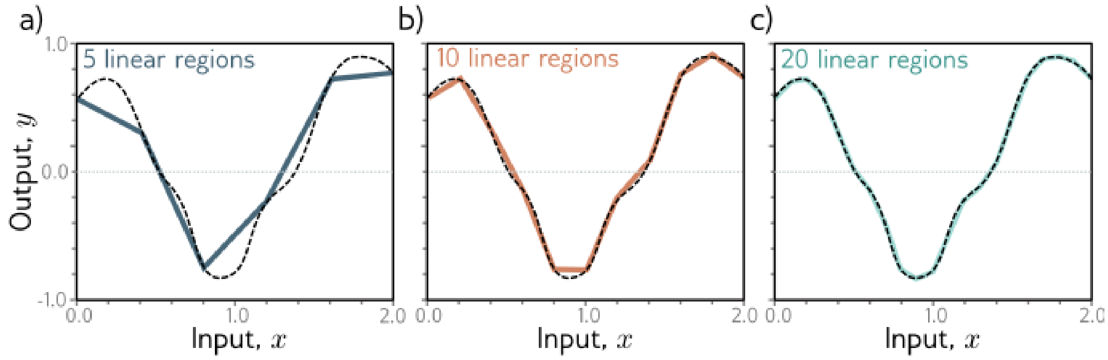
ReLU = a



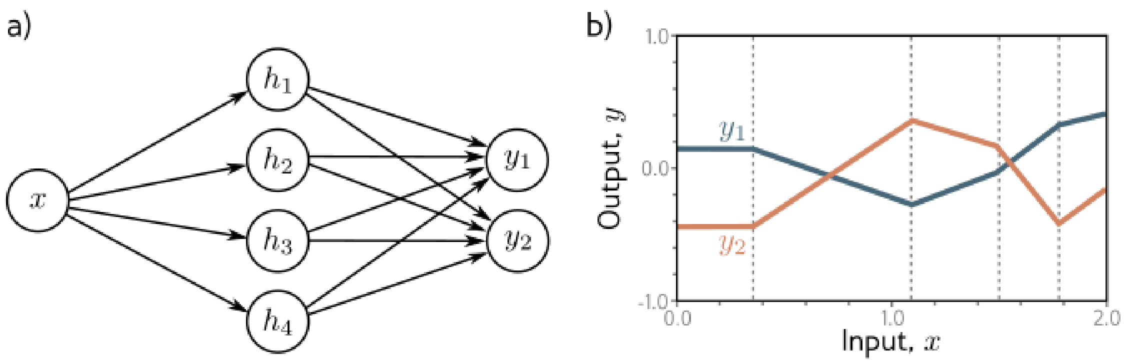
Depicting Neural Networks

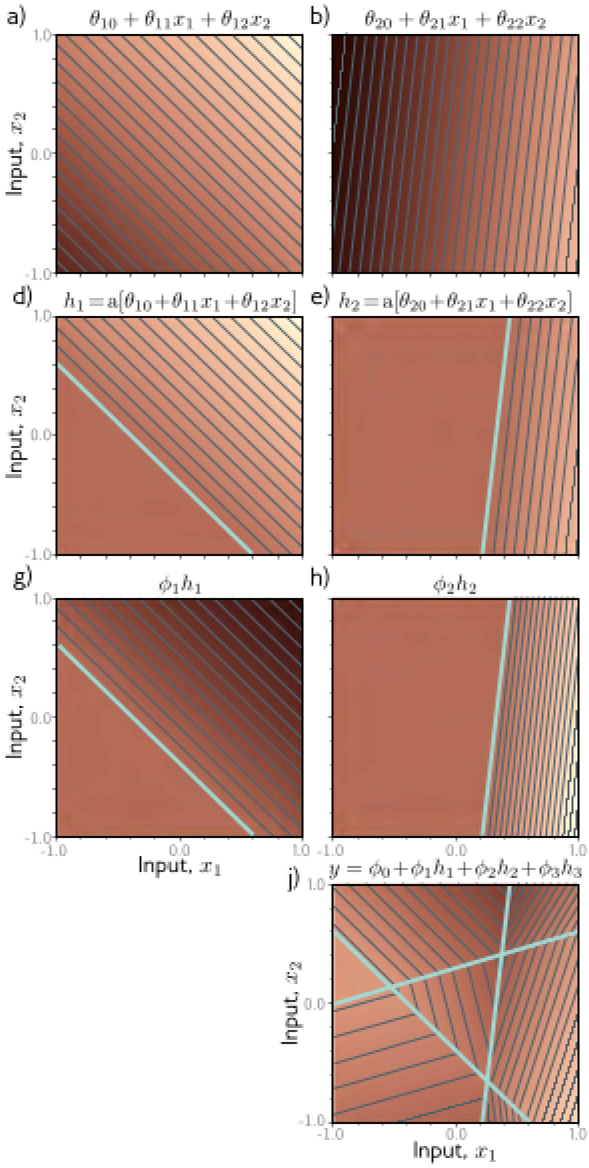


Universal Approximation Theorem

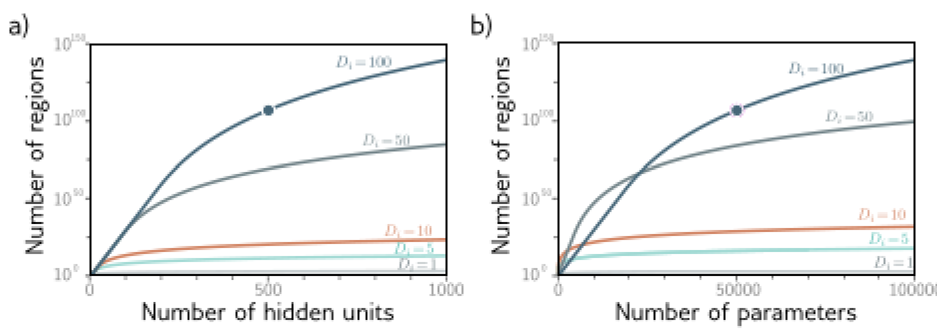


Multivariate outputs

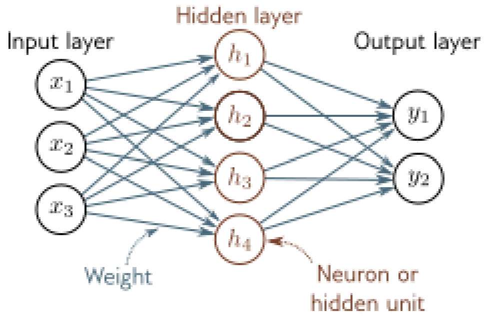




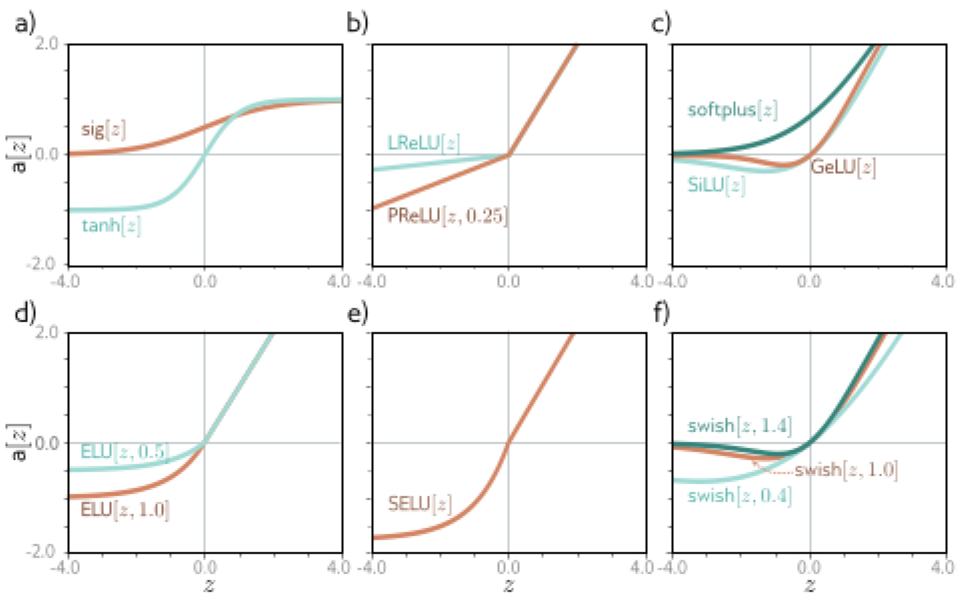
a) $\theta_{10}+\theta_{11}x_1+\theta_{12}x_2$
b) $\theta_{20}+\theta_{21}x_1+\theta_{22}x_2$
c) $\theta_{30}+\theta_{31}x_1+\theta_{32}x_2$
d) $h_1=\mathrm{a}[\theta_{10}+\theta_{11}x_1+\theta_{12}x_2]$
e) $h_2=\mathrm{a}[\theta_{20}+\theta_{21}x_1+\theta_{22}x_2]$
f) $h_3=\mathrm{a}[\theta_{30}+\theta_{31}x_1+\theta_{32}x_2]$
g) $\phi_1 h_1$
h) $\phi_2 h_2$
i) $\phi_3 h_3$
j) $y=\phi_0+\phi_1h_1+\phi_2h_2+\phi_3h_3$
Input, x_1
Input, x_2
-1.0
0.0
1.0



a)
b)
Number of regions
Number of hidden units
Number of parameters
$D_i=100$
$D_i=50$
$D_i=10$
$D_i=5$
$D_i=1$
10^{150}
10^{100}
10^{50}
10^{0}
0
500
1000
50000
100000



Activation functions



Sigmoid

$$\sigma(z) = \frac{1}{1 + exp(-z)}$$

Hyperbolic tangent

$$\tanh(z) = \frac{exp(z) - exp(-z)}{exp(z) + exp(-z)}$$

Parametric Rectified Linear Unit

$$\mathrm{PReLU}(z, \alpha) = \begin{cases} z & \text{if } z > 0 \\ \alpha z & \text{if } z \leq 0 \end{cases}$$

Leaky Rectified Linear Unit

$$\mathrm{LReLU}(z) = \mathrm{PReLU}(z, \alpha = 0.01)$$

Softplus

$$\mathrm{softplus}(z) = \frac{1}{\beta}\log(1 + \exp(\beta z))$$

Gaussian error Linear Units

$$\mathrm{GELU}(z) = z\Phi(z)$$

where $\Phi(z)$ is the error function or the cumulative distribution function of a Gaussian distribution.

Sigmoid Linear Unit

$$\mathrm{SiLU}(z) = z\sigma(z)$$

Exponential Linear Unit

$$\mathrm{ELU}(z, \alpha) == \begin{cases} z & \text{if } z > 0 \\ \alpha(\exp(z) - 1) & \text{if } z \leq 0 \end{cases}$$

Scaled exponential linear unit

$$\mathrm{SELU}(z) = 1.0507 * \mathrm{ELU}(z, 1.673)$$

Swish

$$\mathrm{Swish}(z, \beta) = z\sigma(\beta z)$$

HardSwish

$$\mathrm{HardSwish}(z) = \begin{cases} 0 & z < -3 \\ z(z+3)/6 & -3 \leq z \leq 3 \\ z & z > 3 \end{cases}$$