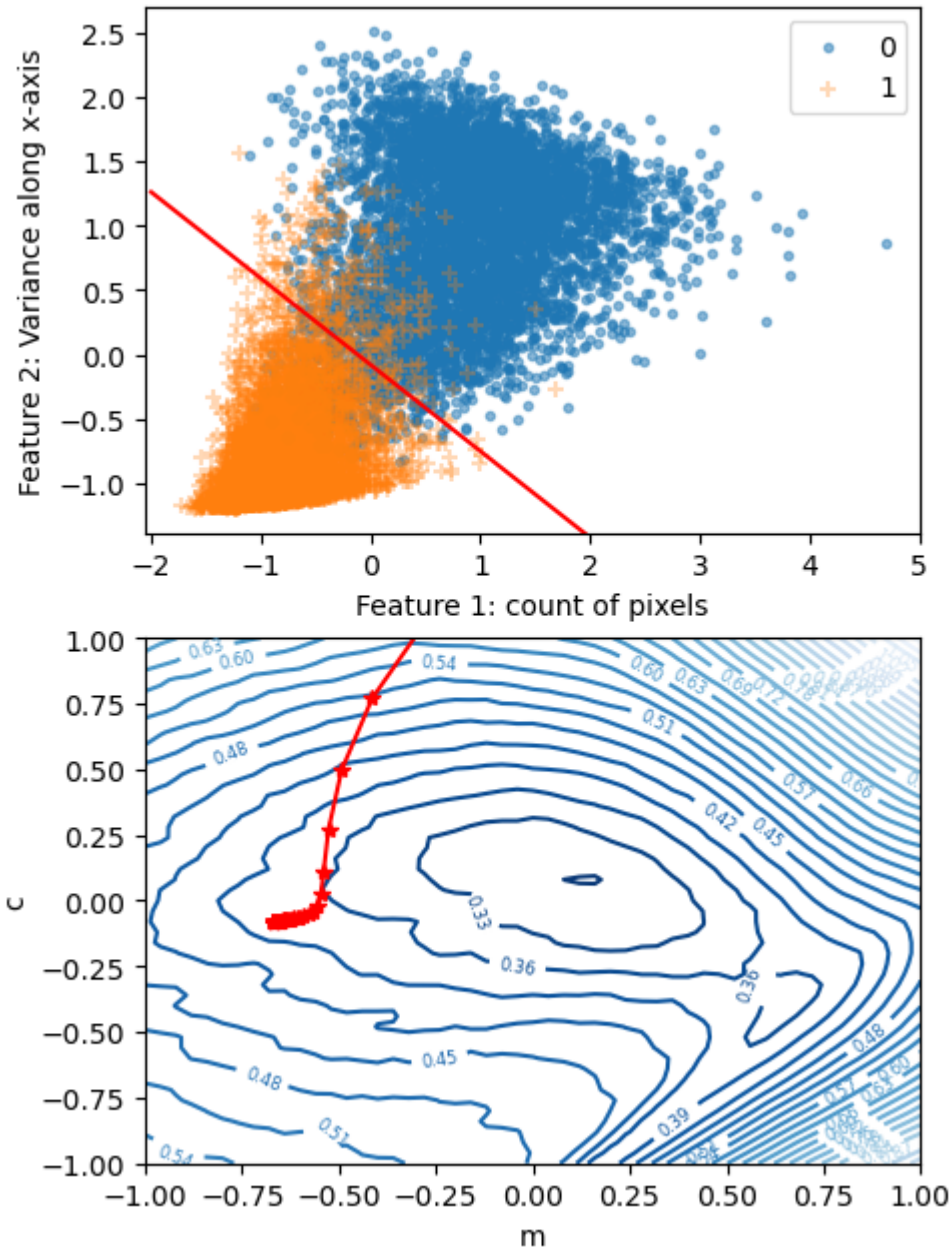
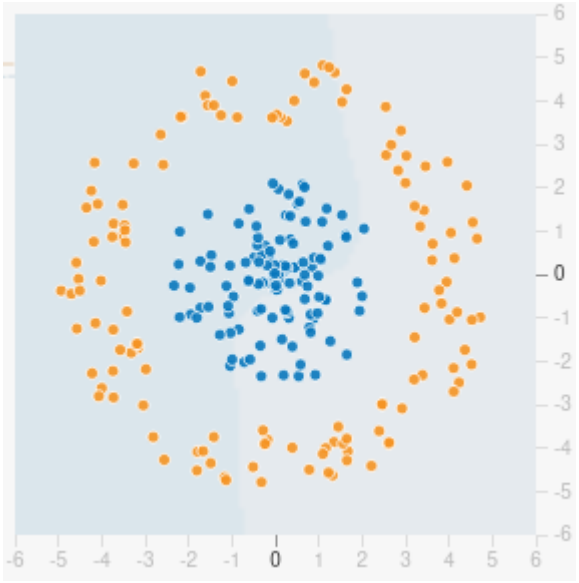
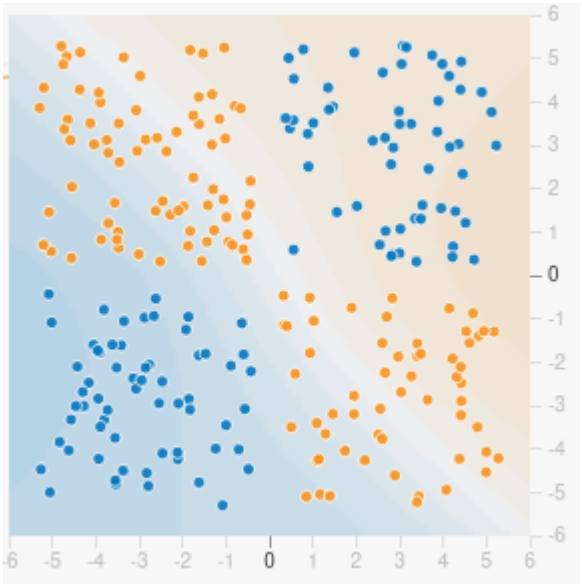
Multi layer Perceptron

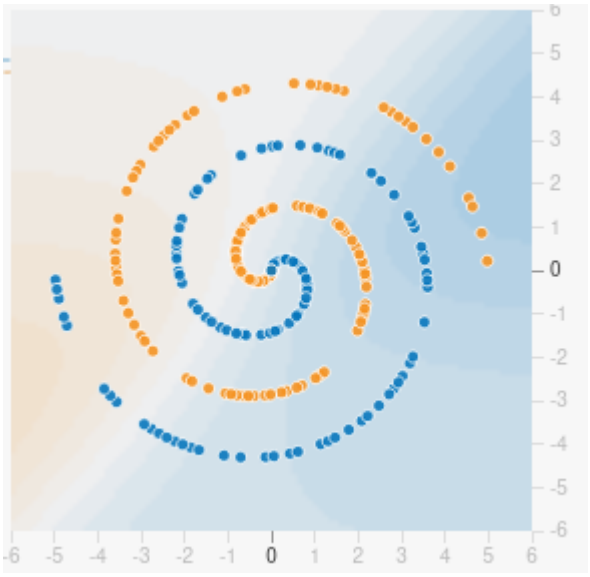
All figures are from Chapter 3 of UDLBook. https://github.com/udlbook/udlbook

Recall the single layer perceptron




6 5 4 3 2 1 0 -1 -2 -3 -4 -5 -6
-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6

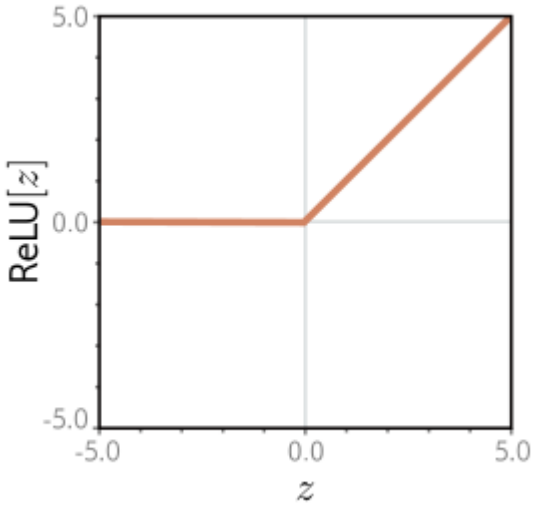

6 5 4 3 2 1 0 -1 -2 -3 -4 -5 -6
-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6


6 5 4 3 2 1 0 -1 -2 -3 -4 -5 -6
-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6

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

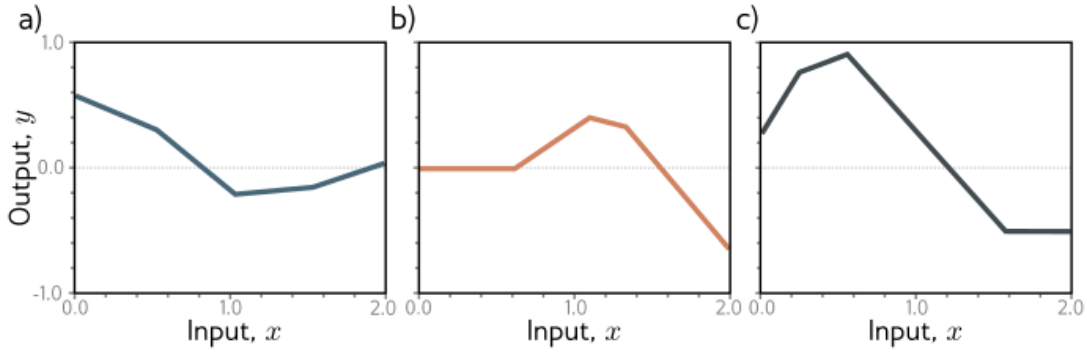
Multi Layer Perceptrons

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



Two layer Perceptron

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



Example

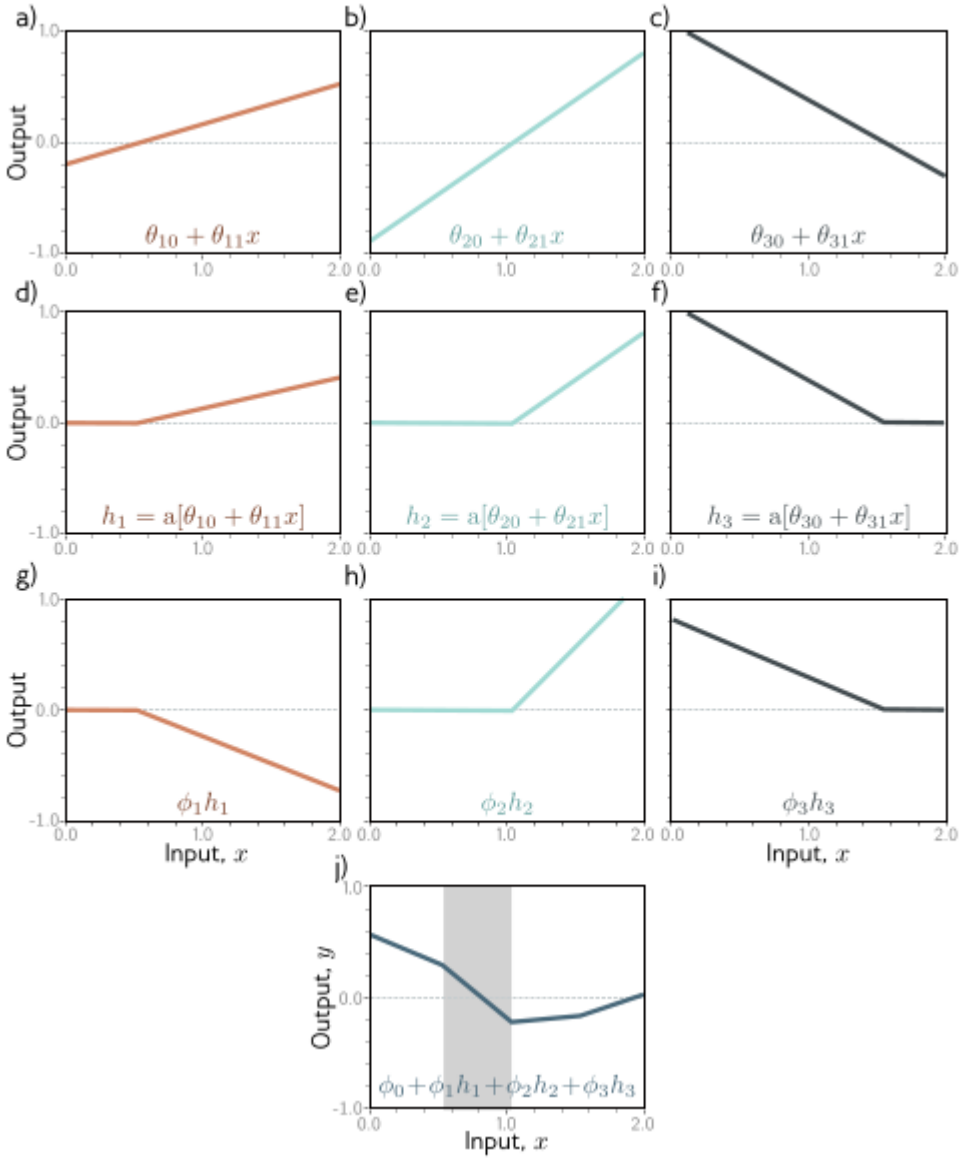
$$a(x) = \text{ReLU}(x)$$

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

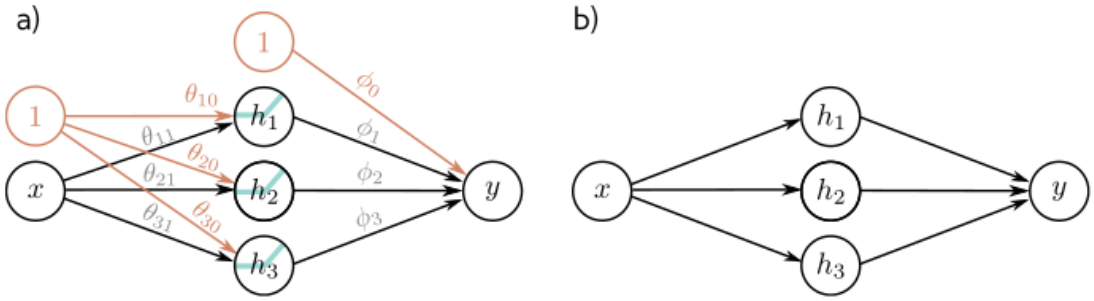
$$h_2 = a(\theta_{20} + \theta_{21}x)$$

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

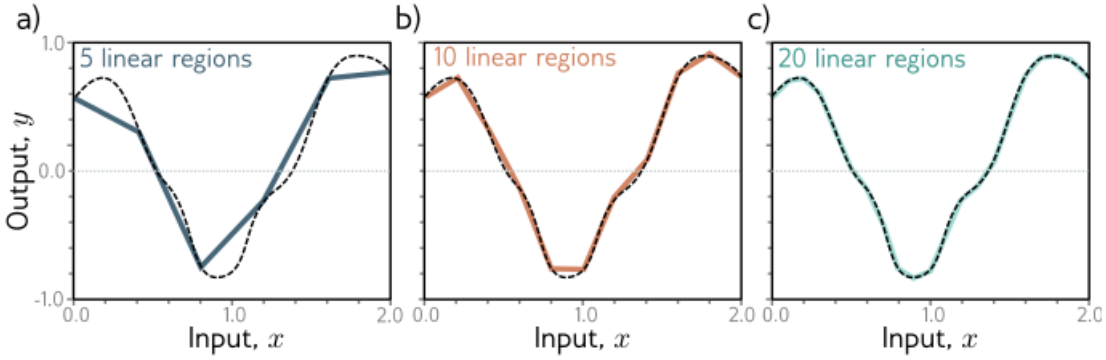
$$y = \phi_0 + \phi_1 h_1 + \phi_2 h_2 + \phi_3 h_3$$



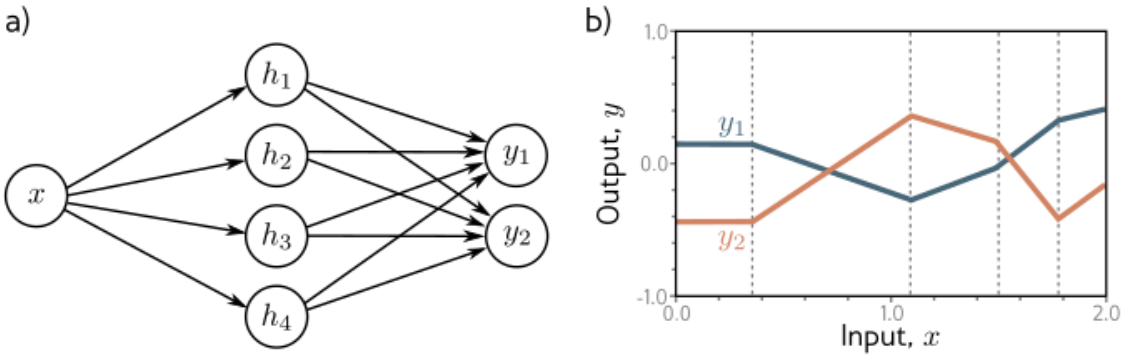
Depicting Neural Networks

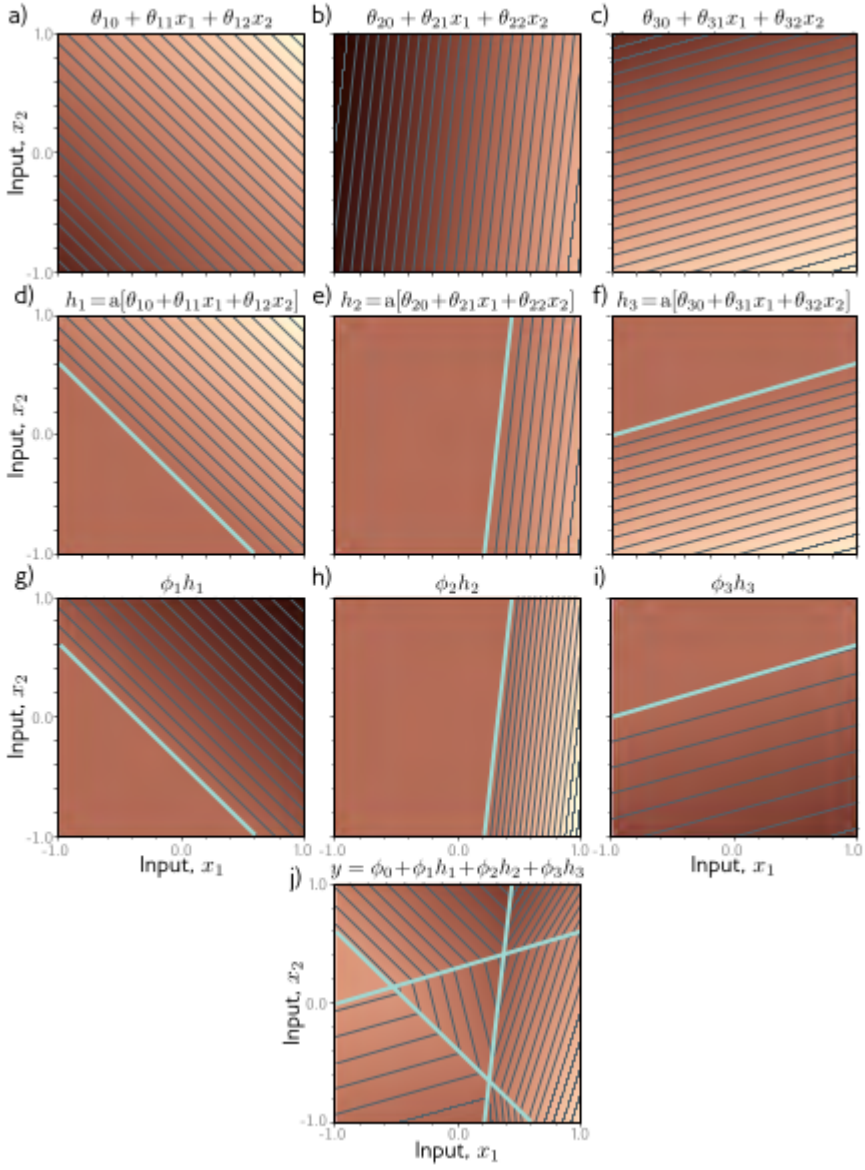


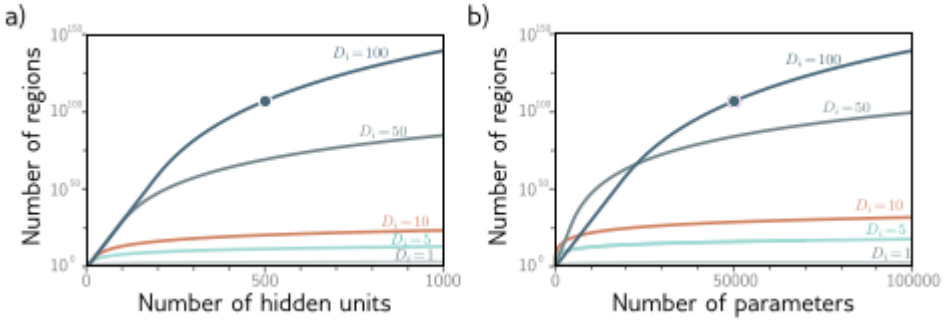
Universal Approximation Theorem

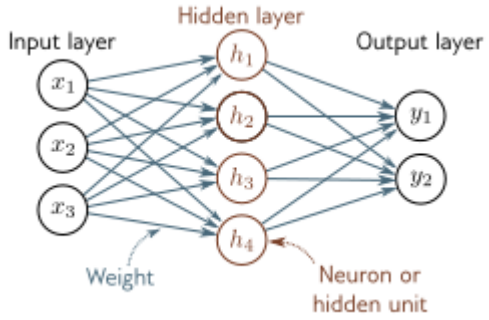


Multivariate outputs

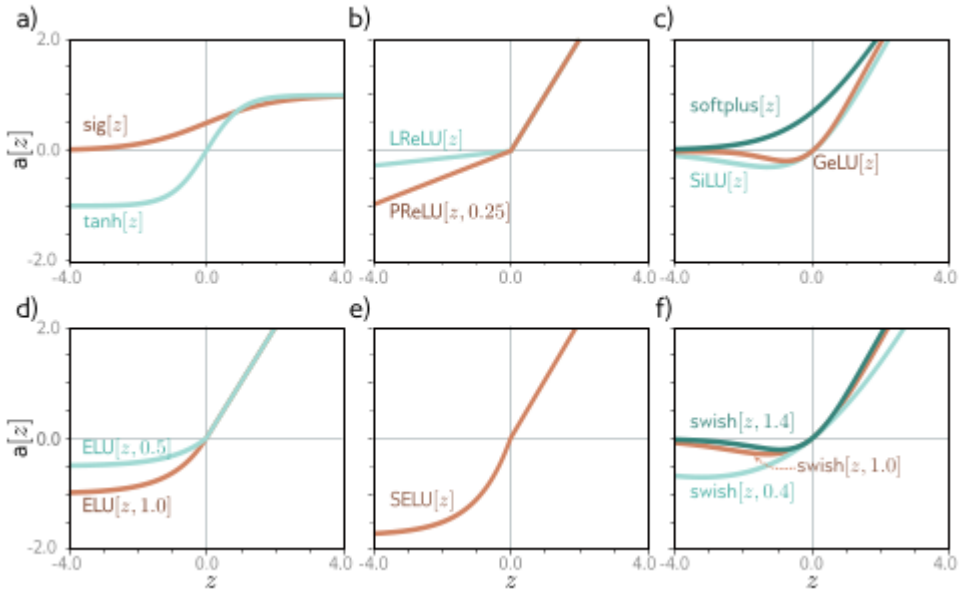



a) $\theta_{10}+\theta_{11}x_1+\theta_{12}x_2$
b) $\theta_{20}+\theta_{21}x_1+\theta_{22}x_2$
c) $\theta_{30}+\theta_{31}x_1+\theta_{32}x_2$
d) $h_1=a[\theta_{10}+\theta_{11}x_1+\theta_{12}x_2]$
e) $h_2=a[\theta_{20}+\theta_{21}x_1+\theta_{22}x_2]$
f) $h_3=a[\theta_{30}+\theta_{31}x_1+\theta_{32}x_2]$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


a)
b)
Number of regions
Number of hidden units
Number of parameters
$D_i=100$
$D_i=50$
$D_i=10$
$D_i=5$
$D_i=1$
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

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

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

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