Backpropagation

Initialize all weights in all units to random value

Repeat until termination condition

For each data point $\mathbf{x_i}, t_i$

Compute z_k and y_ℓ for every unit in the network

For each output unit y_ℓ

$\delta_{y_\ell} = y_\ell(\mathbf{x_i})(1 - y_\ell(\mathbf{x_i}))(t_i - y_\ell(\mathbf{x_i}))$

For each hidden unit z_k

$\delta_{z_k} = z_k(\mathbf{x_i})(1 - z_k(\mathbf{x_i})) \sum_{\ell=1}^{K} w_{\ell k} \delta_{y_\ell}$

For each output unit y_ℓ

For each weight $w_{y_\ell j}$

$w_{y_\ell k} += \eta \delta_{y_\ell} z_k(\mathbf{x_i})$

For each hidden unit z_k

For each weight $w_{z_k j}$

$w_{z_k j} += \eta \delta_{z_k} x_{ij}$

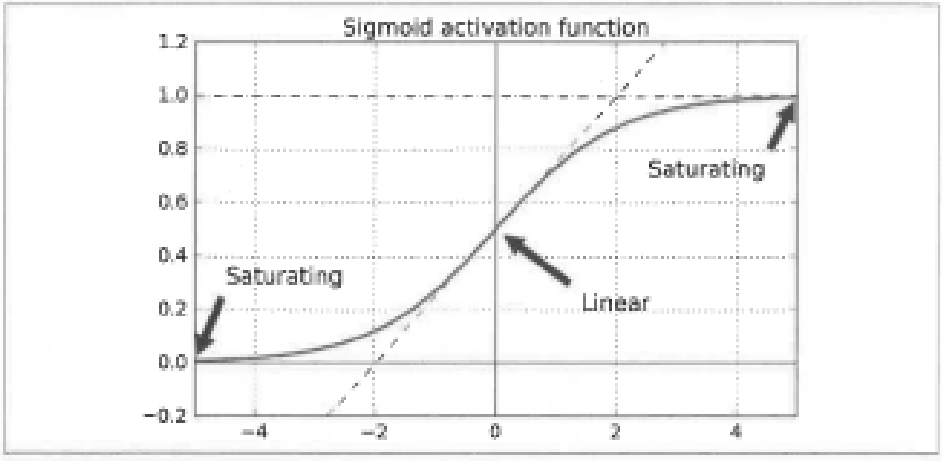

Figure 11-1. Logistic activation function saturation

Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn & TensorFlow*, O'Reilly, 2017. Pg 279.

Table 11-1. Initialization parameters for each type of activation function

Activation function	Uniform distribution [−r, r]	Normal distribution
Logistic	$r = \sqrt{\frac{6}{n_{inputs} + n_{outputs}}}$	$\sigma = \sqrt{\frac{2}{n_{inputs} + n_{outputs}}}$
Hyperbolic tangent	$r = 4\sqrt{\frac{6}{n_{inputs} + n_{outputs}}}$	$\sigma = 4\sqrt{\frac{2}{n_{inputs} + n_{outputs}}}$
ReLU (and its variants)	$r = \sqrt{2}\sqrt{\frac{6}{n_{inputs} + n_{outputs}}}$	$\sigma = \sqrt{2}\sqrt{\frac{2}{n_{inputs} + n_{outputs}}}$

Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn & TensorFlow*, O'Reilly, 2017. Pg 280.

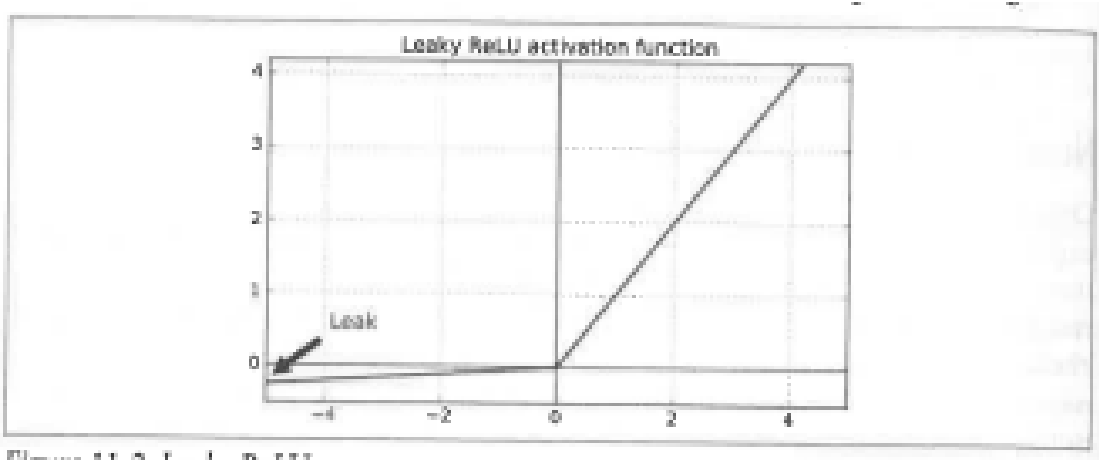

Figure 11-2. Leaky ReLU

Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn & TensorFlow*, O'Reilly, 2017. Pg 282.

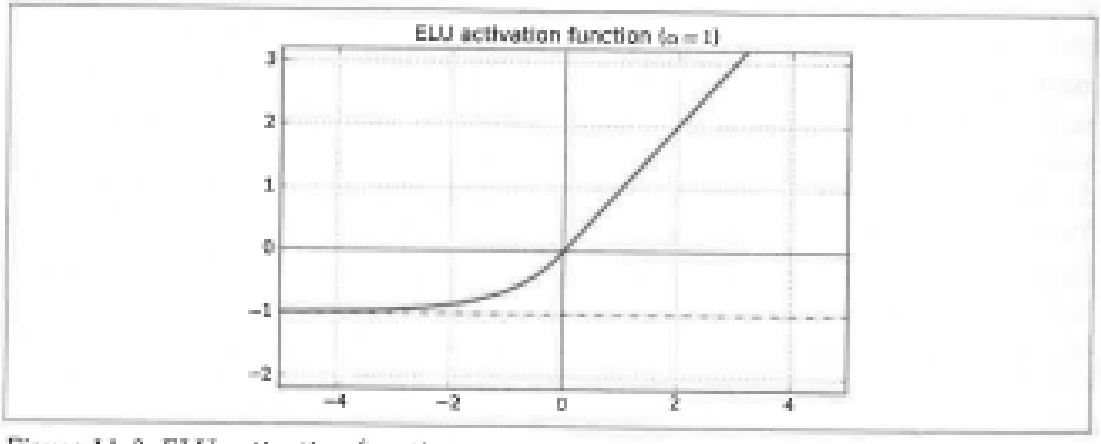

Figure 11-3. ELU activation function

Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn & TensorFlow*, O'Reilly, 2017. Pg 282.