Delta rule

From Wikipedia, the free encyclopedia

In machine learning, the **delta rule** is a gradient descent learning rule for updating the weights of the inputs to artificial neurons in single-layer neural network.[1] It is a special case of the more general backpropagation algorithm. For a neuron j with activation function $g(x)$, the delta rule for j's ith weight w_{ji} is given by

$$\Delta w_{ji} = \alpha(t_j - y_j)g'(h_j)x_i,$$

where

α is a small constant called *learning rate*
$g(x)$ is the neuron's activation function
t_j is the target output
h_j is the weighted sum of the neuron's inputs
y_j is the actual output
x_i is the ith input.

It holds that $h_j = \sum x_i w_{ji}$ and $y_j = g(h_j)$.

The delta rule is commonly stated in simplified form for a neuron with a linear activation function as

$$\Delta w_{ji} = \alpha(t_j - y_j)x_i$$

While the delta rule is similar to the perceptron's update rule, the derivation is different. The perceptron uses the Heaviside step function as the activation function $g(h)$, and that means that $g'(h)$ does not exist at zero, and is equal to zero elsewhere, which makes the direct application of the delta rule impossible.

Derivation of the delta rule

The delta rule is derived by attempting to minimize the error in the output of the neural network through gradient descent. The error for a neural network with j outputs can be measured as

$$E = \sum_{j} \frac{1}{2}(t_j - y_j)^2.$$

In this case, we wish to move through "weight space" of the neuron (the space of all possible values of all of the neuron's weights) in proportion to the gradient of the error function with respect to each weight. In order to do that, we calculate the partial derivative of the error with respect to each weight. For the ith weight, this derivative can be written as

$$\frac{\partial E}{\partial w_{ji}}.$$

Because we are only concerning ourselves with the jth neuron, we can substitute the error formula above while omitting the summation:

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial \left(\frac{1}{2} \left(t_j - y_j \right)^2 \right)}{\partial w_{ji}}$$

Next we use the chain rule to split this into two derivatives:

$$= \frac{\partial \left(\frac{1}{2} \left(t_j - y_j \right)^2 \right)}{\partial y_j} \frac{\partial y_j}{\partial w_{ji}}$$

To find the left derivative, we simply apply the general power rule:

$$= - \left(t_j - y_j \right) \frac{\partial y_j}{\partial w_{ji}}$$

To find the right derivative, we again apply the chain rule, this time differentiating with respect to the total input to j, h_j:

$$= - \left(t_j - y_j \right) \frac{\partial y_j}{\partial h_j} \frac{\partial h_j}{\partial w_{ji}}$$

Note that the output of the jth neuron, y_j, is just the neuron's activation function g applied to the neuron's input h_j. We can therefore write the derivative of y_j with respect to h_j simply as g's first derivative:

$$= - \left(t_j - y_j \right) g'(h_j) \frac{\partial h_j}{\partial w_{ji}}$$

Next we rewrite h_j in the last term as the sum over all k weights of each weight w_{jk} times its corresponding input x_k:

$$= - \left(t_j - y_j \right) g'(h_j) \frac{\partial \left(\sum_k x_k w_{jk} \right)}{\partial w_{ji}}$$

Because we are only concerned with the ith weight, the only term of the summation that is relevant is $x_i w_{ji}$. Clearly,

$$\frac{\partial x_i w_{ji}}{\partial w_{ji}} = x_i,$$

giving us our final equation for the gradient:

$$\frac{\partial E}{\partial w_{ji}} = - \left(t_j - y_j \right) g'(h_j) x_i$$

As noted above, gradient descent tells us that our change for each weight should be proportional to the gradient. Choosing a proportionality constant α and eliminating the minus sign to enable us to move the weight in the negative direction of the gradient to minimize error, we arrive at our target equation:

$$\Delta w_{ji} = \alpha(t_j - y_j)g'(h_j)x_i.$$

See also

- Stochastic gradient descent
- Backpropagation

References

1. Russell, Ingrid. "The Delta Rule" (http://uhavax.hartford.edu/compsci/neural-networks-delta-rule.html). University of Hartford. Retrieved 5 November 2012.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Delta_rule&oldid=611122358"

Categories: Artificial neural networks

- This page was last modified on 1 June 2014, at 19:22.
-