# Loss



where and are the weights and biases in the network, the regularization strength.

# Activation

## Hyperbolic tangent (tanh)





# Example

We use the following example to test this implementation.



The regressor vector is thus



# Gradient descent

## Naive stochastic gradient descent

where is the learning rate.

## Momentum

where is the forgetting factor leading to exponential averaging.