Deep Neural Networks

Apprentissage par réseaux de neurones artificiels

ANNEN Rayane, MARTINS Alexis

HEIG VD

12.04.2023

Contents

Digit recognition from raw data

What is the learning algorithm being used to optimize the weights of the neural networks? What are the parameters (arguments) being used by that algorithm? What cost function is being used ? please, give the equation(s)

The algorithm used to optimize the weight is RMSprop (Root Mean Square Propagation).

The parameters are the following:

```
tf.keras.optimizers.RMSprop(
    learning_rate=0.001,
    rho=0.9,
    momentum=0.0,
    epsilon=1e-07,
    centered=False,
    weight_decay=None,
    clipnorm=None,
    clipvalue=None,
    global_clipnorm=None,
    use_ema=False,
    ema_momentum=0.99,
    ema_overwrite_frequency=100,
    jit_compile=True,
    name="RMSprop",
    **kwargs
)
```

The following equations are used:

$$E[g^2]_t = \beta E[g^2]_{t-1} + (1-\beta)\left(\frac{\partial C}{\partial w}\right)^2$$

$$w_t = w_{t-1} - \frac{\eta}{\sqrt{E[g^2]_t}}\frac{\partial C}{\partial w}$$

where η is the learning rate, w_t the new weight, β is the moving average parameter, $E[g]$ is the moving average of squared gradients and $\frac{\partial C}{\partial w}$ is the derivative of the cost function with respect to the weight.

The cost function is the categorical cross-entropy loss function :

$$\mathsf{CE} = -\frac{1}{N}\sum_{k=0}^{N} \log \vec{p_i}[y_i]$$

where N is the number of samples, $\vec{p_i}$ is the neural network output and y_i is the target class index.