
Algorithm 2: OnMiniBatchUpdate_Callback.

Input: $MODEL$ as the full-precision input model

Input: E_{size} as the target exponent bits size

Input: M_{size} as the target mantissa bits size

Input: D_{train} as the training data set

Input: D_{val} as the validation data set

Input: N_{ep} as the number of epochs

Input: B_{size} as the mini-batch size

Output: $MODEL$ as the quantized output model

// Quantize

1 $MODEL \leftarrow \text{QuantizeTraining}(MODEL, E_{size}, M_{size})$

2 **if** $1 < epoch$ **then**

 // Update model after first epoch

3 $mse_v \leftarrow \text{Evaluate}(MODEL, D_{val})$

4 **if** $mse_v < mse_i$ **then**

5 Update(MODEL)

6 $mse_i \leftarrow mse_v$

7 **end**

8 **end**
