



Linköping University

TDDC17 ARTIFICIAL INTELLIGENCE **Lab 6: Deep Learning**

Robin Andersson (roban591)
Lawrence Thanakumar Rajappa (lawra776)

October 15, 2019

Part 2

Q1. In the gradient descent learning code below, please complete the gradient computation by inputting the correct variables where there are question marks. We are using Tensorflow to automatically compute the gradient with GradientTape (tape) similar to how you were taught above, but to do supervised learning which tensorflow node do we compute the gradient of, and with regard to which variable? Please fill this in below.

Part 3

Q2. Show the math for why the first Dense layer has 100,480 parameters with these inputs and number of neurons

The input for the first layer is (28,28) and number of neurons in the dense(hidden) layer is 128 neurons, Number of biases for hidden layer is 128 neurons, hence

$$(\text{Number of inputs} * \text{Number of Neurons in the hidden layer}) + \text{Number of biases for hidden layer} \\ (28*28*128)+128 = 100,480 \text{ parameters.}$$

Q3. Here you will evaluate different mini-batch sizes for stochastic gradient descent (see the deep learning lecture). Please separately run the training code above with batch sizes of 1, 10, 100, 1000 and 60000. Write down the training times (you can use the first number in seconds, not the per sample time) and the training set accuracy reached, both in the first line of the output. This can randomly vary a bit between runs but it should give you an idea. In your lab report, plot both curves and reason about which batch size produced the most accuracy given the time spent, i.e. which batch size would be best to start the training with?

| Batch Size | Time Duration in (seconds) | Accuracy % |
|------------|----------------------------|------------|
| 1 | 109 | 81.23 |
| 10 | 15 | 82.56 |
| 100 | 3 | 81.78 |
| 1000 | 2 | 72.81 |
| 60000 | 2 | 6.16 |