# AI5031: Machine learning, exercise sheet 9

## 1 Implementation of gradient descent by hand

Allocate a TF variable with two elements, which we shall denote $w_1$ and $w_2$. Implement a loss function $\mathcal{L} = w_1^2 + w_2^2$. Create an instance of GradientDescentOptimizer, and use it to perform 20 iterations of gradient descent with a step size of $\epsilon = 0.1$. Compare the first three steps to a manual gradient descent calculation!

## 2 Training a DNN

Train a DNN with three hidden affine+ReLU layers of sizes 200-200-200 on MNIST, using a batch size of 50. Training should be conducted for 200 iterations with a learning rate of 0.01. After each training step, the classification error on the test set should be computed and displayed. It is advantageous to use the *AdamOptimizer* class for performing SGD.

## 3 Training a DNN, outlook

**a)** Use a learning rate of 0.1, what do you observe? And what do you think is the reason? **b)** Use a learning batch size of 1, what do you observe? And what do you think is the reason? **c)** Reduce the layer size to 20, what do you observe?