

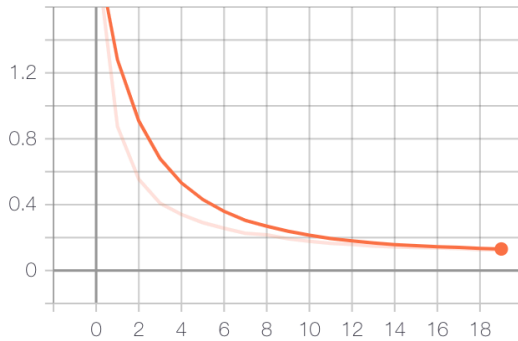
Exercise 1: MLP in Eager Mode – Group 24

Sebastian Häni haeniseb@students.zhaw.ch, Raffael Affolter affolraf@students.zhaw.ch, Benjamin Mäder maedeben@students.zhaw.ch

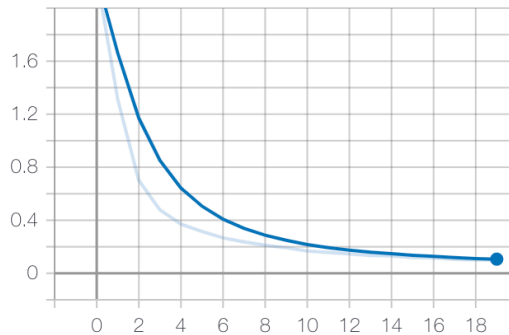
Training with `stdev_scale_glorot`

Layers: 50, 50, 50, 10

avg_test_loss
tag: avg_test_loss

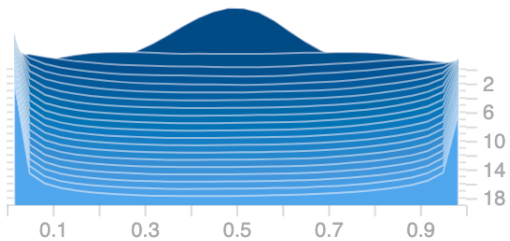


avg_train_loss
tag: avg_train_loss



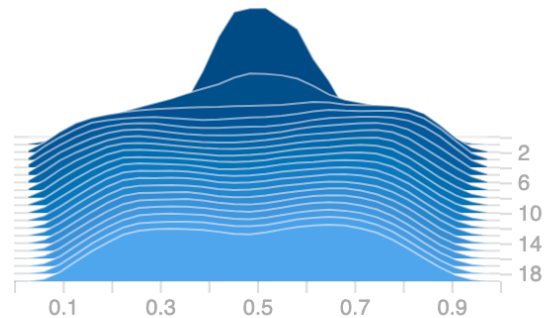
activ_0
tag: activ_0

20191110-164511/glorot_initialized/train



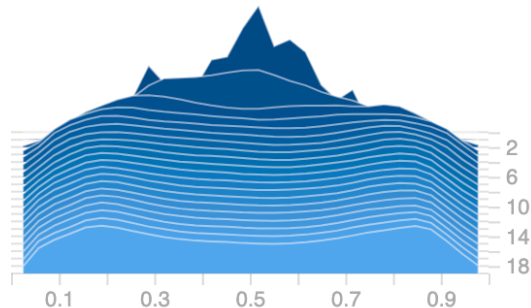
activ_1
tag: activ_1

20191110-164511/glorot_initialized/train



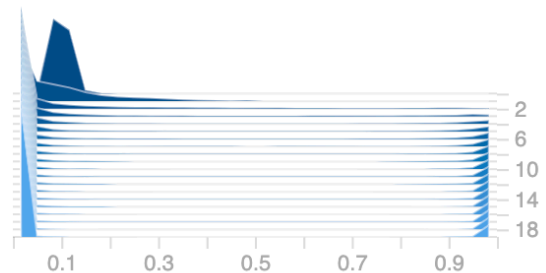
activ_2
tag: activ_2

20191110-164511/glorot_initialized/train



activ_3
tag: activ_3

20191110-164511/glorot_initialized/train

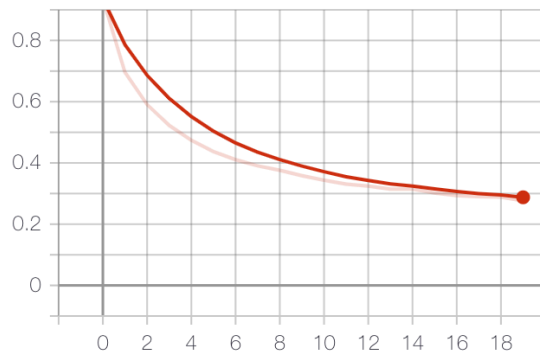


⇒ The activations are not saturated with Glorot initialization.

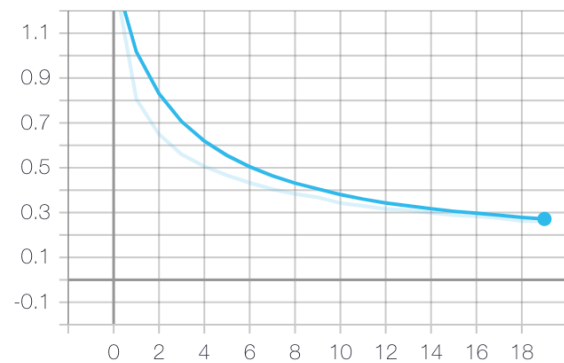
Training with no std scale initialization

Layers: 50, 50, 50, 10

avg_test_loss
tag: avg_test_loss

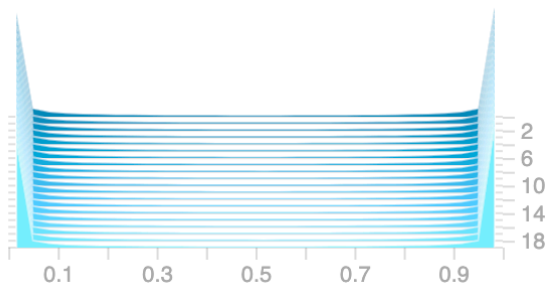


avg_train_loss
tag: avg_train_loss



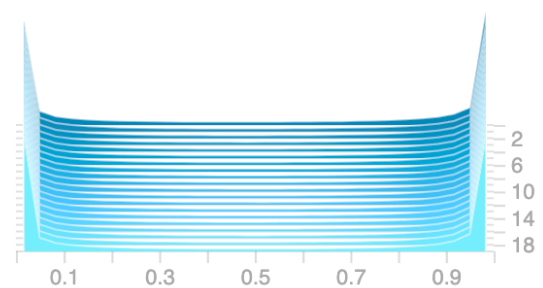
activ_0

20191110-170252/no_scaled_stdev_initialization/train
tag: activ_0



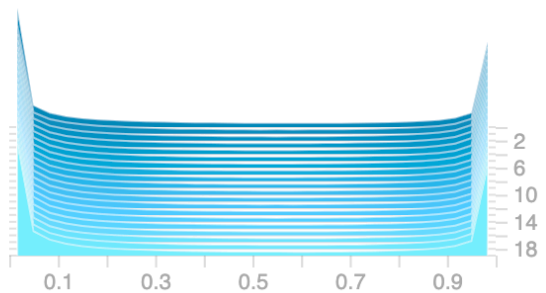
activ_1

20191110-170252/no_scaled_stdev_initialization/train
tag: activ_1



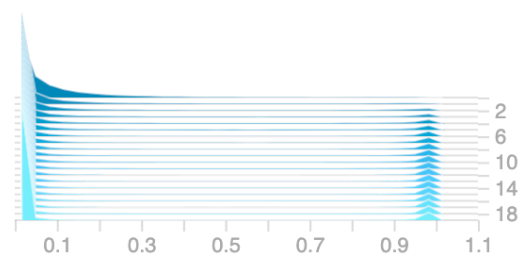
activ_2

20191110-170252/no_scaled_stdev_initialization/train
tag: activ_2



activ_3

20191110-170252/no_scaled_stdev_initialization/train
tag: activ_3

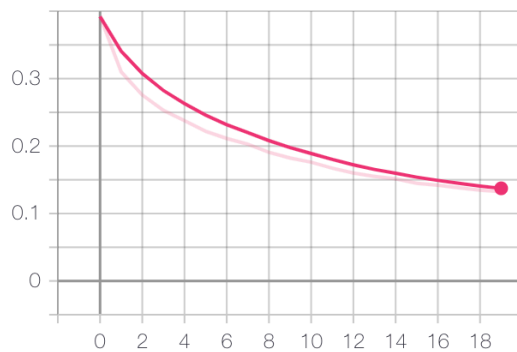


⇒ The activations clearly are saturated.

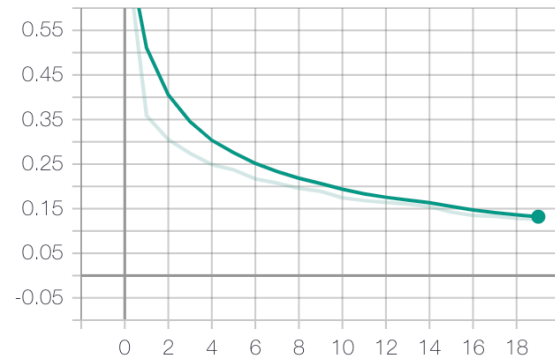
Simple model

Now we try what happens if we simplify the model to 1 hidden layer of 50 units and an output of 10 units. We use Glorot initialization

avg_test_loss
tag: avg_test_loss

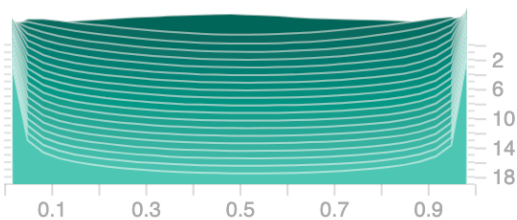


avg_train_loss
tag: avg_train_loss



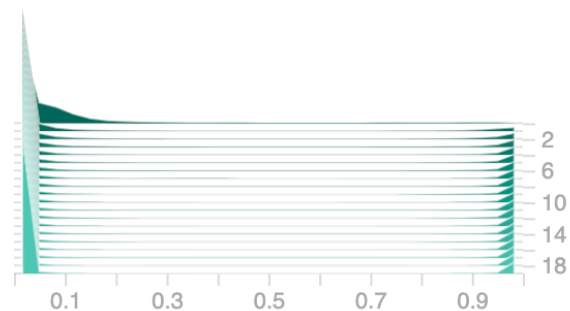
activ_0
tag: activ_0

20191110-171741/simple_model/train



activ_1
tag: activ_1

20191110-171741/simple_model/train

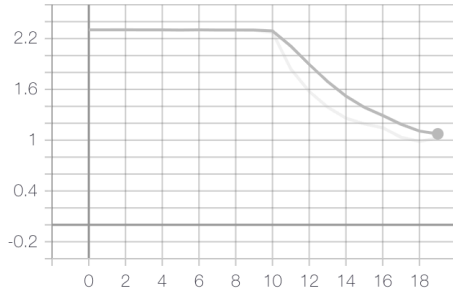


⇒ The activations are clearly saturated even though we used Glorot initialization. The model didn't learn well because it cannot cope with the data complexity as it's too simple.

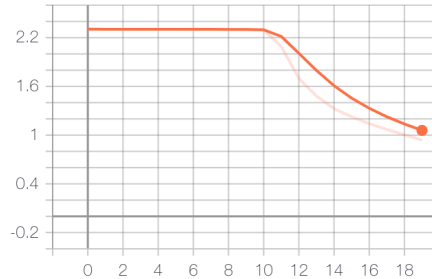
Complex model

Now we try what happens if we use a model with 5 hidden layers of 20 units and an output of 10 units. We use Glorot initialization.

avg_test_loss
tag: avg_test_loss

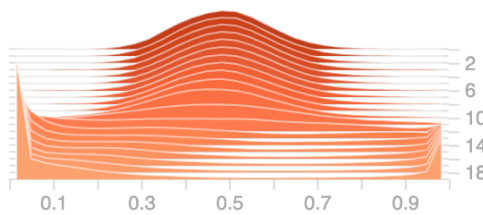


avg_train_loss
tag: avg_train_loss



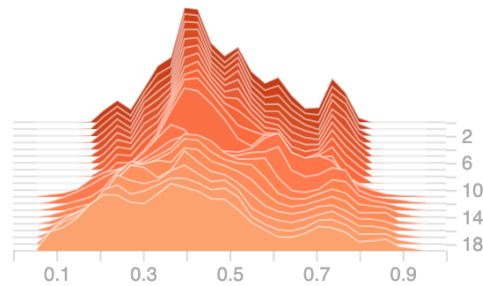
activ_0
tag: activ_0

20191110-173152/complex_model/train



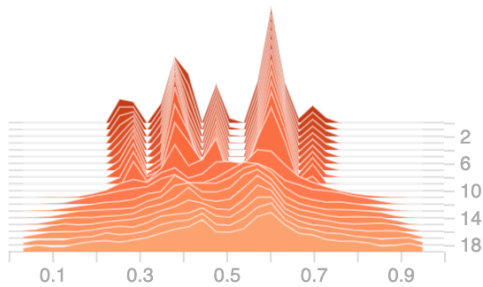
activ_1
tag: activ_1

20191110-173152/complex_model/train



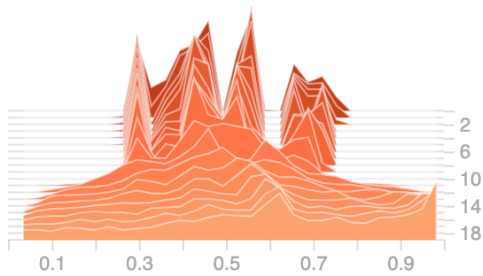
activ_2
tag: activ_2

20191110-173152/complex_model/train



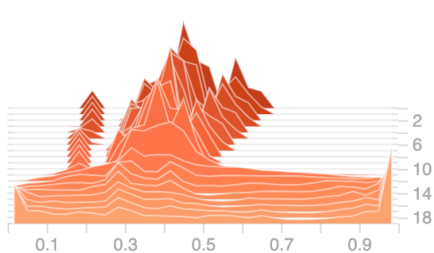
activ_3
tag: activ_3

20191110-173152/complex_model/train



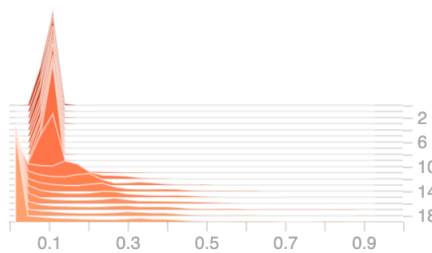
activ_4
tag: activ_4

20191110-173152/complex_model/train



activ_5
tag: activ_5

20191110-173152/complex_model/train



⇒ Even though there is no saturation, the layers had too few units to properly fit the model.