

Generalization error at optimal regularisation

The plot shows the generalisation error at optimally chosen regularization parameter $\lambda \in [1e-5, 1e5]$ during the NN gradient descent (initialization emphasized) and for ridge regression (dashed).

Epoch



10m

100m

1

10

↑ Generalization error

