Practical Machine Learning
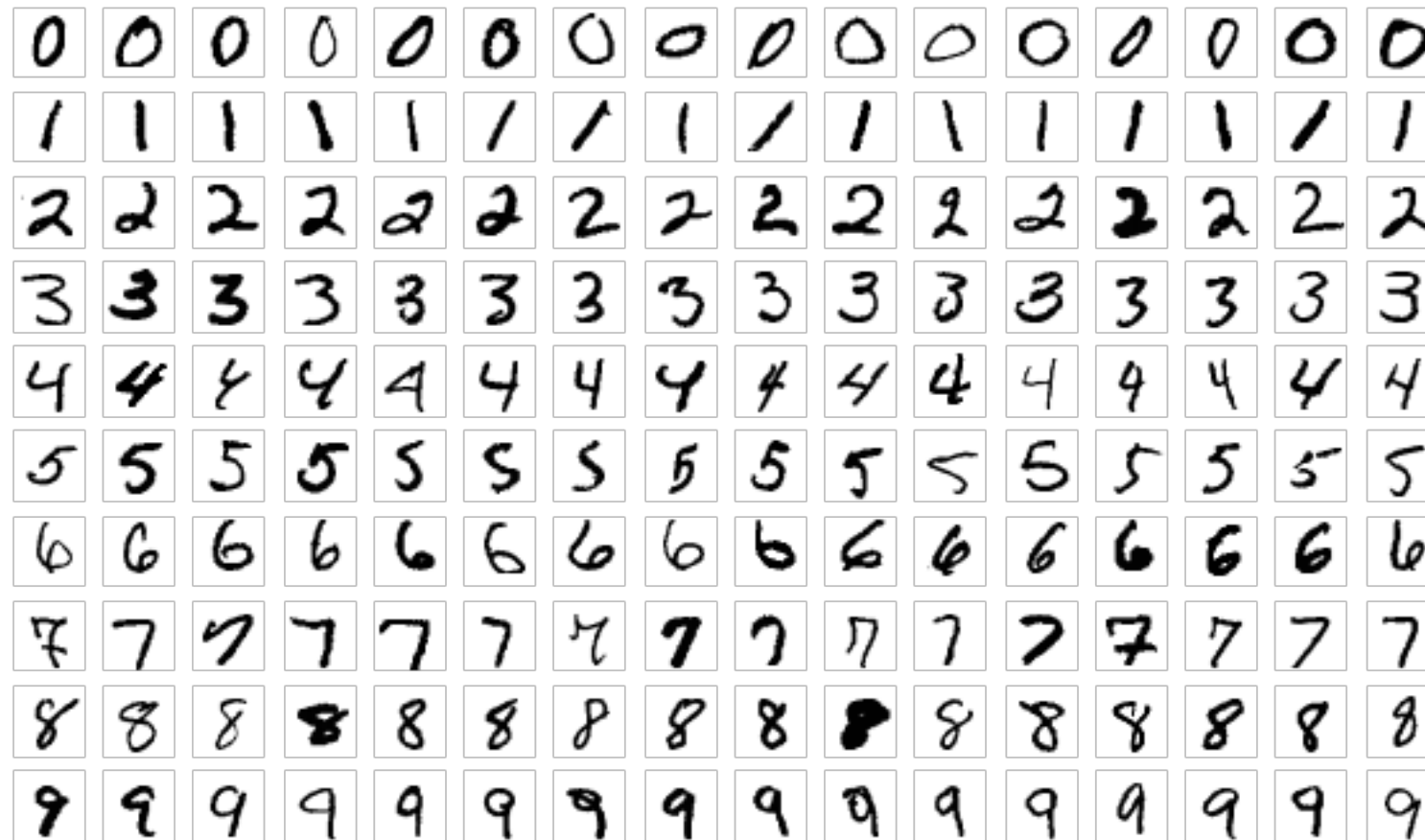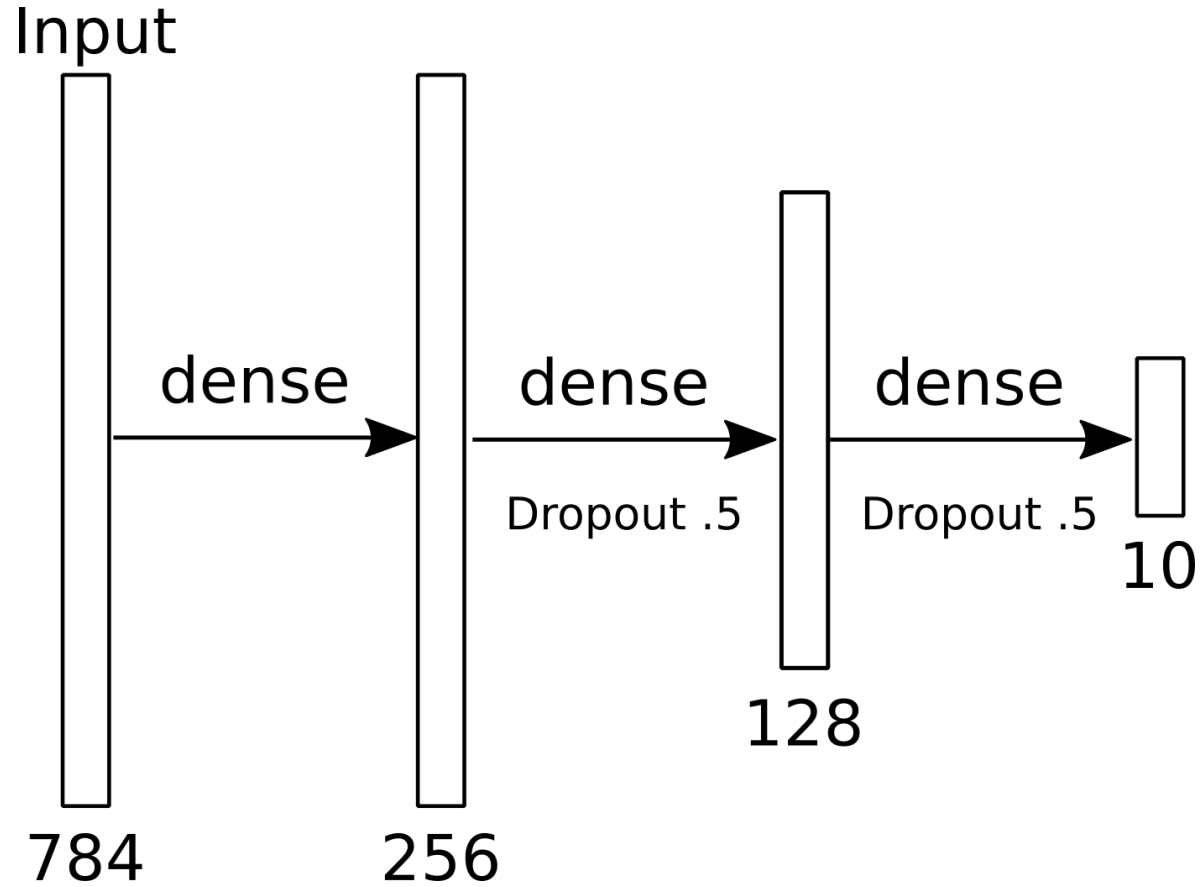
Optimizer and Hyperparameter

Sven Mayer

Impact of Hyperparameter on the Models

MNIST data as showcase

Sven Mayer

Test Model

Input

dense → dense → dense →

784   256   128   10

Dropout .5   Dropout .5
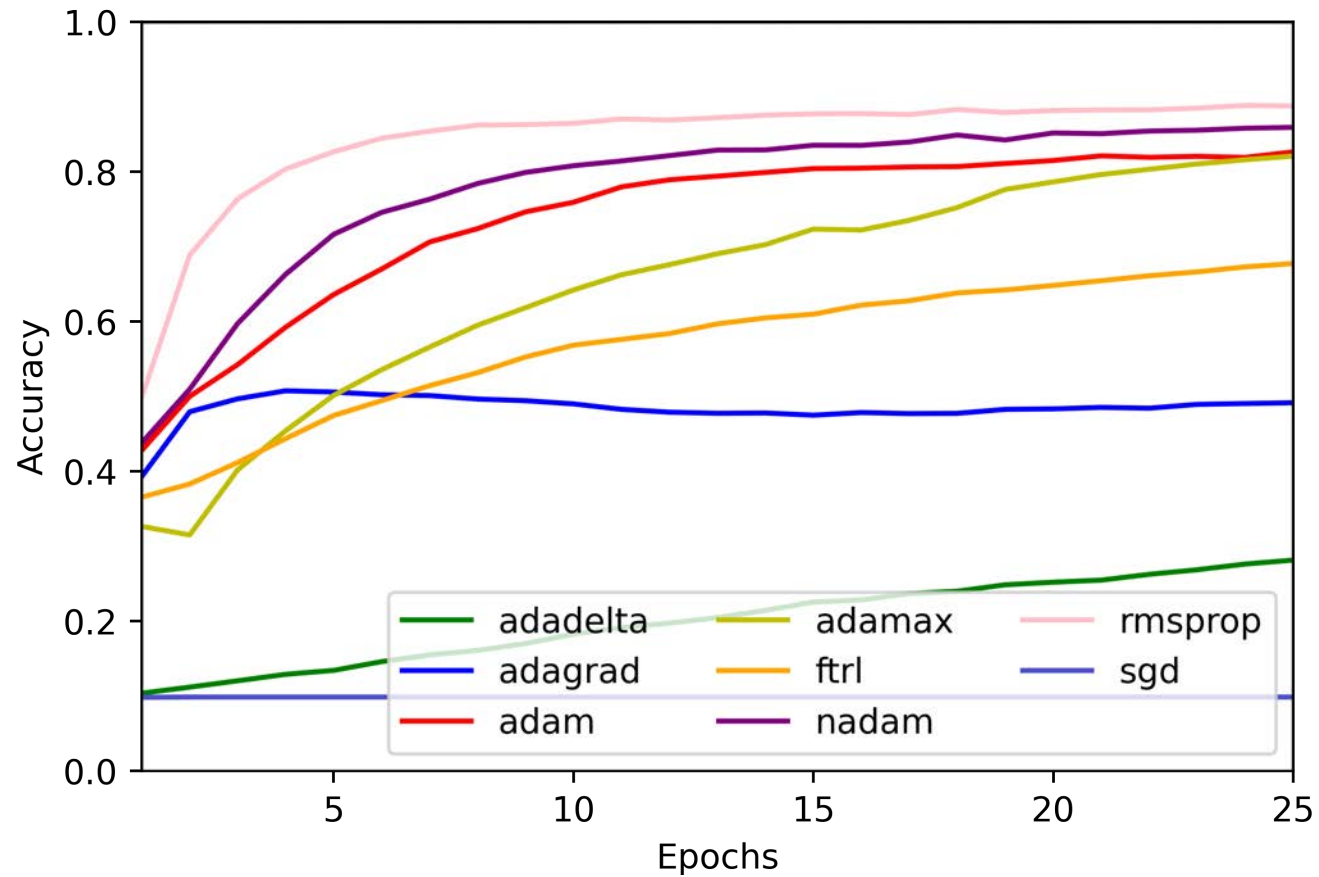
Optimizer

Optimizer

The optimizer is an algorithm that decided on how to adjust the weights and biases during backpropagation. This is done by minimizing the cost described by the loss function.

Sven Mayer

Optimizer

- Adadelta (2012 https://arxiv.org/abs/1212.5701)

- Adagrad (2011 https://www.jmlr.org/papers/volume12/duchi11a/duchi11a.pdf)

- Adam (v1: 2014 v9: 2017 https://arxiv.org/abs/1412.6980)

- Adamax (version of Adam https://arxiv.org/abs/1412.6980)

- FTRL (2013 https://research.google.com/pubs/archive/41159.pdf)

- Nadam (2015 http://cs229.stanford.edu/proj2015/054_report.pdf)

- Root Mean Square Propagation (RMSprop) (2012 http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf)

- Stochastic gradient descent (SGD)

Sven Mayer

Optimizer Test

With default settings

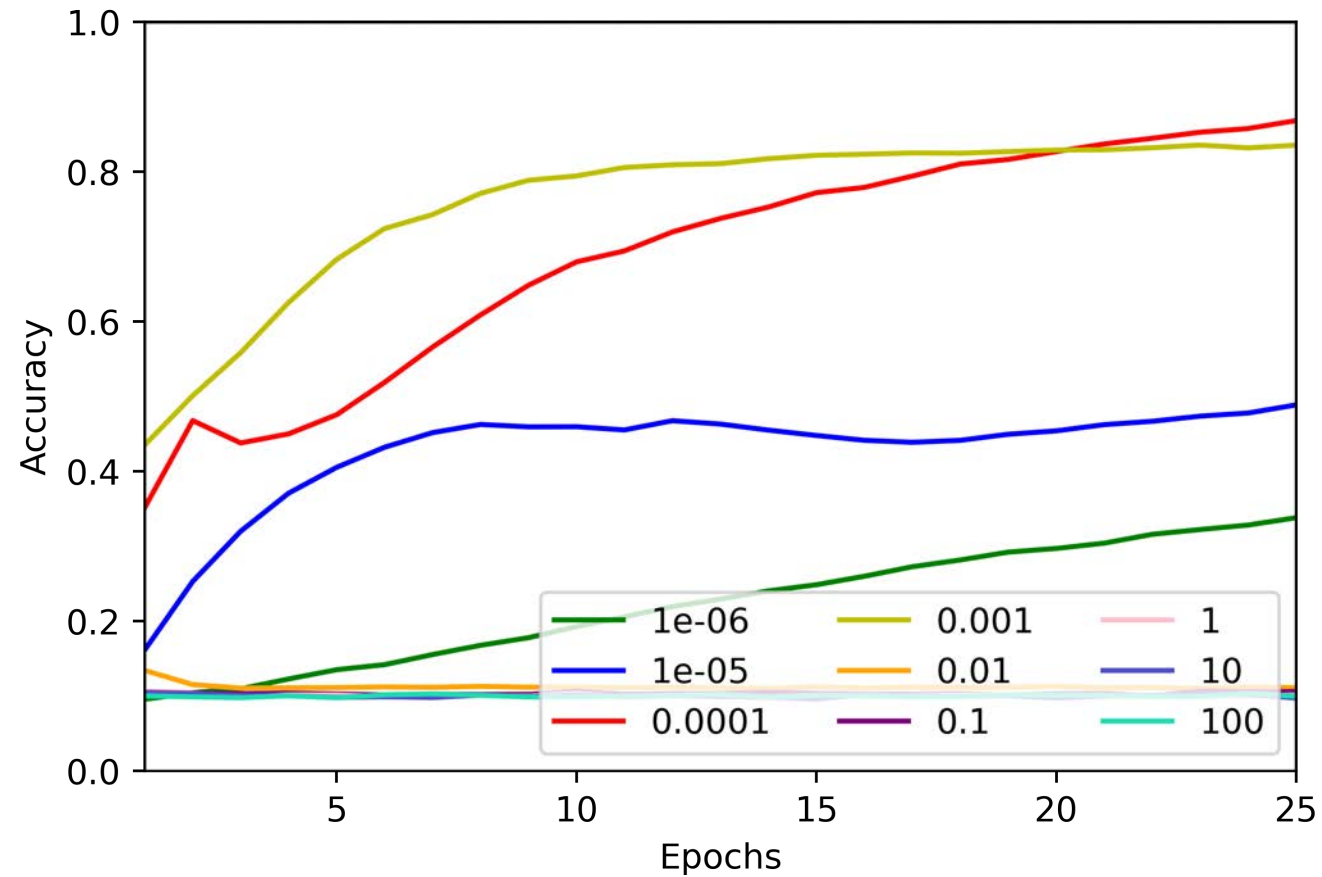2 dense layers with dropput tested on a subset of the MNIST dataset

Learning Rate

Learning Rate

The learning rate is the rate on how much the optimizer adjusts the weights and biases in each backpropagation step.

Sven Mayer

Learning Rate Test
Effect of the Initial Learning Rate Using Adam

2 dense layers with dropput tested on a subset of the MNIST dataset

Sven Mayer

Adaptive Learning Rate

- Learning rate decay over time

- Learning rate decay on no change (plateau)

- Use of "*momentum*" to accelerate training
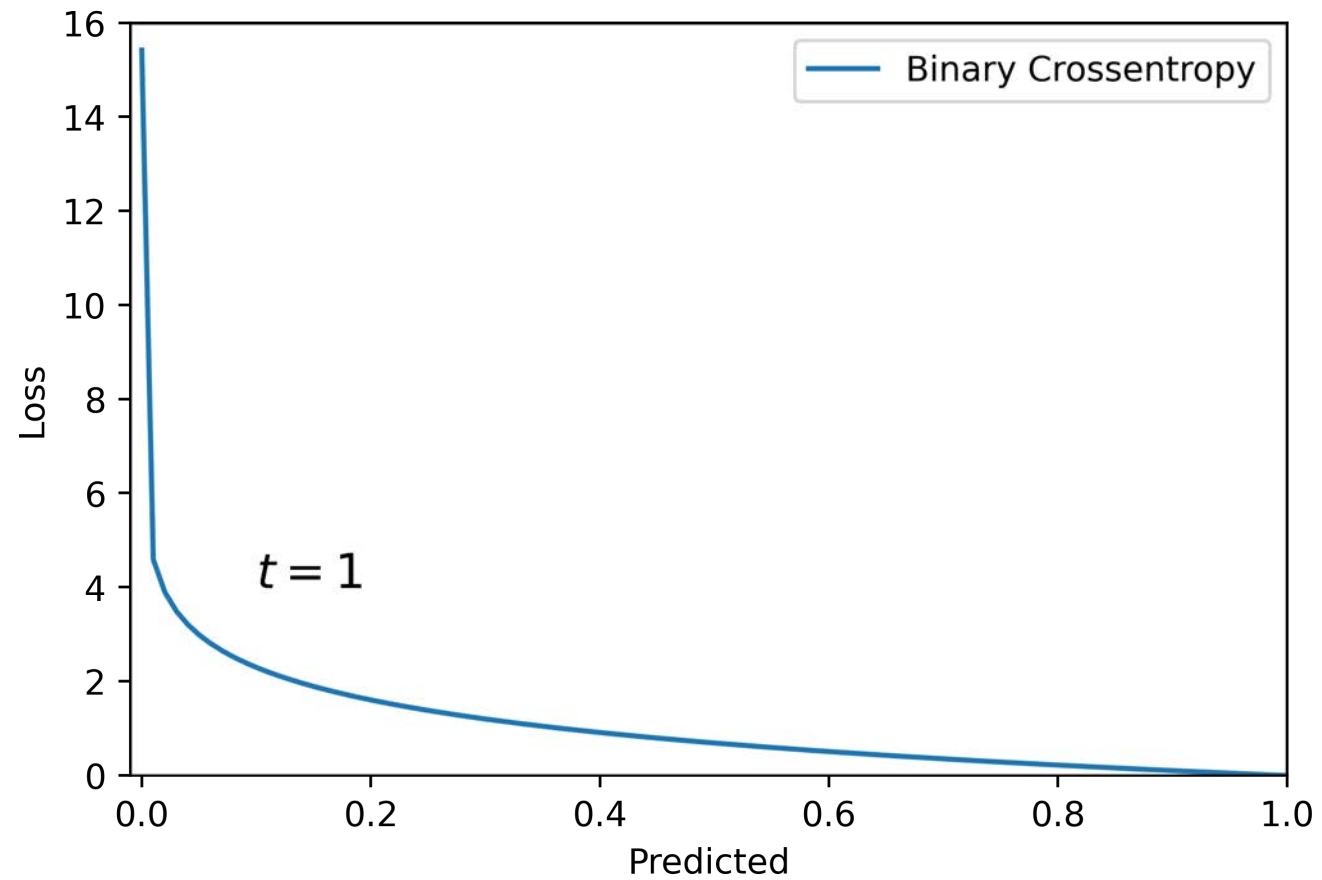
Loss Function

Loss Function

A loss function (cost function) is a function to determine how different the prediction is with respect to the ground truth. The optimizer's goal is to minimize the loss.

Sven Mayer

Loss Functions

- Discrete Data
  - Binary Cross Entropy
  - Categorical Cross Entropy
  - Hinge loss
- Continues Data
  - Mean Absolute Error (MAE) (also know as L1 Loss)
  - Mean Squared Error (MSE) (also know as L2 Loss)
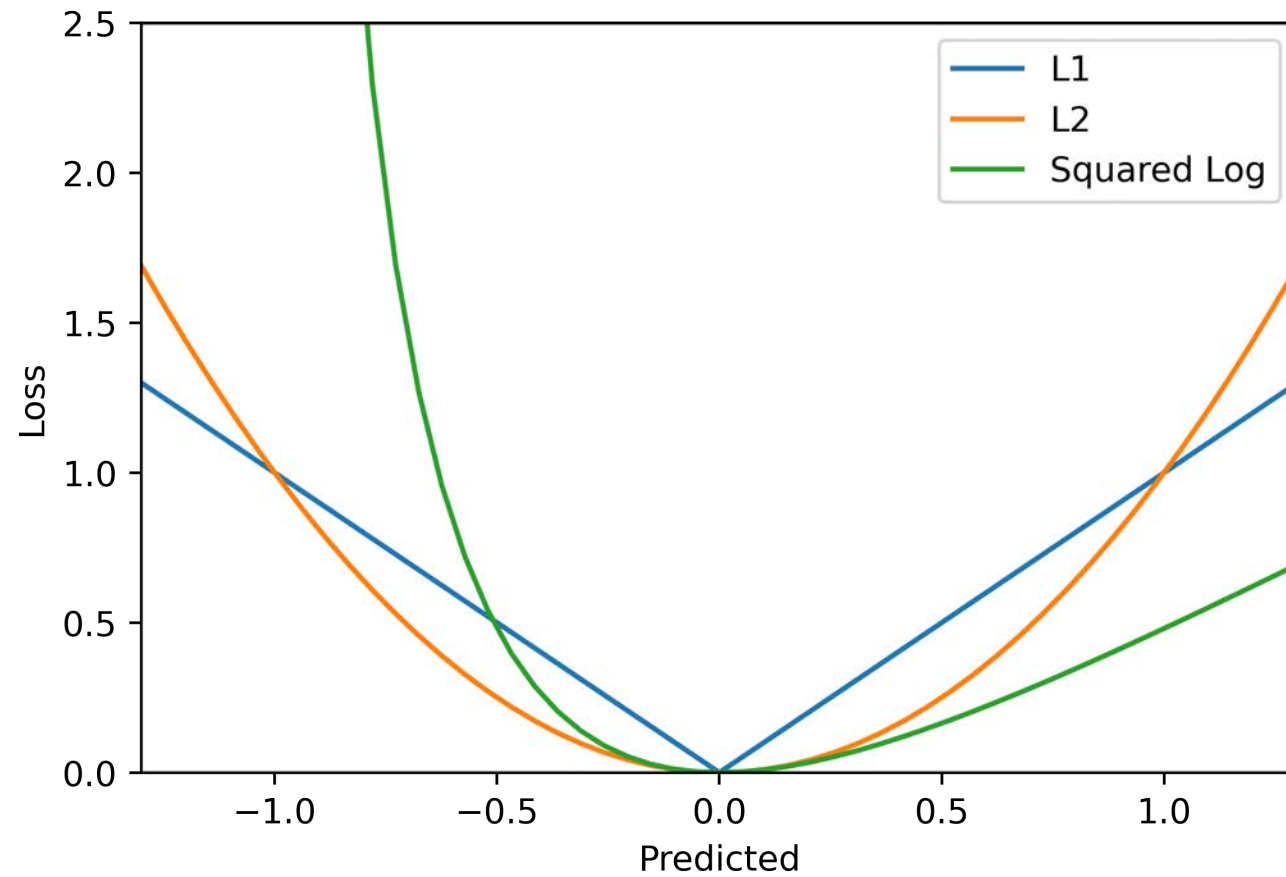 - Mean Squared Logarithmic Error

Sven Mayer

Loss Functions

Classificaion

Loss Functions

Regression

Sven Mayer

Loss Functions
Advanced Loss

- Pixel-wise loss

- Feature matching (FM), e.g. [1]

$$loss = \lambda_1 * MSE + \lambda_2 * FM$$

In the image domain, such approaches are used to counteract blurry images.

[1] Florian Schroff, Dmitry Kalenichenko, James Philbin; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815-823

Sven Mayer

Conclusion

Layers

- Optimizer
- Learning Rate
- Adaptive Learning Rate
- Loss Function

License

This file is licensed under the Creative Commons

Attribution-Share Alike 4.0 (CC BY-SA) license:

https://creativecommons.org/licenses/by-sa/4.0

Attribution: Sven Mayer