

Dropout

Overfitting

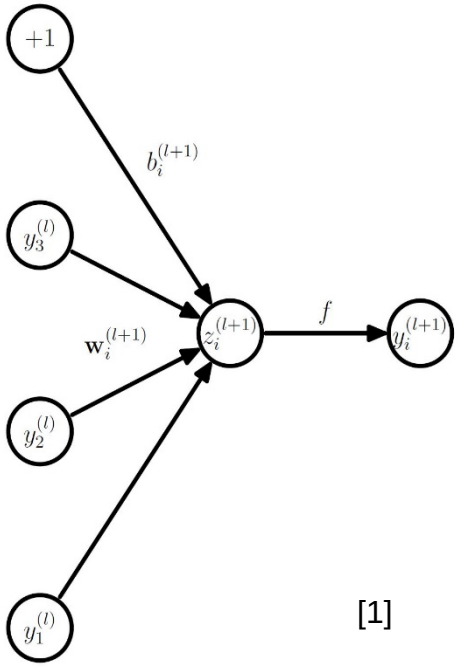
2 equation per neuron in an artificial neural network

- Calculation of activation
- Calculation of state using bias and weighted input from prior layer

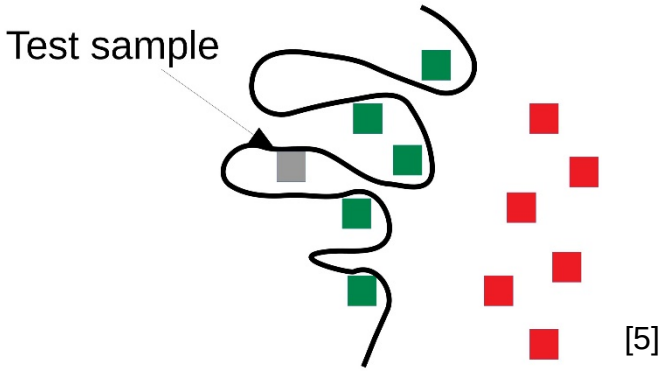
each neuron has its on input weights
each weight is a network parameter

adding a neuron to a layer means adding several parameters to the network

too many parameters can lead to overfitting

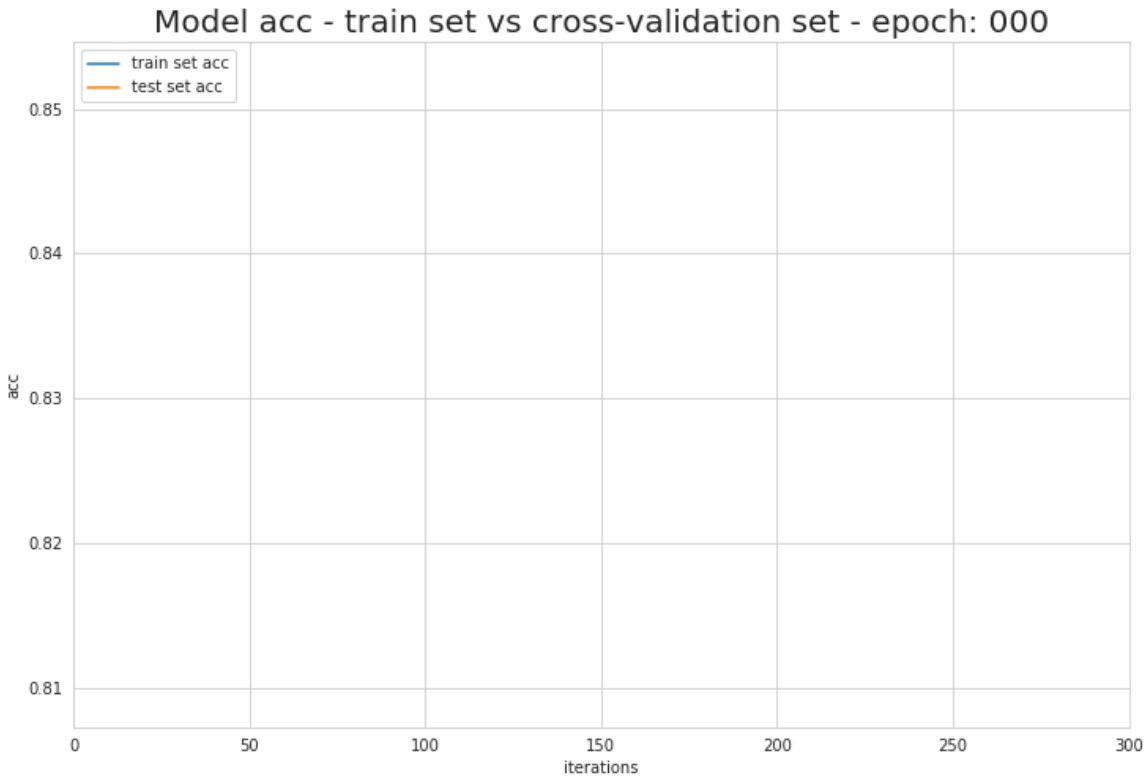


[1]



[5]

Overfitting



[4]

Overfitting

- network with too many parameters can adapt to noise in training data
- noise different to test data
- “memorization” of training data
- co-adaption between layers through chain rule:

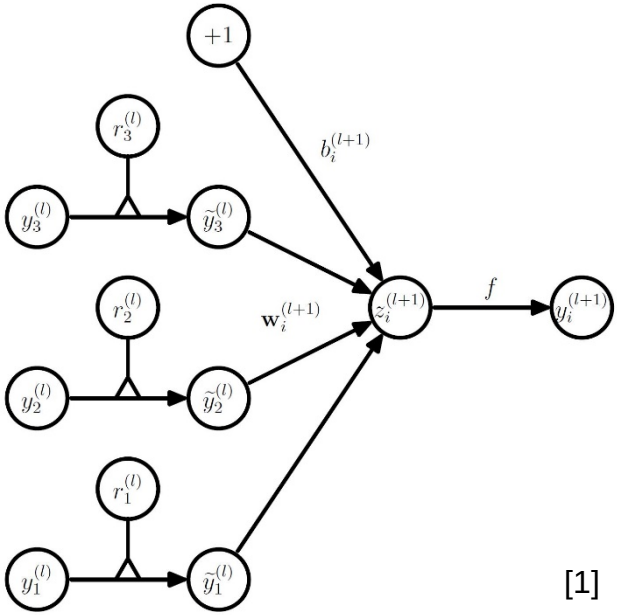
$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b)$$

[5]

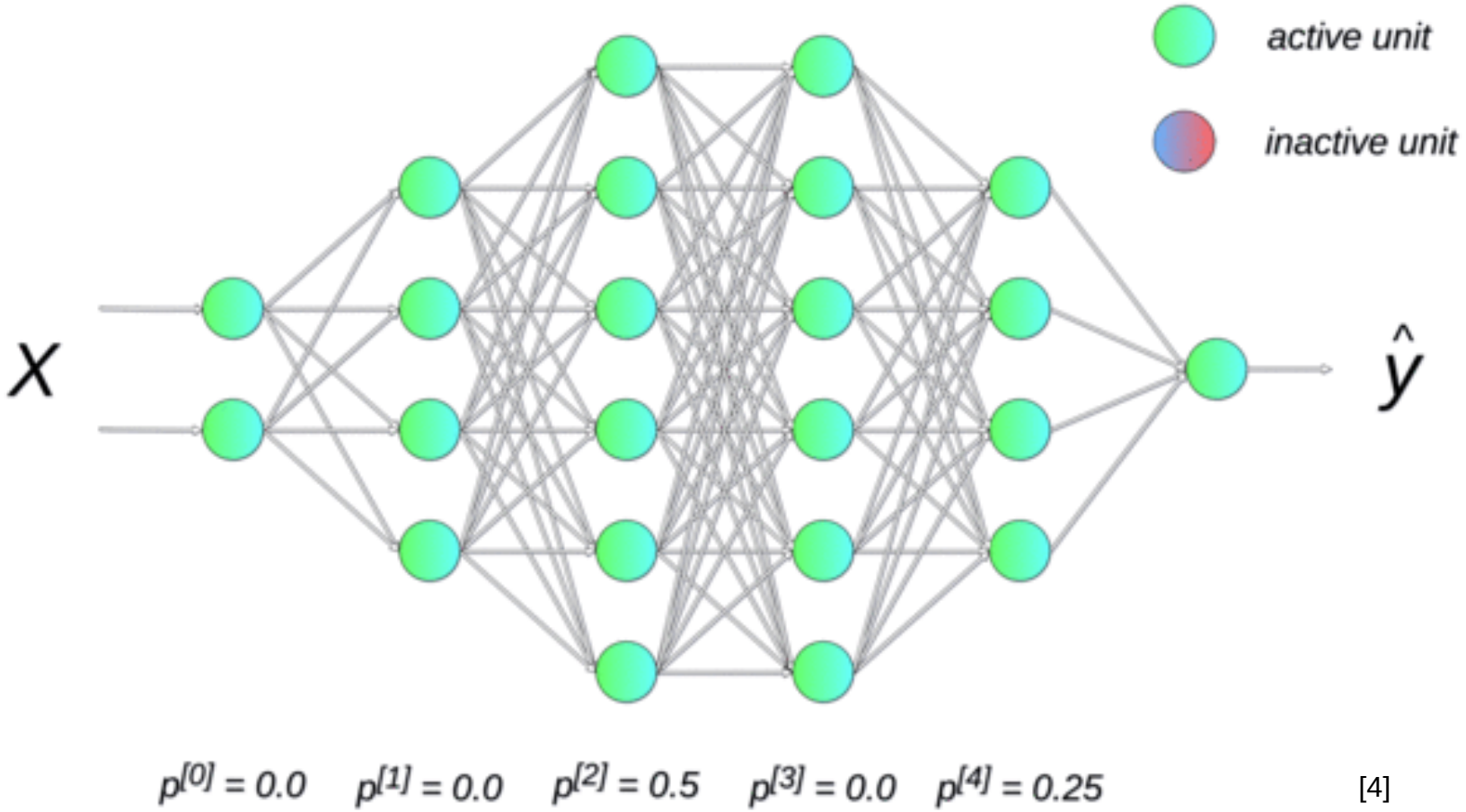
- neuron updated according to the behavior of the other neurons
- neurons may change so they correct the mistakes of other neurons

Dropout

- solution:
  randomly dropping neurons of the net to train with thinner networks and to stop co-adaption
- adding a layer that drops a neuron with the probability p



[1]

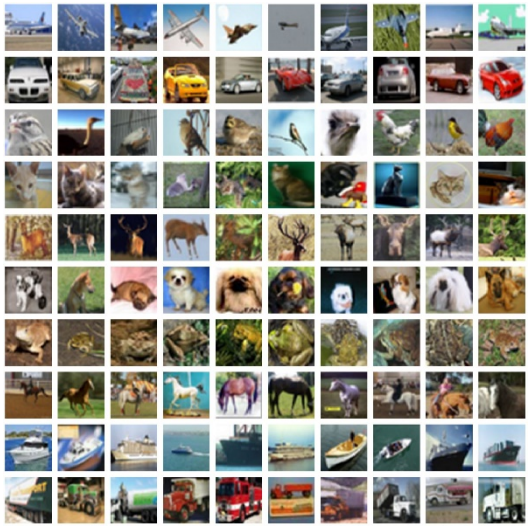


[4]

Experiments

setup

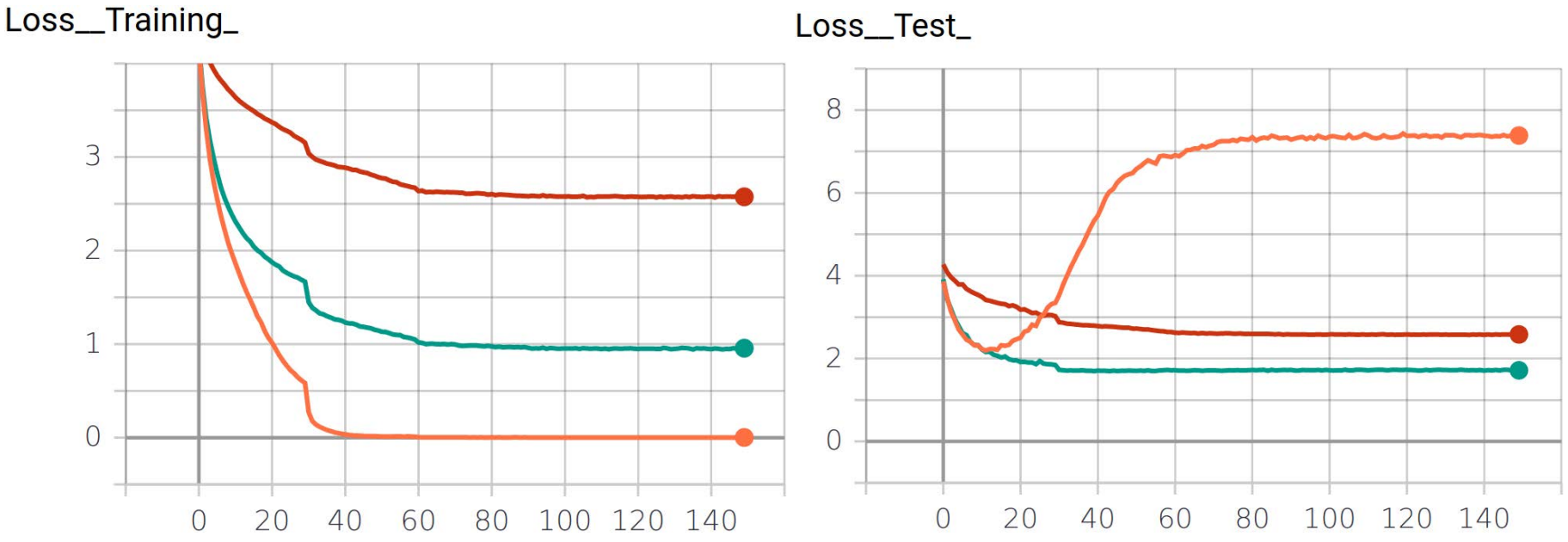
- Modell: Alexnet
  more than 650000 neurons and over 62 million parameters
  won 2012 ImageNet LSVRC-2012 competition with 15.3% error rate
  (vs 26.2 % second place) [3]
- data:
  CIFAR 10/100
  60000 images
  3-channel color and 32x32 pixels
  10 different classes with 6000 images each
  100 different classes with 600 images each[1]



[1]

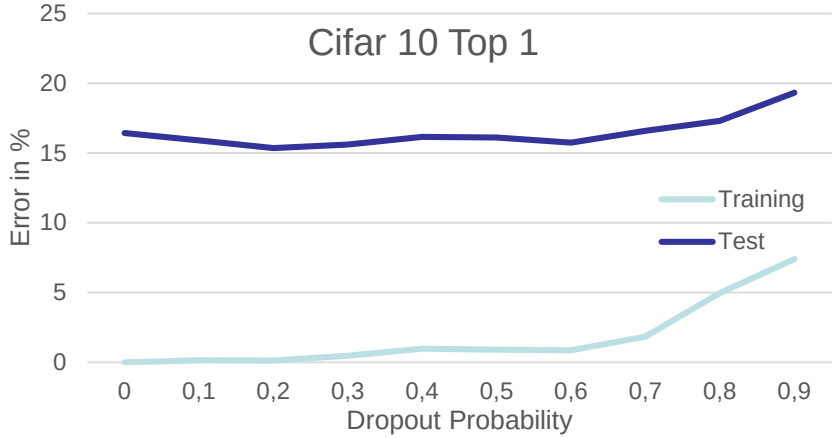
Loss function comparison

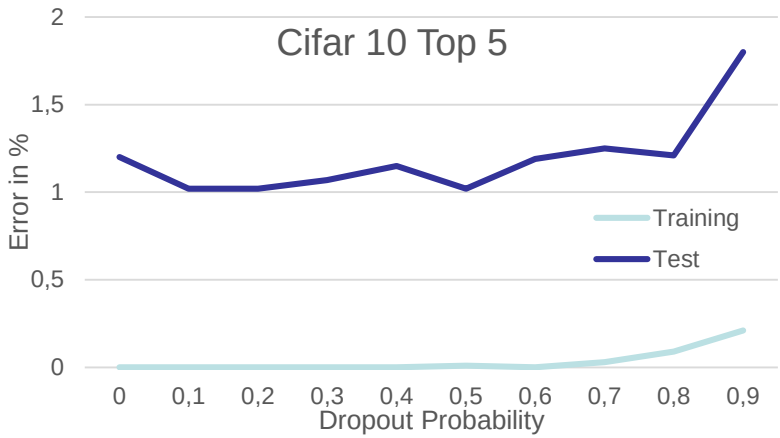
- Loss function with and without dropout

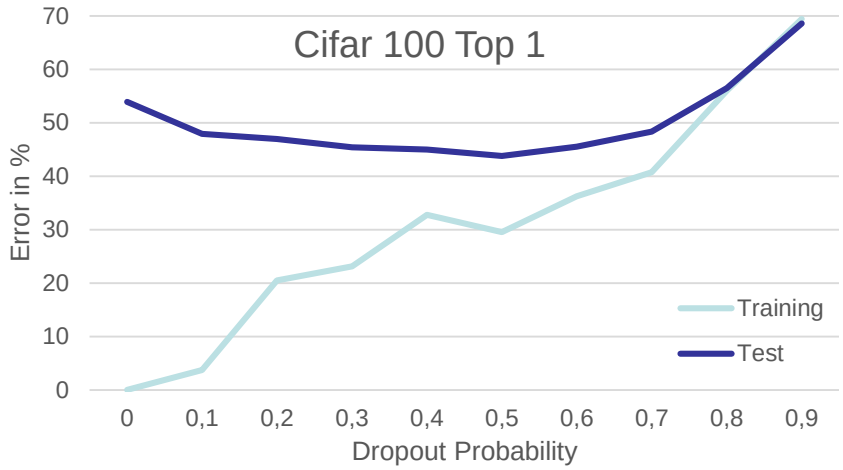


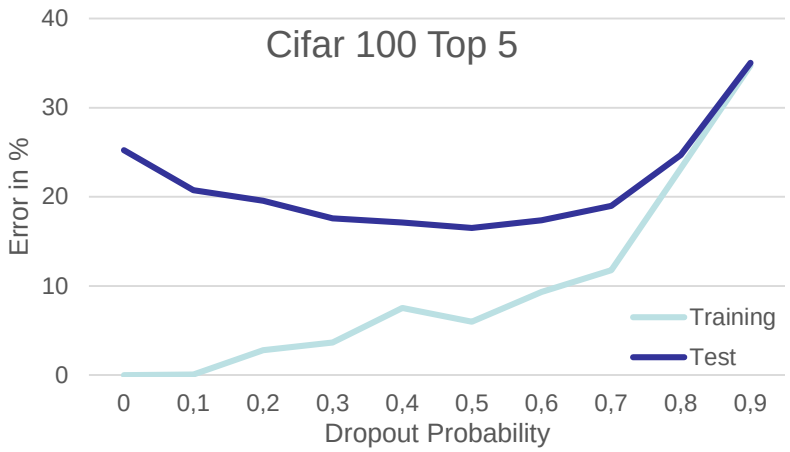
Effect of dropout probability

correlation between error rate and dropout probability



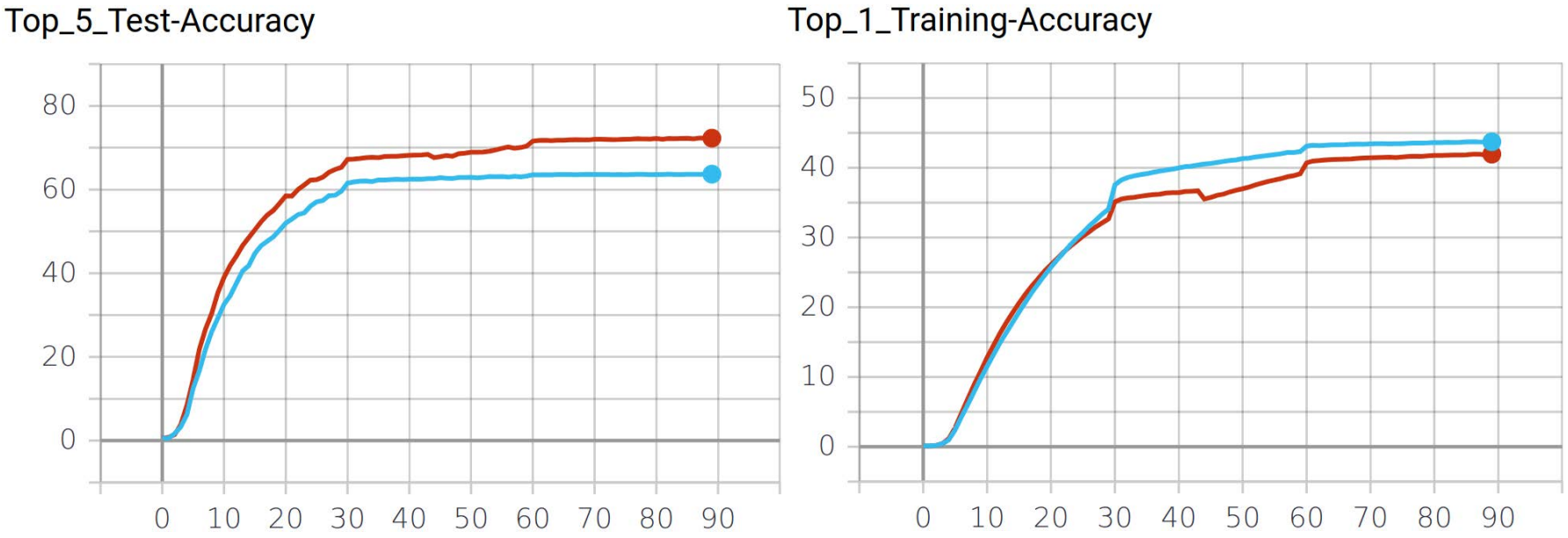






Effect of dropout probability

- Effect of dropout in correlation to data size



0.5 Dropout Probability

No Dropout

Conclusion

- dropout can prevent overfitting
- dropout can decrease error rate by using the principle of combining several trained thinner networks
- dropout adds noise to the input signal
- dropout rate has a sweet spot
- dropout increases learning time
- the effect of dropout on error rate is dependent on the available size of training data

Sources

[1] „Dropout: A Simple Way to Prevent Neural Networks from Overfitting" Srivastava, Hinton, Krizhevsky et. al. (2014): https://www.cs.toronto.edu/~hinton/absps/JMLRdropout.pdf

[2] GitHub Project: https://github.com/iceboy910447/ALEXNET-CIFAR10

[3] „A Walk- through of Alexnet", Hao Gao (2017): https://medium.com/@smallfishbigsea/a-walk-through-of-alexnet-6cbd137a5637

[4] „Preventing Deep Neural Network from Overfitting", Piotr Skalski (2018) : https://towardsdatascience.com/preventing-deep-neural-network-from-overfitting-953458db800a

[5] „Deep Learning – An Introduction Beyond Buzzwords" Janis Keuper (2019): https://elearning.hs-offenburg.de/moodle/pluginfile.php/429832/mod_resource/content/1/1_Deep_Learning_Intro.pdf