Comparison analysis of dropout methods for regression task in deep learning

Defence presentation

Zafarzhon Irismetov
Scientific supervisor: Dr. Sc. Comp. Evalds Urtans

Topicality of the thesis topic

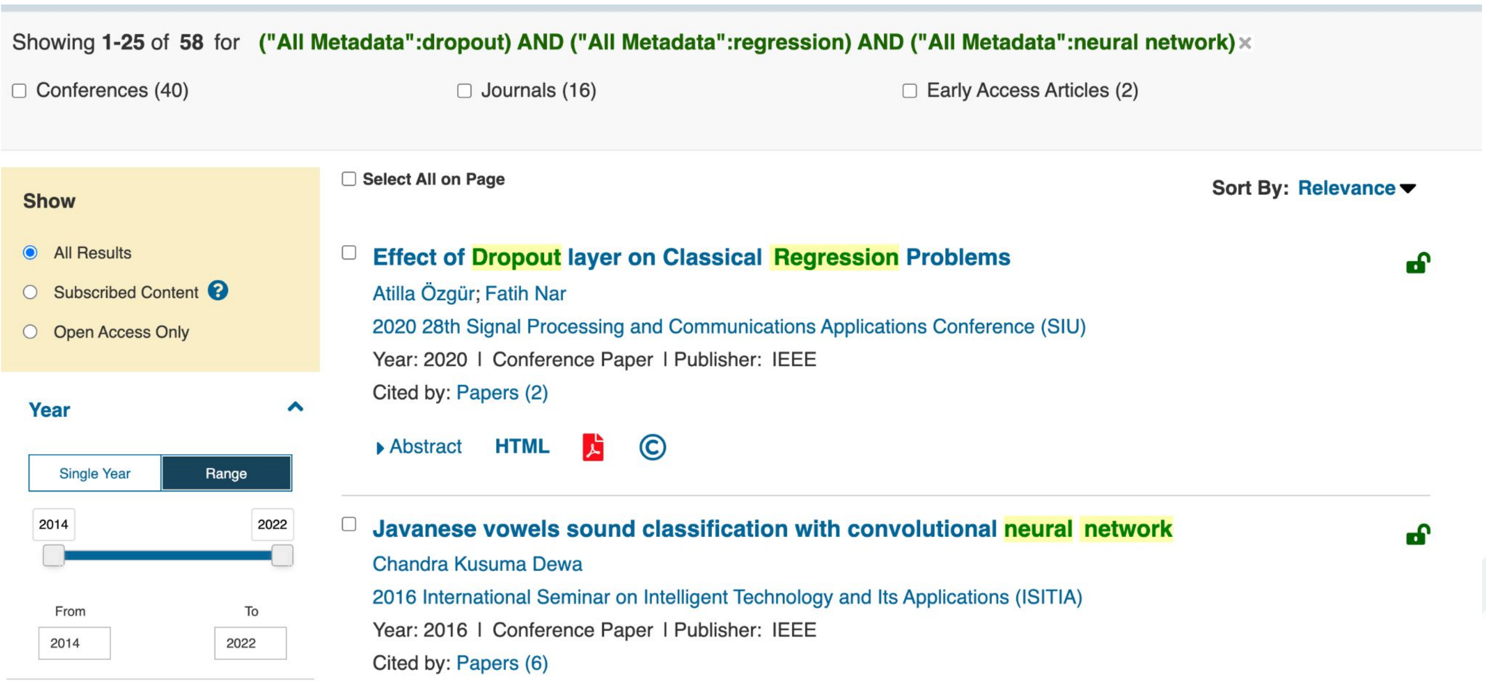
The dropout method prevents overfitting in deep neural network. Not much researches have been made regarding dropout effect on regression tasks in deep models.

During litrachure search only one paper with similar research was found. The information there was not sufficient, because first “old” version of dropout method was tested.

Most of the researches that test the effect of dropout methods are made on classification, image and speech recognition tasks.

Papers: https://www.researchgate.net/publication/344274687_Effect_of_Dropout_Layer_on_Classical_Regression_Problems (similar topic)
https://arxiv.org/pdf/2010.05244.pdf(Advanced dropout)

Topicality of the thesis topic literature surfing



https://ieeexplore-ieee-org.resursi.rtu.lv/search/searchresult.jsp?action=search&newsearch=true&matchBoolean=true&queryText=(%22All%20Metadata%22:dropout)%20AND%20(%22All%20Metadata%22:regression)%20AND%20(%22All%20Metadata%22:neural%20network)

The goal of the thesis

- Type 1

- The purpose of this work is to perform a comparison analysis of Simple, Drop-Connect, Gaussian and Advanced dropout methods in a regression tasks with four datasets in deep learning model. (Boston houses, California housing price, Weather is Szeged, BNG).

Hypothesis and tasks

Hypothesis:

1. The dropout functions prevent overfitting in regression tasks.
2. The modern ‘Advanced’ dropout function (Xie, et al., 2021) reduce overfitting better than its predecessors.

Tasks:

- To study the background information of neural network, regression, overfitting, and dropout function
- Analyze the open source framework ‘PyTorch’ to build a deep neural model for this experiment.
- Develop the methodology of the experiment.
- Compare and analyze the results of four dropout functions in different datasets

Overfitting problem

Overfitting is a problem in machine learning when the model learned patterns specific to the training data, which are irrelevant to other data. In other words, the model is unable to produce accurate predictions for real data.

The overfitting may appear due to these factors:

- Data for deep models may contain many errors.
- Model complexity is high

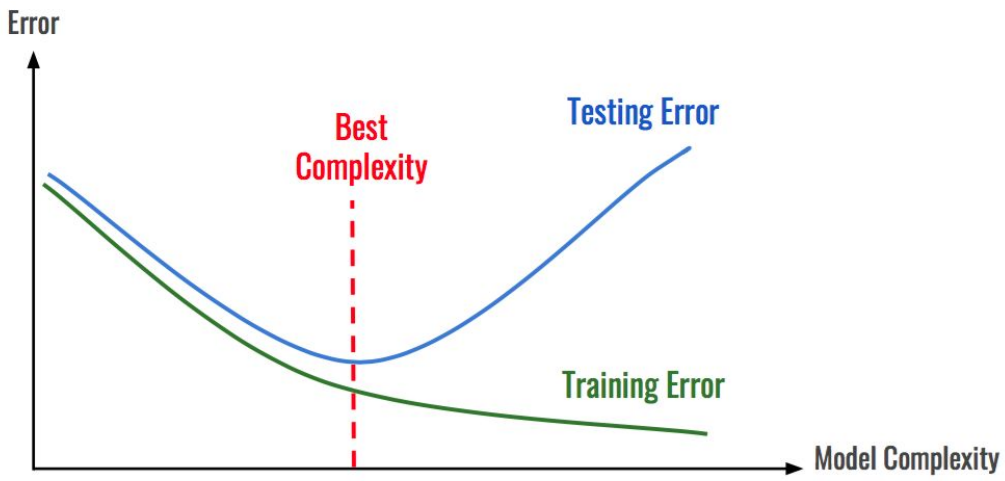
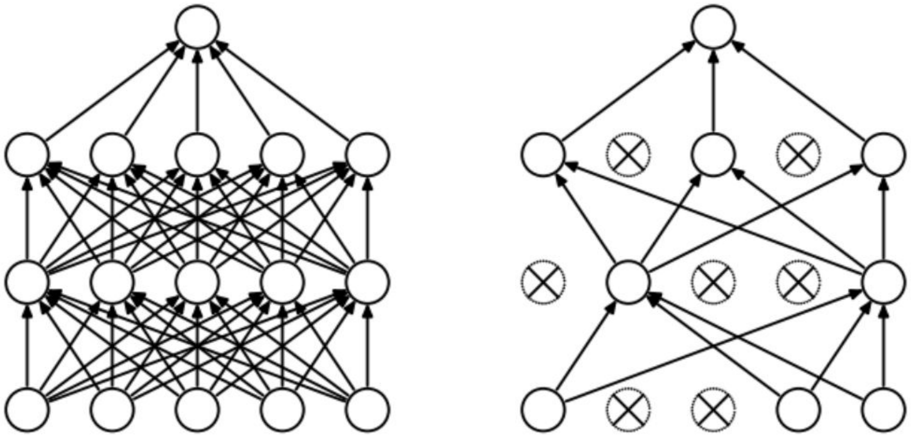


Image source: https://stackoverflow.com/questions/59856614/overfitting-and-data-leakage-in-tensorflow-keras-neural-network

Dropout method

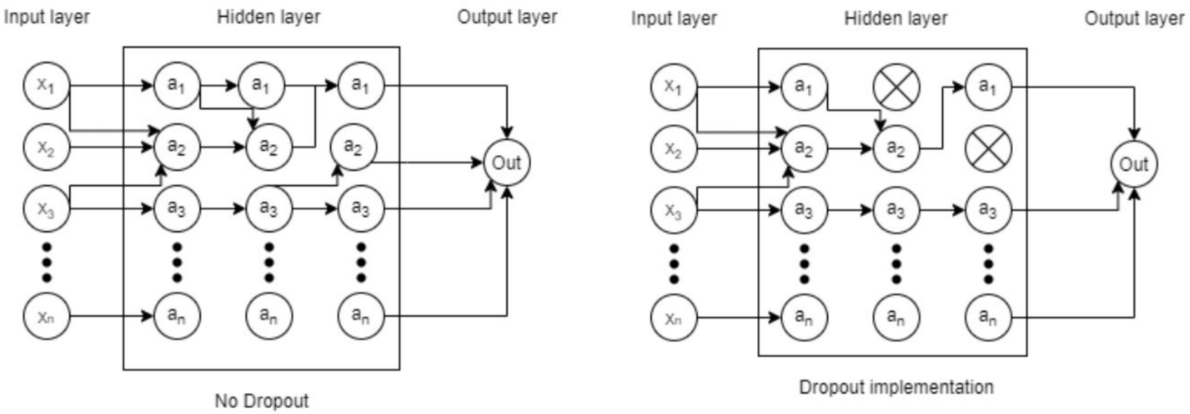
One method of preventing overfitting is the dropout. It was invented in 2012 and since then, researchers have continued to upgrade it. This method is often used in models for picture and voice recognition, because it showed it's practically usefulness for long time.



(a) Standard Neural Net (b) After applying dropout.

Image source: https://medium.com/@amarbudhiraja/https-medium-com-amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5

Simple dropout



$y = P * a\,(\,\omega \times X + b\,)$

y - output

P - Bernoulli(P) (probability of drop)

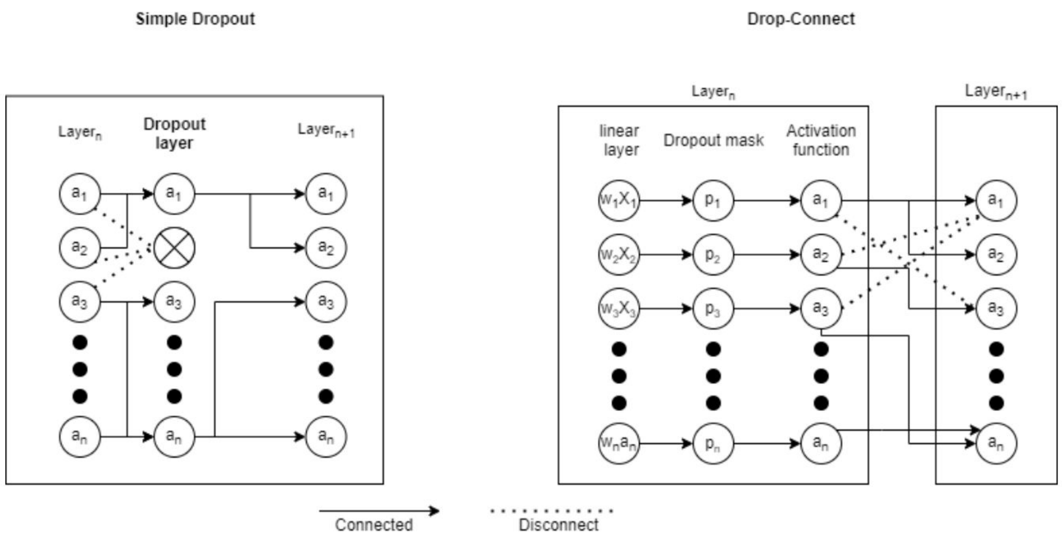
a - activation function

w - weight

X - input

b - bias

Drop-Connect



$y = a(\dot{\omega} \times X + b)$

$\dot{\omega} = \omega \times P$

y - output

P - Bernoulli(P) (probability of drop)

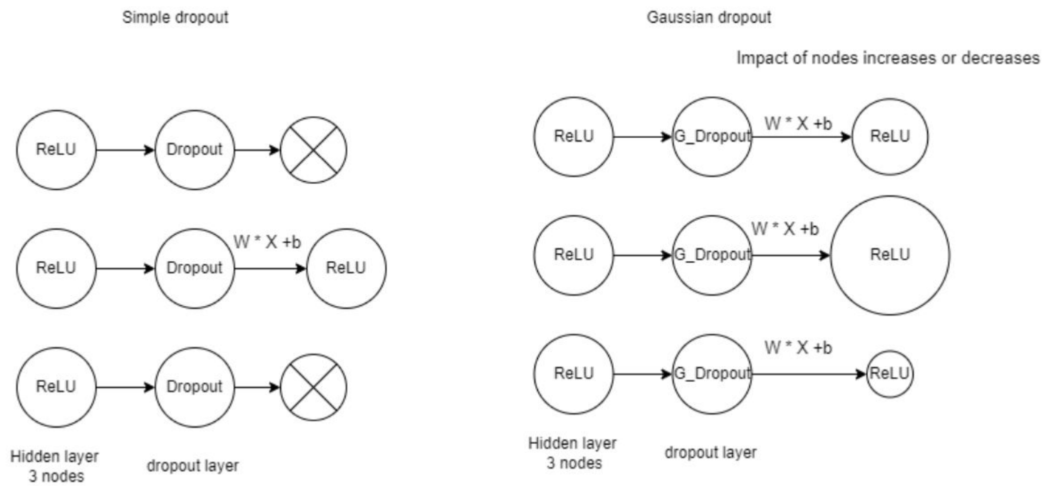
a - activation function

ω - weight

X - input

b - bias

Gaussian Dropout



$$y = a(\,\omega \times X + b\,) * Mg$$

$$M \sim \mathcal{N}(1,\, p/(1-p)\,)$$

y - output

M - dropout mask

a - activation function

ω - weight

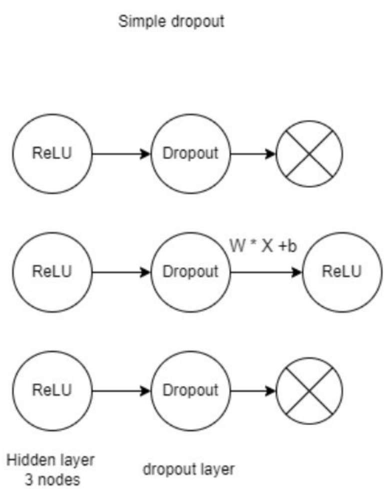
X - input

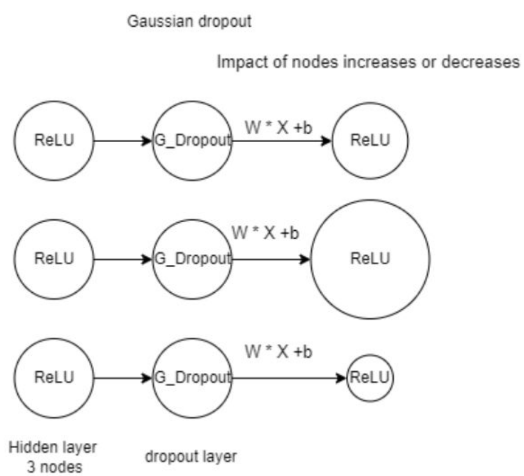
b - bias

Advanced dropout

Same behavior as gaussian dropout, but different formula, with 2 learnable parameters (mu, sigma)

$y = a(\omega \times X + b) * M\text{j}$





y - output

M - dropout mask

a - activation function

ω - weight

X - input

b - bias

$$m_j^{(l)} = k(r_j^{(l)}) = \text{Sigmoid}(r_j^{(l)}) = \frac{1}{1 + e^{-r_j^{(l)}}},$$

$$r_j^{(l)} \sim \mathcal{N}(\mu_l, \sigma_l^2),$$

Model architecture, without dropout

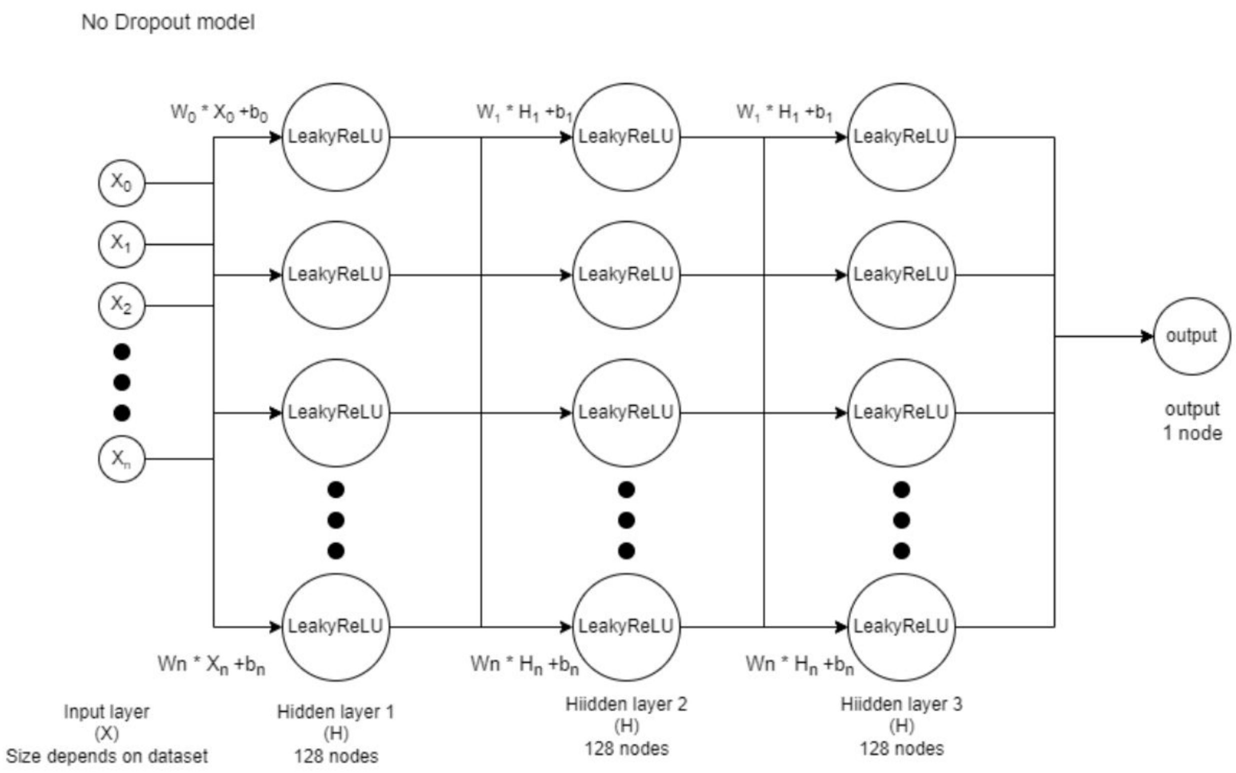


Figure 10 Experiment model without dropout

Simple dropout model

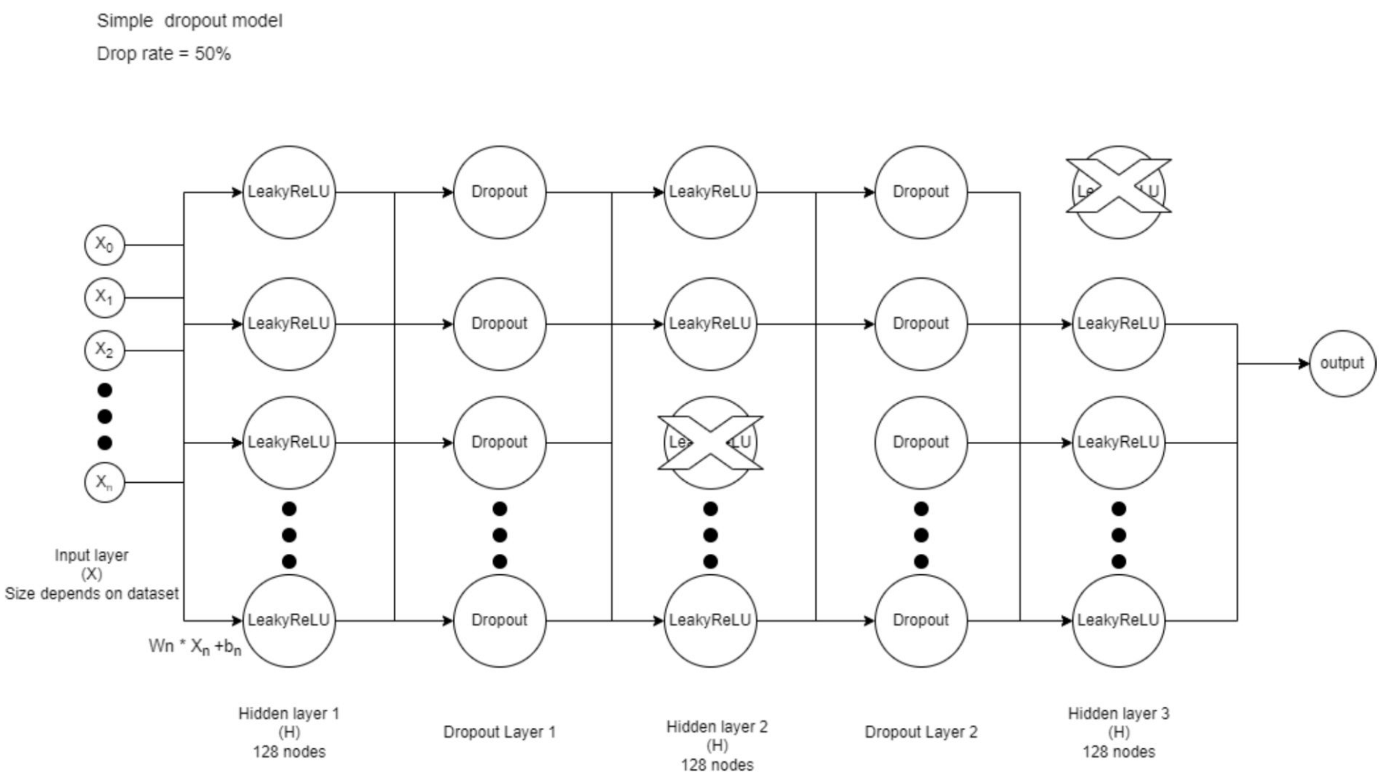


Figure 11 Simple Dropout functions placement

Drop-connect model

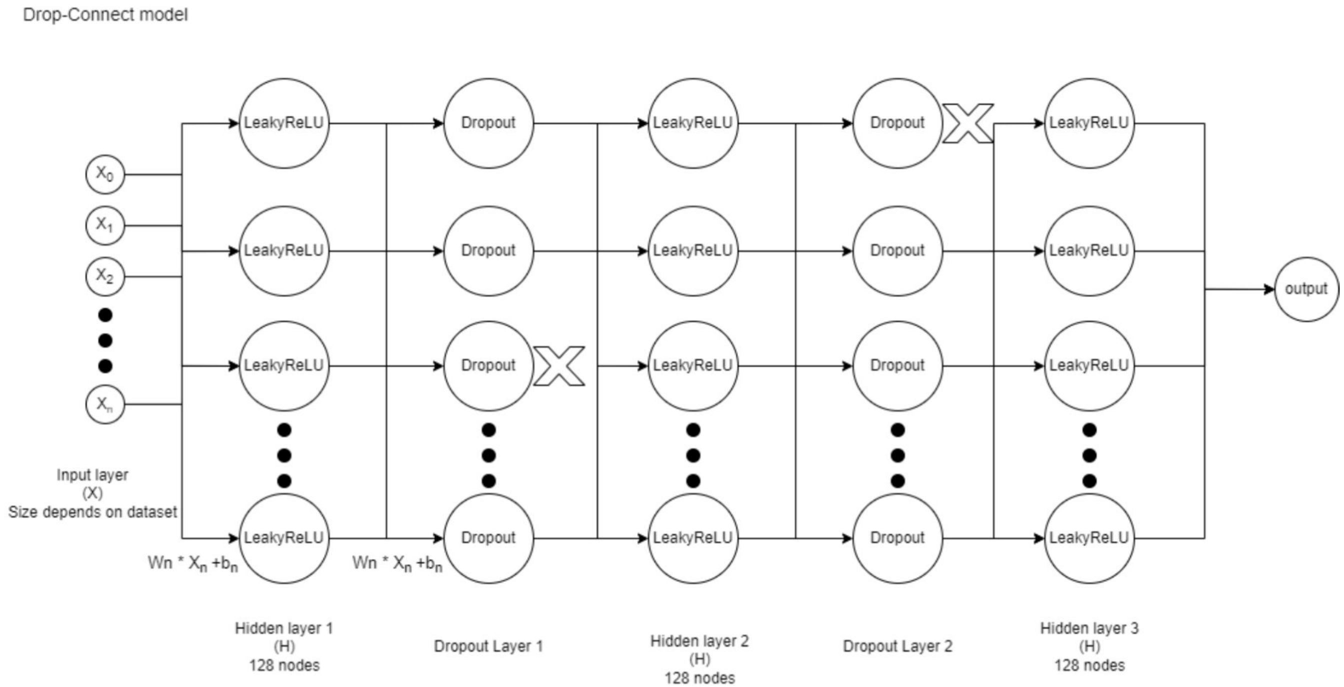


Figure 13 Drop-Connect functions placement

Gaussian and Advanced dropout models

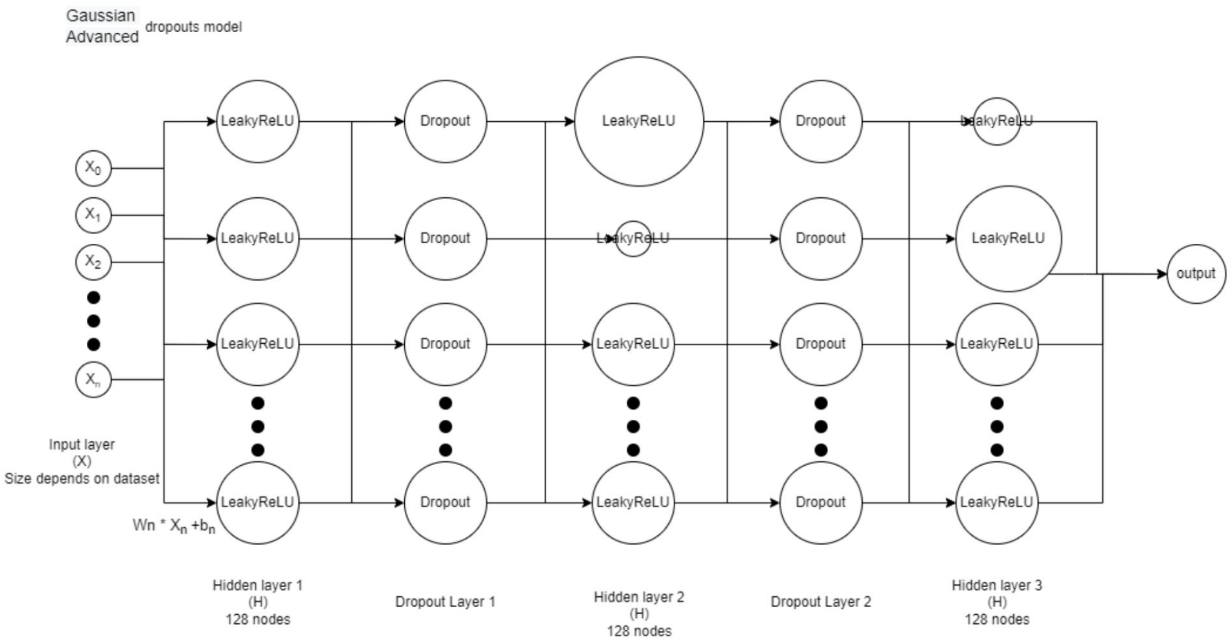


Figure 12 Gaussian and Advanced Dropout functions placement

Parameters grid-search method

Dataset name	Learning rate	Epochs	Amount of data to test (%)	Batch size
Boston houses	0.001	2500	40%	8
California houses	0.001	2500	40%	16
Weather in Szeged	0.001	2500	40%	32
BNG	0.001	2500	40%	32

Dropout method	Drop rate
Simple dropout	50%
Gaussian dropout	50%
Drop-Connect	50%
Advanced dropout	parameters learning rate: 0.0001

Table 2. Dataset information

Name	Total samples	Features	Link
Boston houses	506	14	http://lib.stat.cmu.edu/datasets/boston
California houses	20,640	9	https://www.kaggle.com/camnugent/california-housing-prices
Weather in Szeged	96,540	4	https://www.kaggle.com/budincsevity/szeged-weather
BNG	1,000,000	18	https://www.openml.org/d/1191

Results, boston houses

Boston houses contains data regarding the real estate situation in Boston in 1978.

The best accuracy with standard model was 75.86%. Advanced and Simple dropout increased max accuracy by 15% and reached 90.51%. During advanced dropout testing, the highest accuracy was reached in 300 epochs.

Table 3 Boston houses dataset, results overview

Dropout method	Dropout probability (p)	Best Loss in test (Smaller the better)	Best R^2 score in test (Higher the better)	P – value (With regard to the NoDropout model and Dropout models)
No Dropout	-	0.2955	75.86%	-
Simple Dropout	0.5	0.2218	90.26%	0.022
Drop-Connect	0.5	0.365	66.88%	0.0
Gaussian Dropout	0.5	0.2454	88.03%	$1.8 * 10^{-184}$
Advanced Dropout	-	**0.2096**	**90.51%**	$1.3 * 10^{-91}$

Boston houses, Loss/R2 plot

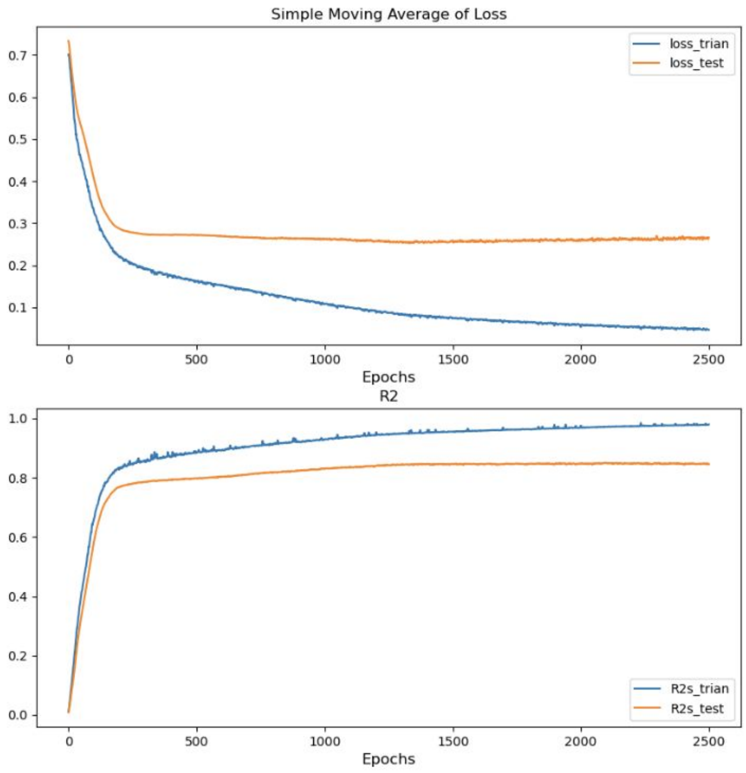


Figure 14 Boston dataset, no dropout

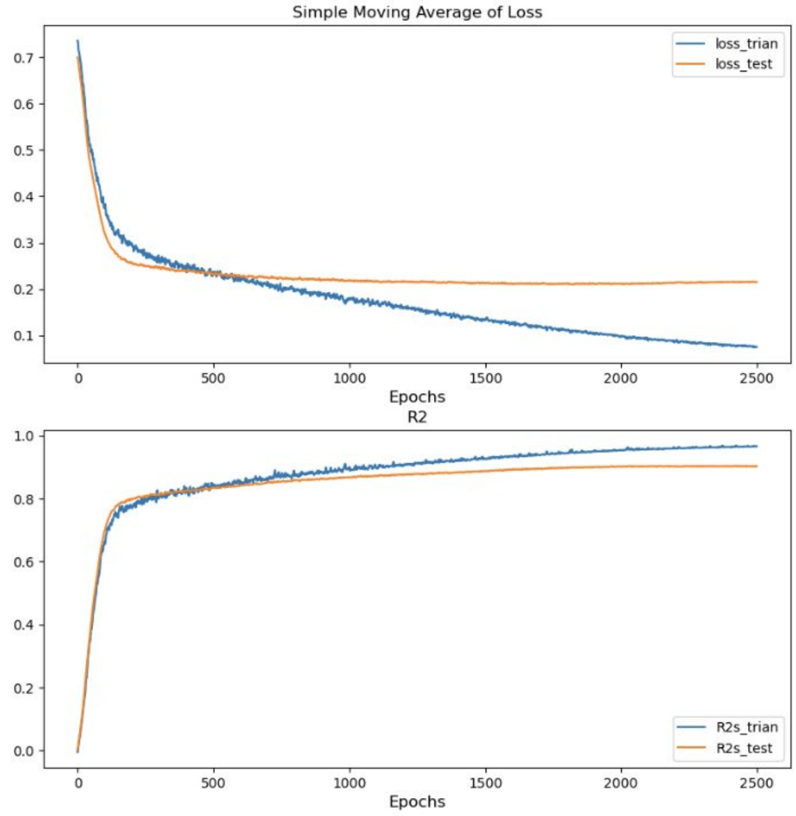


Figure 18 Boston dataset with Advanced dropout

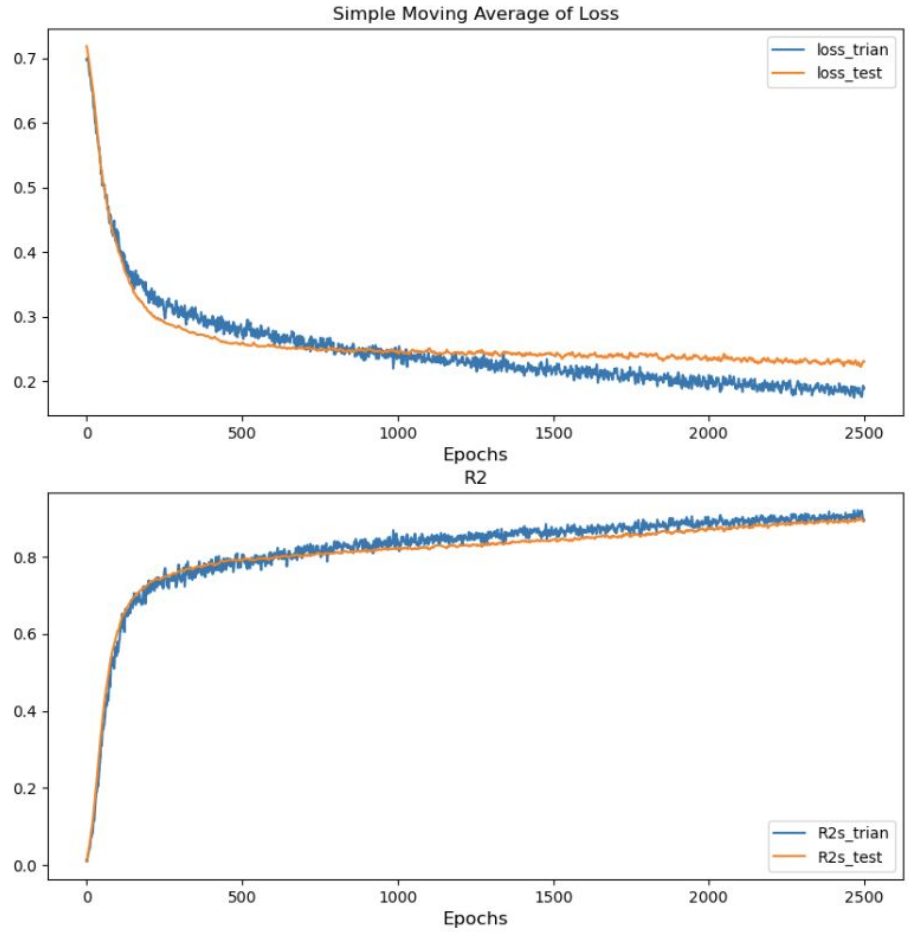


Figure 15 Boston dataset with Simple Dropout

Results, california houses

California houses, it contains the data regarding real estate situation in California in 1990.

Simple dropout performed the best in this case, best accuracy was 80.06%, very close to standard model's result.

The Drop-Connect performed very bad, the R2 score results have a large variance.

Advanced dropout reached the max accuracy during 500 epochs.

Table 4 California Housing dataset, overview of results

Dropout method	Dropout probability (p)	Best Loss in test (Smaller the better)	Best R^2 score in test (Higher the better)	P – value (With regard to NoDropout with Dropout models)
No Dropout	-	**0.2861**	**80.35%**	-
Simple Dropout	0.5	0.2894	*80.06%*	$3.9 * 10^{-21}$
Drop-Connect	0.5	0.4705	56.9%	0.0
Gaussian Dropout	0.5	0.3333	75.59%	0.0
Advanced Dropout	-	0.2938	79.37%	$6 * 10^{-43}$

California housing, Loss/R2 plot

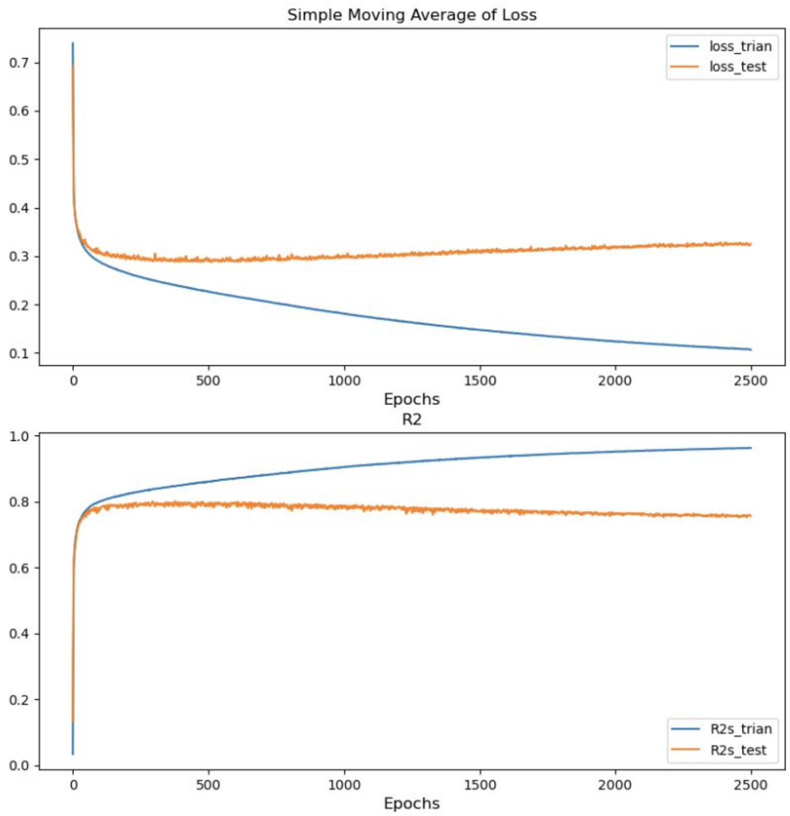


Figure 19 California Housing - overfitting in ANN model

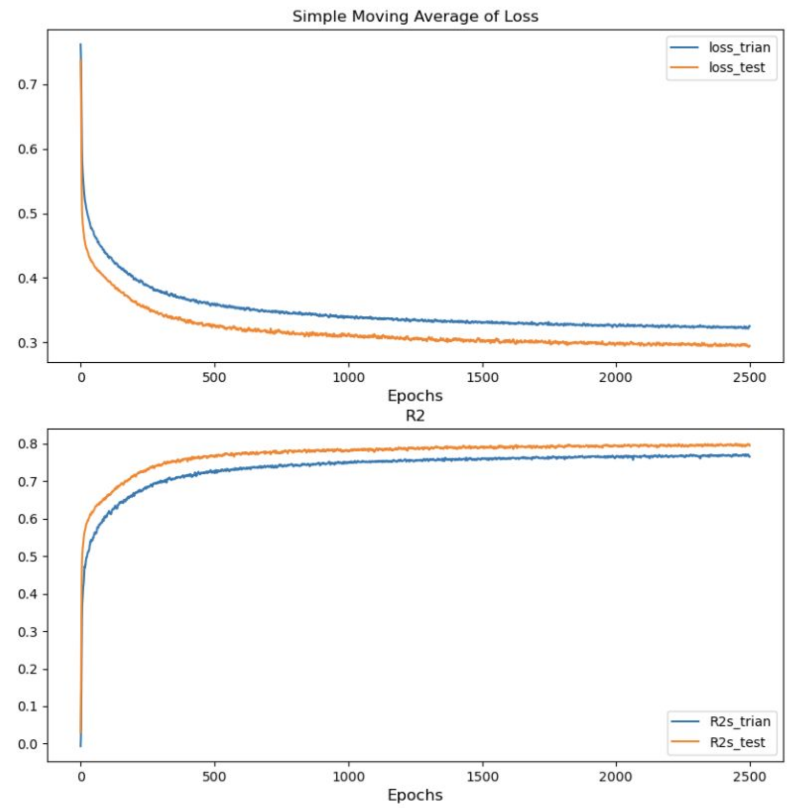


Figure 20 California Housing dataset, Simple dropout

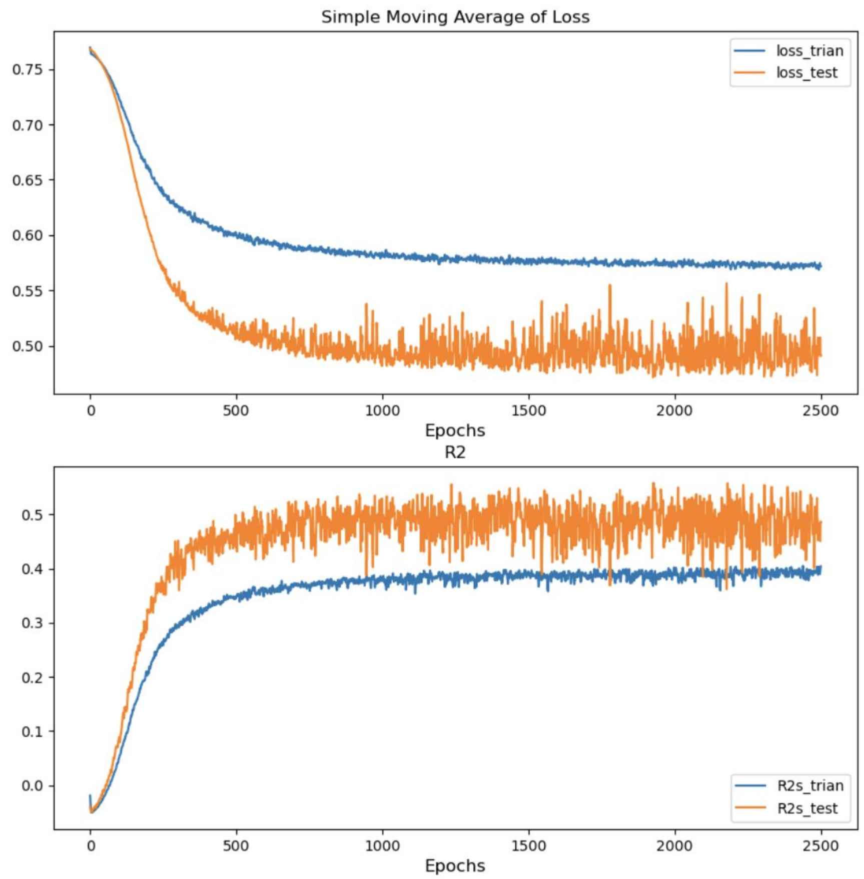


Figure 21 California Housing dataset with Drop-Connect

Advanced dropout, california

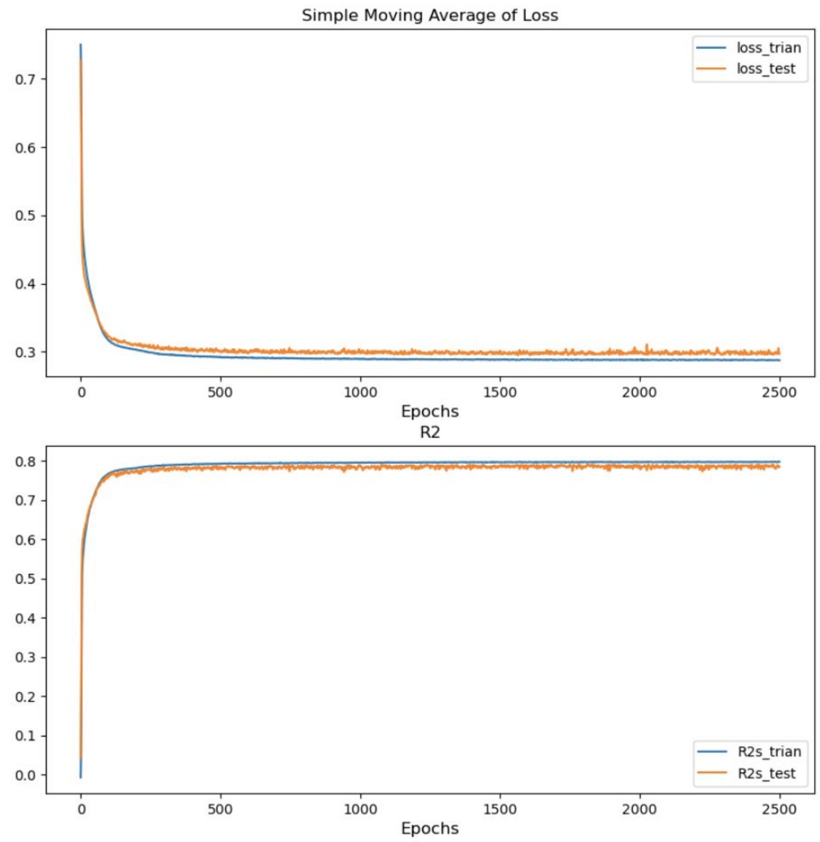


Figure 23 Advanced Dropout, California Housing dataset

Results, weather in szeged

Weather in Szeged is the weather storage dataset, which contains weather data from 2006 to 2016 in the Czech Republic area. Predict the temperature by “Wind Speed (km/h)”, “Humidity”, “Wind Bearing (degrees)” factors.

Simple dropout is the best solution with best accuracy 44.36%.

Gaussian dropout had many accuracy drops, untrustable results.

Advanced dropout reached its maximum during 500 epochs.

Table 5. Dropout functions results review, weather dataset

Dropout method	Dropout probability (p)	Best Loss in test (Smaller the better)	Best R^2 score in test (Higher the better)	P – value (With regard to NoDropout with Dropout models)
No Dropout	-	**0.58**	**44.96%**	-
Simple Dropout	0.5	0.5981	*44.36%*	$3.8 * 10^{-109}$
Drop-Connect	0.5	0.6068	41.83%	$1.6 * 10^{-280}$
Gaussian Dropout	0.5	0.6095	42.32%	0.0
Advanced Dropout	-	0.5952	43.71%	$8.5 * 10^{-260}$

Weather dataset, Loss/R2 plot

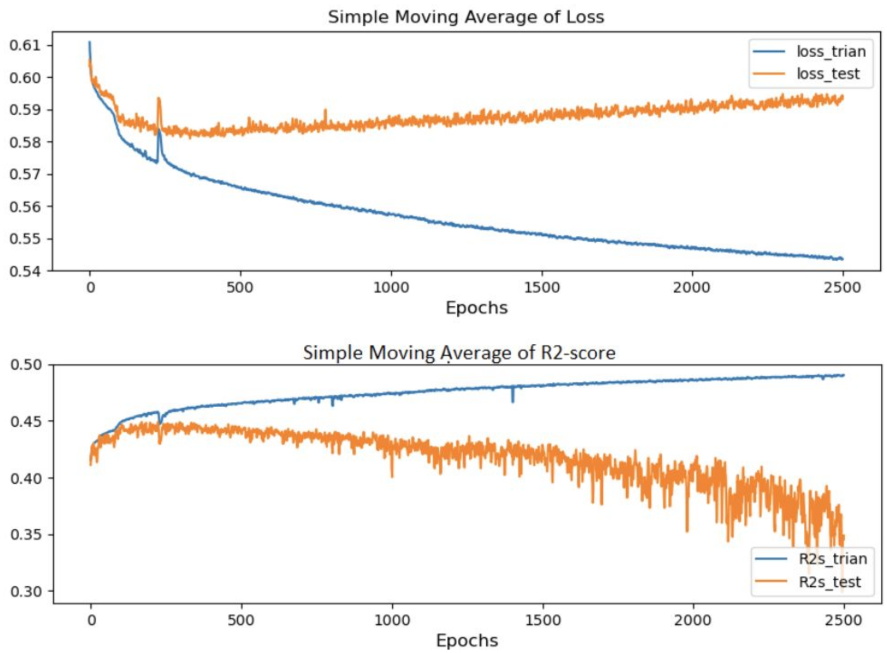


Figure 24 Dataset 'Weather in Szeged' - Overfitting in the ANN model.

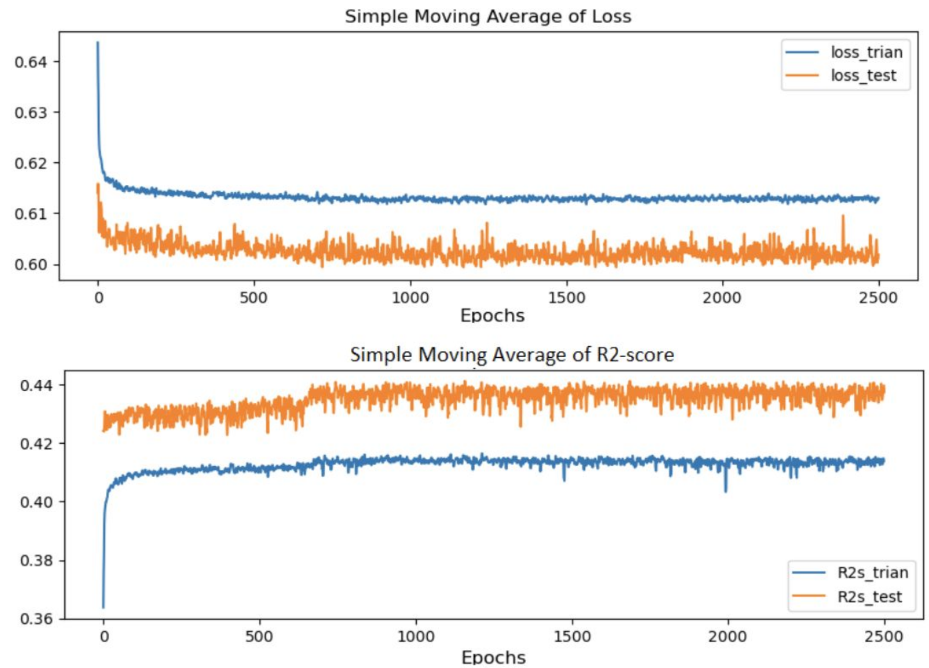


Figure 25 Simple Dropout, weather dataset

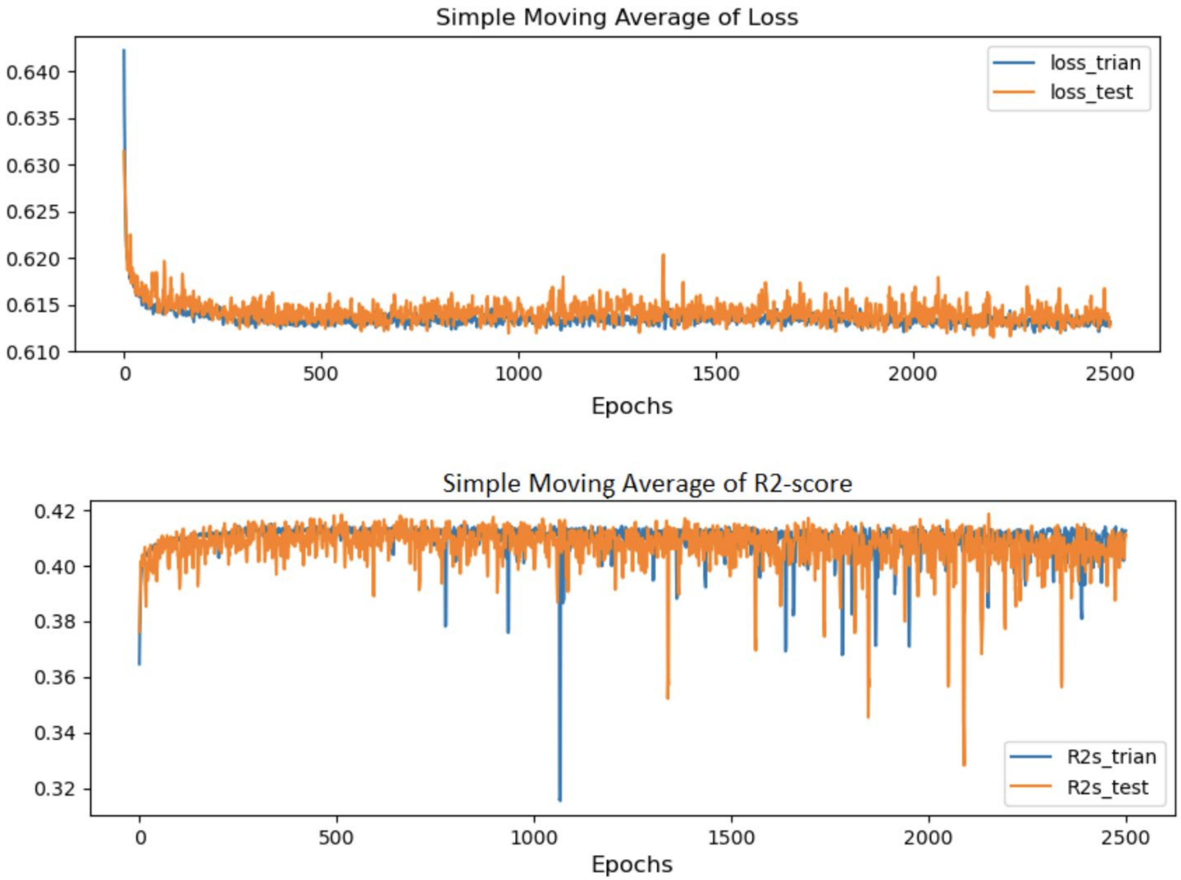


Figure 27 Gaussian Dropout, weather dataset

Gaussian dropout in california and boston

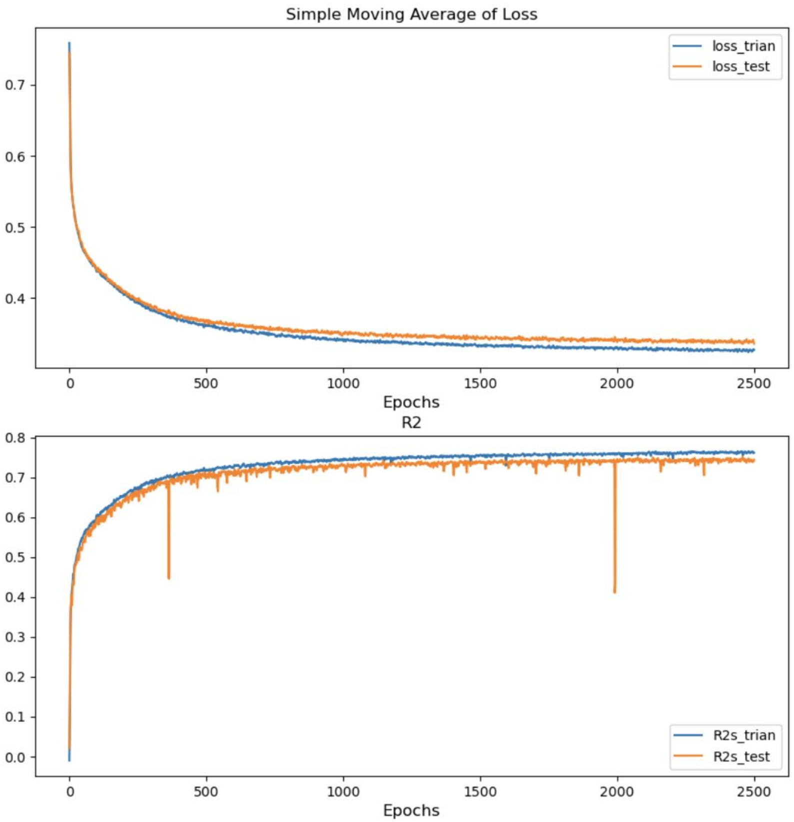


Figure 22 Gaussian Dropout, California Housing dataset

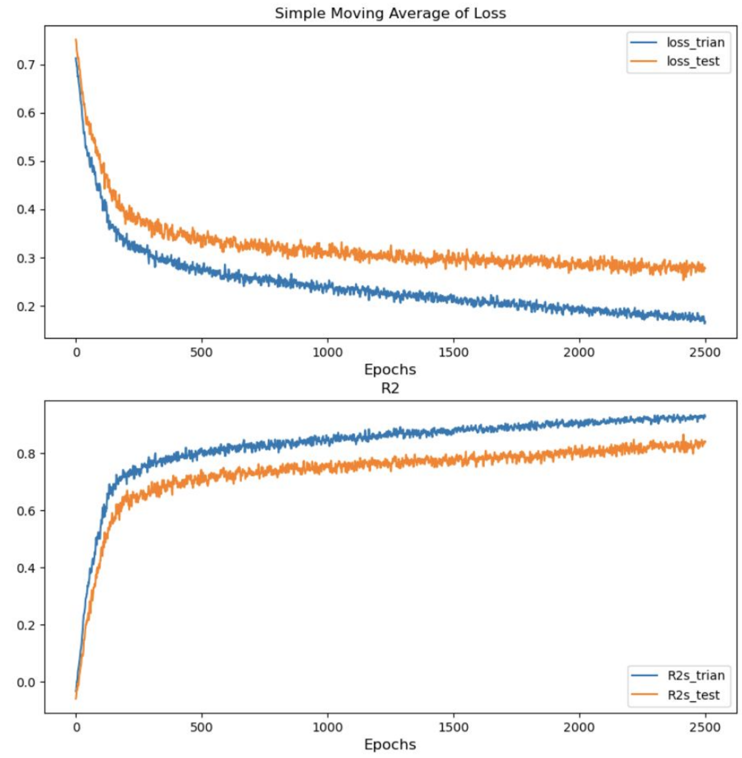


Figure 17 Boston dataset with Gaussian dropout

Results, BNG

BNG is synthetic dataset, generated by Bayesian Network. Many features and many samples.

The best solution was Simple dropout with R2 score 42.89%.

However, Advanced dropout prevented overfitting and reached max accuracy in 300 epochs.

Table 6 BNG dataset, results overview

Dropout method	Dropout probability (p)	Best Loss in test (Smaller the better)	Best R^2 score in test (Higher the better)	P-value (With regard to NoDropout with Dropout models)
No Dropout	-	0.5559	41.46%	-
Simple Dropout	0.5	**0.5412**	**42.89%**	0.0
Drop-Connect	0.5	0.6415	31.80%	0.0
Gaussian Dropout	0.5	0.5561	40.52%	$2.7 * 10^{-156}$
Advanced Dropout	-	0.5817	39.40%	$3 * 10^{-106}$

BNG Loss/R2 plots

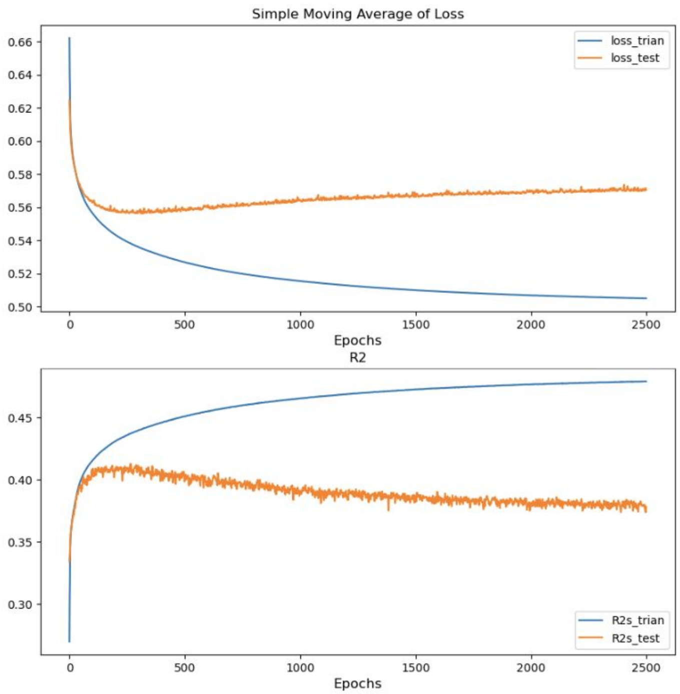


Figure 29 BNG dataset without dropout

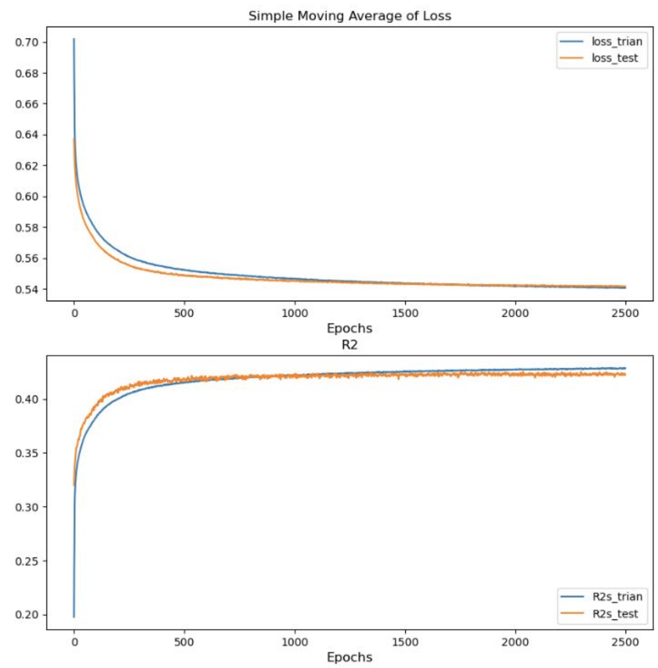


Figure 30 BNG dataset with Simple dropout

Further research

Make more diverse comparison research with classification, speech recognition, image classification or recognition tasks. Make comparison analysis between simple and advanced dropouts in those tasks.

Beside dropout, add other regularization methods: L1 and L2 norms and batch normalizations. Test them together.

Conclusions

First hypothesis is proved. No matter size or type of dropout method, it is preventing overfitting.

Second hypothesis is disproved. The Simple dropout had better R2 score results in 3 out of 4 datasets.

Drop-Connect is the worst method to implement. It was able to prevent the overfitting, however the R2 scores were the much lower in comparison with other dropout methods.

Advanced dropout is the good solution to replace gaussian dropout, as it produces more precise results and the max accuracy achieved faster, that other dropout methods .

Thank you for attention!