

MNIST Training

Document Author

Contents

1	Summary	1
2	Training reports	2
2.1	Model 1: ConvNet2layers	2
2.2	Model 2: ConvNet2layers	3
2.3	Model 3: ConvNet2layers	5
2.4	Model 4: ConvNet2layers	7
2.5	Model 5: MLP2layers	9
2.6	Model 6: MLP2layers	11
2.7	Model 7: MLP2layers	13
2.8	Model 8: MLP2layers	15
2.9	Model 9: MLP2layers	17
2.10	Model 10: MLP5layers	19
2.11	Model 11: MLP5layers	21
3	Model Architectures	24
3.1	MLP2layers	24
3.2	MLP5layers	25
3.3	ConvNet2layers	27

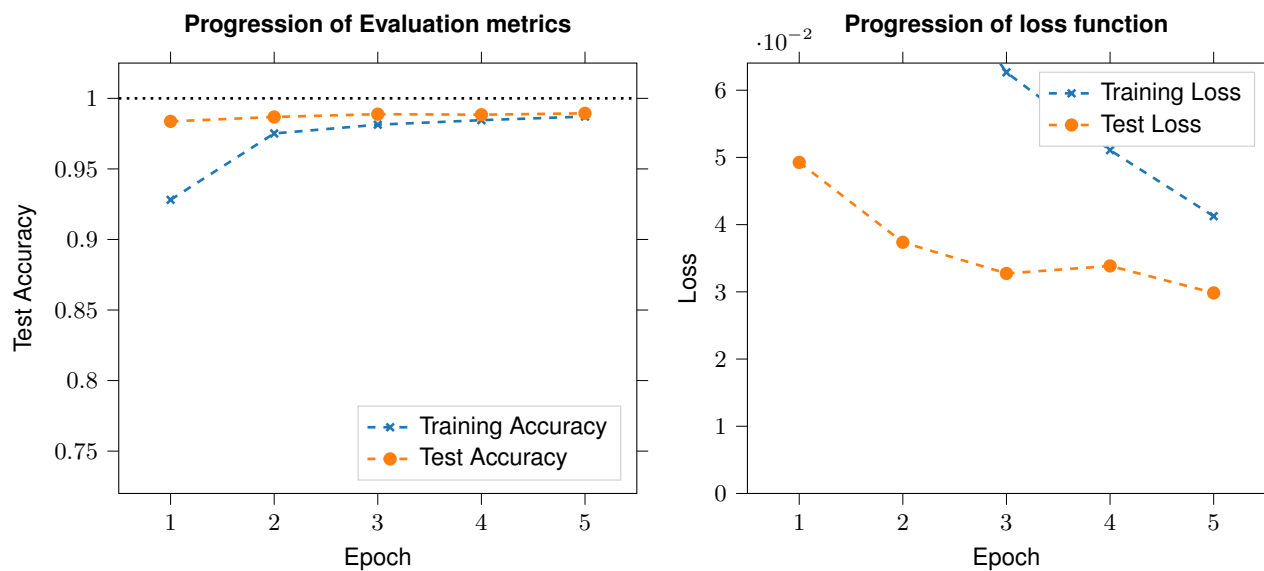
1 Summary

Nº	Model name	#Parameters	#Epochs	Batch size	Test Acc.	Training Acc.
1	ConvNet2layers	1 199 882	5	128	98.94 %	98.71 %
2	ConvNet2layers	1 199 882	5	128	99.02 %	98.73 %
3	ConvNet2layers	1 199 882	5	128	99.07 %	98.66 %
4	ConvNet2layers	1 199 882	5	128	99.06 %	98.66 %
5	MLP2layers	669 706	1	128	89.83 %	88.29 %
6	MLP2layers	669 706	5	128	92.72 %	92.66 %
7	MLP2layers	669 706	5	128	91.21 %	90.94 %
8	MLP2layers	669 706	5	128	92.62 %	92.78 %
9	MLP2layers	669 706	5	128	90.79 %	91.43 %
10	MLP5layers	430 602	5	128	91.93 %	89.04 %
11	MLP5layers	430 602	5	128	90.34 %	86.2 %

2 Training reports

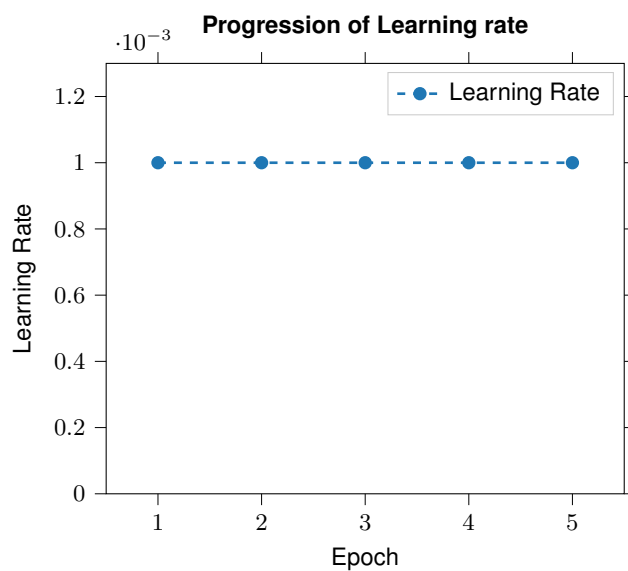
2.1 Model 1: ConvNet2layers

Training history See Figure 1.



(a) Accuracy learning process for model 1.

(b) Loss learning process for model 1.



(c) Learning rate per epoch for model 1.

Figure 1: Training and evaluation metrics for model 1.

Dataset

Name MNIST

Train-Test-Dev split: *Training set:* 60000, *Test set:* 10000, *Dev set:* 0,

Image size [28, 28]

Training

Number of epochs 5

Optimizer Adam (Kingma et al., 2015)

Learning Rate 0.0010000000474974513

Beta 1 0.89999999761581421

Beta 2 0.9990000128746033

Decay 0.0

Epsilon 1e-07

Amsgrad False

Loss Categorical crossentropy

Batch size 128

Shuffle Yes

Training time 34 sec

Platform

Weights exported to path weights\ConvNet2layers_5ep_MNIST.h5

Device used GPU (GeForce GTX 1060 6GB)

CPU Intel(R) Xeon(R) CPU E3-1245 v5 @ 3.50GHz, X86_64

Python Version 3.7.2.final.0 (64 bit)

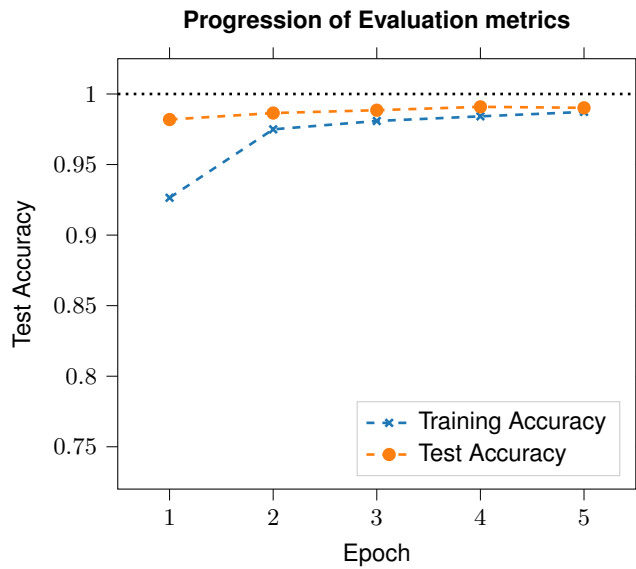
Keras Version 2.2.5 (Backend: tensorflow)

Tensorflow Version 1.14.0

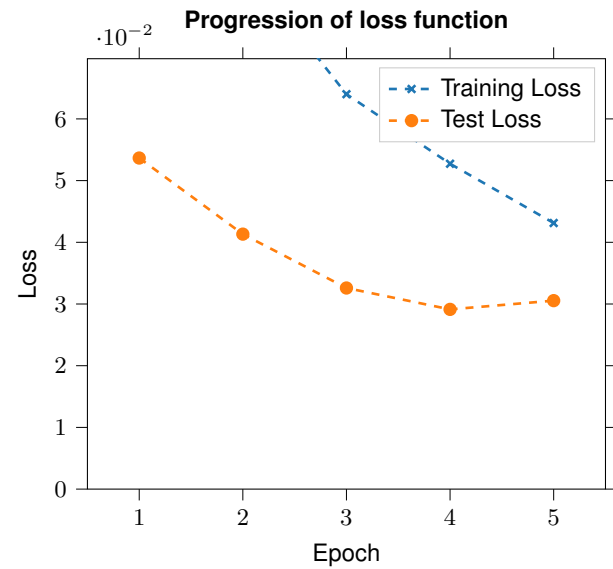
Timestamp 25.09.2019 at 16:02

2.2 Model 2: ConvNet2layers

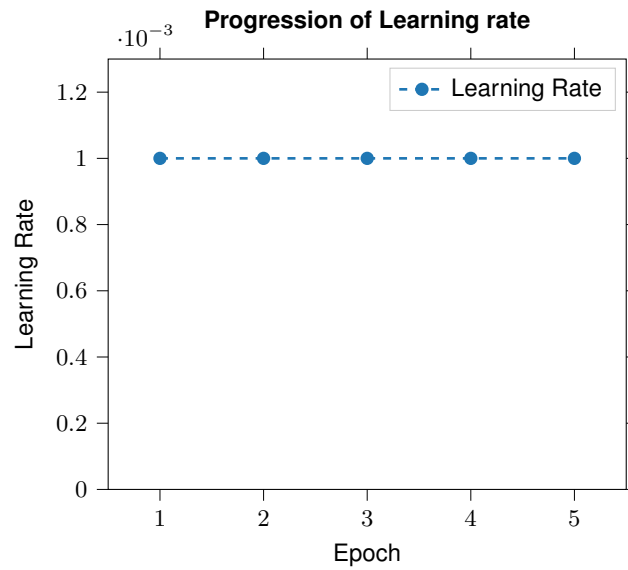
Training history See Figure 2.



(a) Accuracy learning process for model 2.



(b) Loss learning process for model 2.



(c) Learning rate per epoch for model 2.

Figure 2: Training and evaluation metrics for model 2.

Dataset

Name MNIST

Train-Test-Dev split: *Training set:* 60000, *Test set:* 10000, *Dev set:* 0,

Image size [28, 28]

Training

Number of epochs 5

Optimizer Adam (Kingma et al., 2015)

Learning Rate	0.0010000000474974513
Beta 1	0.89999999761581421
Beta 2	0.9990000128746033
Decay	0.0
Epsilon	1e-07
Amsgrad	False

Loss Categorical crossentropy

Batch size 128

Shuffle Yes

Training time 39 sec

Platform

Weights exported to path weights\ConvNet2layers_5ep_MNIST.h5

Device used GPU (GeForce GTX 1060 6GB)

CPU Intel(R) Xeon(R) CPU E3-1245 v5 @ 3.50GHz, X86_64

Python Version 3.7.2.final.0 (64 bit)

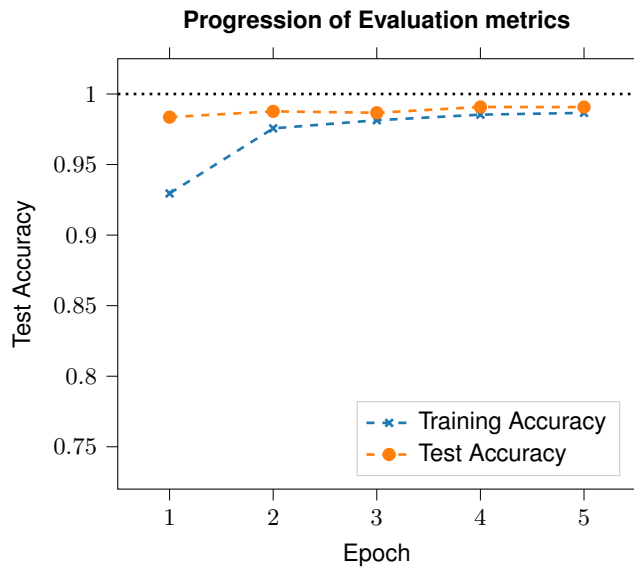
Keras Version 2.2.5 (Backend: tensorflow)

Tensorflow Version 1.14.0

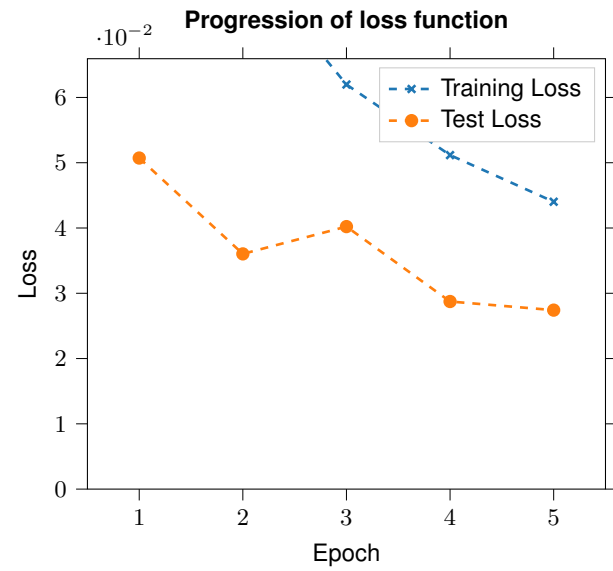
Timestamp 25.09.2019 at 16:04

2.3 Model 3: ConvNet2layers

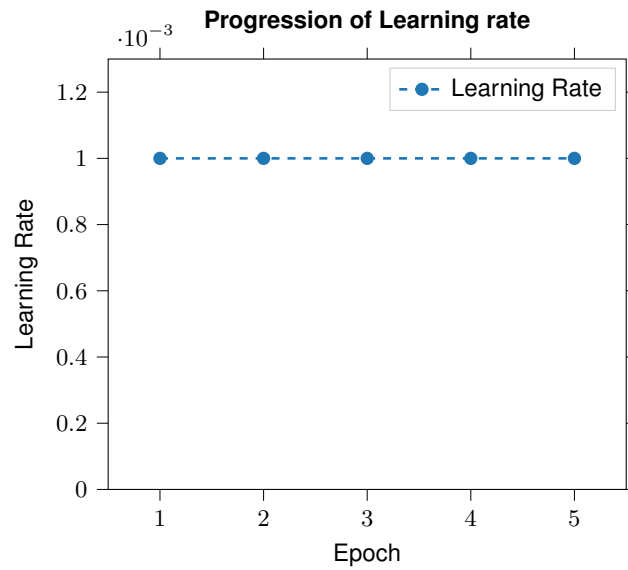
Training history See Figure 3.



(a) Accuracy learning process for model 3.



(b) Loss learning process for model 3.



(c) Learning rate per epoch for model 3.

Figure 3: Training and evaluation metrics for model 3.

Dataset

Name MNIST

Train-Test-Dev split: *Training set:* 60000, *Test set:* 10000, *Dev set:* 0,

Image size [28, 28]

Training

Number of epochs 5

Optimizer Adam (Kingma et al., 2015)

Learning Rate	0.0010000000474974513
Beta 1	0.89999999761581421
Beta 2	0.9990000128746033
Decay	0.0
Epsilon	1e-07
Amsgrad	False

Loss Categorical crossentropy

Batch size 128

Shuffle Yes

Training time 36 sec

Platform

Weights exported to path weights\ConvNet2layers_5ep_MNIST.h5

Device used GPU (GeForce GTX 1060 6GB)

CPU Intel(R) Xeon(R) CPU E3-1245 v5 @ 3.50GHz, X86_64

Python Version 3.7.2.final.0 (64 bit)

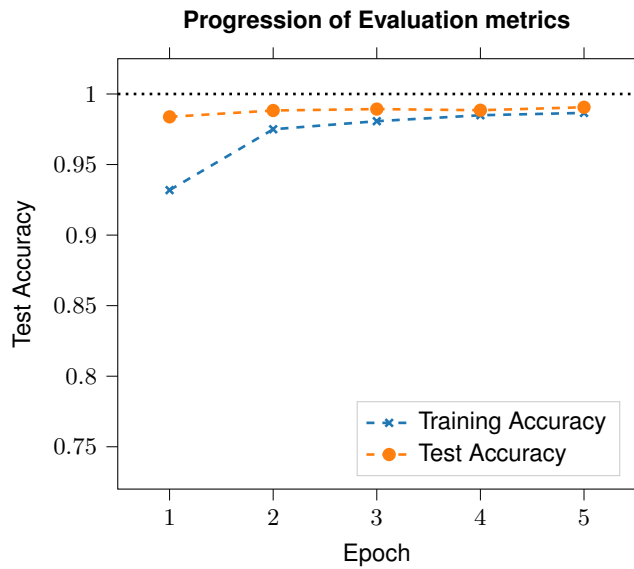
Keras Version 2.2.5 (Backend: tensorflow)

Tensorflow Version 1.14.0

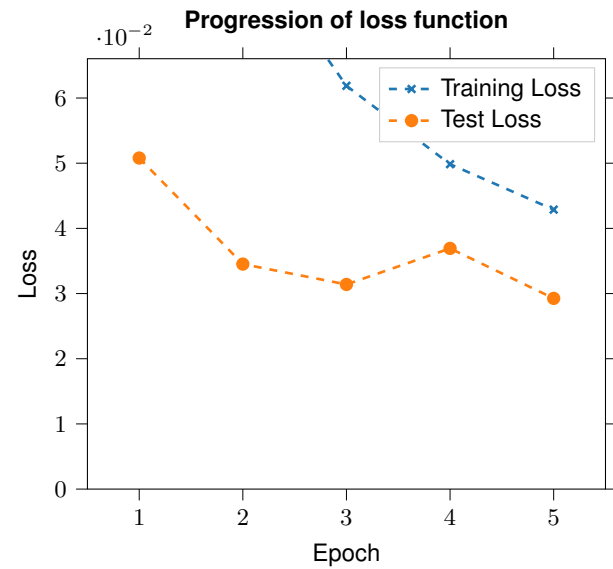
Timestamp 25.09.2019 at 16:07

2.4 Model 4: ConvNet2layers

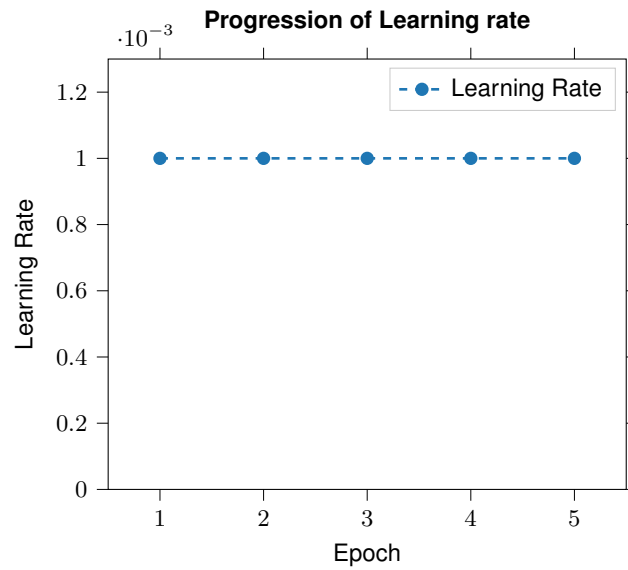
Training history See Figure 4.



(a) Accuracy learning process for model 4.



(b) Loss learning process for model 4.



(c) Learning rate per epoch for model 4.

Figure 4: Training and evaluation metrics for model 4.

Dataset

Name MNIST

Train-Test-Dev split: *Training set:* 60000, *Test set:* 10000, *Dev set:* 0,

Image size [28, 28]

Training

Number of epochs 5

Optimizer Adam (Kingma et al., 2015)

Learning Rate	0.0010000000474974513
Beta 1	0.89999999761581421
Beta 2	0.9990000128746033
Decay	0.0
Epsilon	1e-07
Amsgrad	False

Loss Categorical crossentropy

Batch size 128

Shuffle Yes

Training time 39 sec

Platform

Weights exported to path weights\ConvNet2layers_5ep_MNIST.h5

Device used GPU (GeForce GTX 1060 6GB)

CPU Intel(R) Xeon(R) CPU E3-1245 v5 @ 3.50GHz, X86_64

Python Version 3.7.2.final.0 (64 bit)

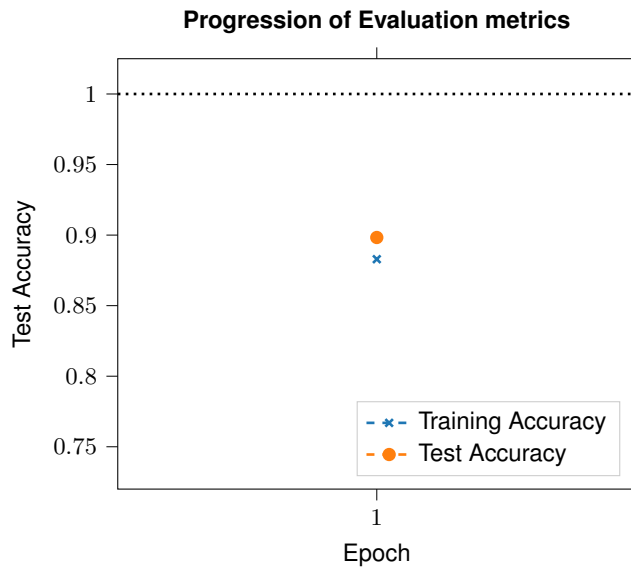
Keras Version 2.2.5 (Backend: tensorflow)

Tensorflow Version 1.14.0

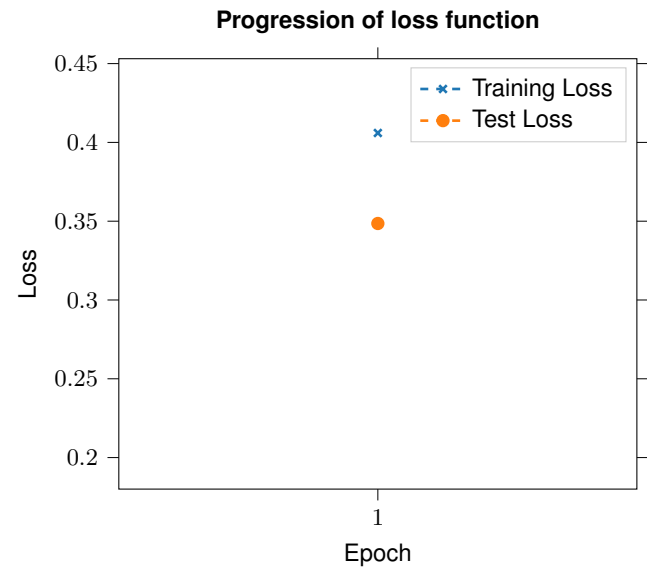
Timestamp 25.09.2019 at 16:10

2.5 Model 5: MLP2layers

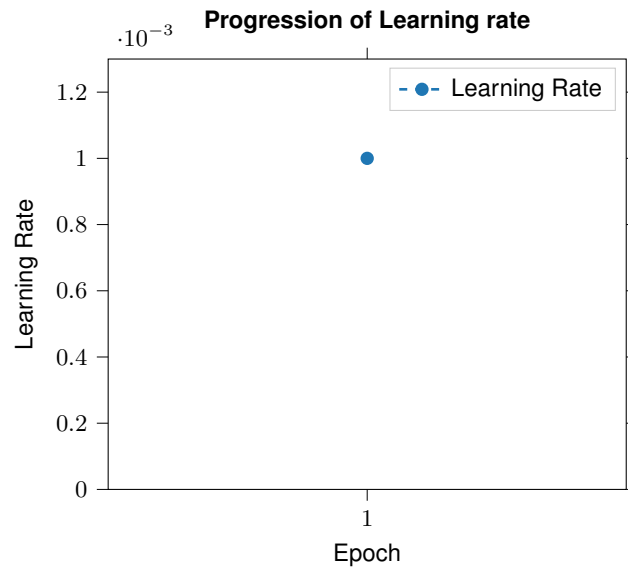
Training history See Figure 5.



(a) Accuracy learning process for model 5.



(b) Loss learning process for model 5.



(c) Learning rate per epoch for model 5.

Figure 5: Training and evaluation metrics for model 5.

Dataset

Name MNIST

Train-Test-Dev split: *Training set:* 60000, *Test set:* 10000, *Dev set:* 0,

Image size [28, 28]

Training

Number of epochs 1

Optimizer Adam (Kingma et al., 2015)

Learning Rate	0.0010000000474974513
Beta 1	0.89999999761581421
Beta 2	0.9990000128746033
Decay	0.0
Epsilon	1e-07
Amsgrad	False

Loss Categorical crossentropy

Batch size 128

Shuffle Yes

Training time 5 sec

Platform

Weights exported to path weights\MLP2layers_1ep_MNIST.h5

Device used GPU (GeForce GTX 1060 6GB)

CPU Intel(R) Xeon(R) CPU E3-1245 v5 @ 3.50GHz, X86_64

Python Version 3.7.2.final.0 (64 bit)

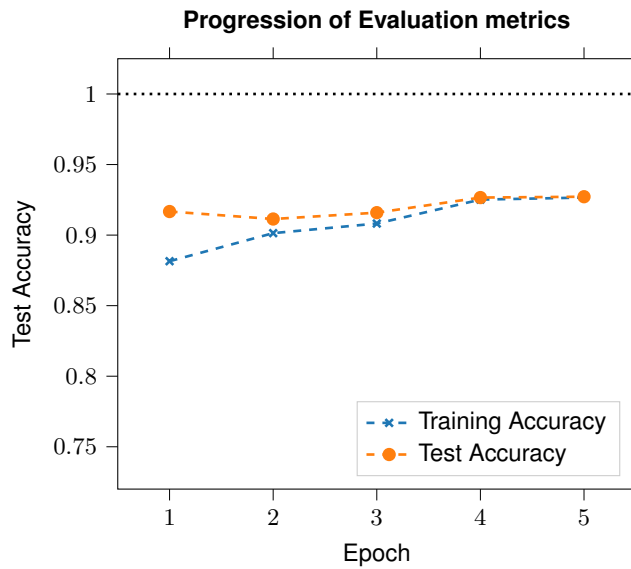
Keras Version 2.2.5 (Backend: tensorflow)

Tensorflow Version 1.14.0

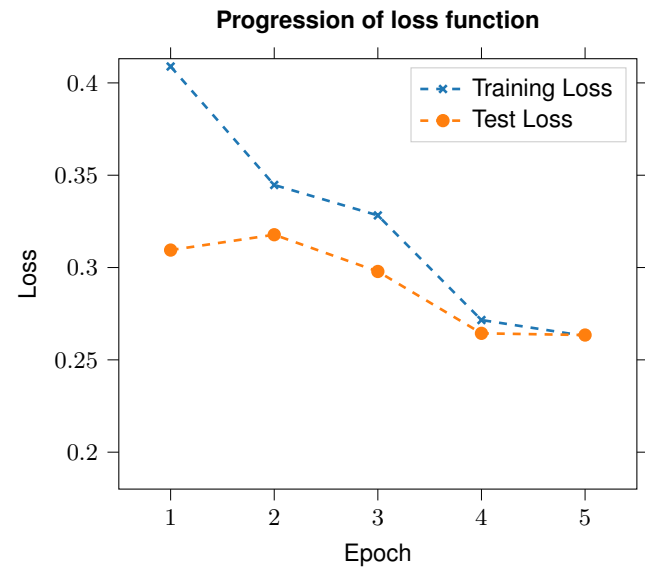
Timestamp 25.09.2019 at 15:16

2.6 Model 6: MLP2layers

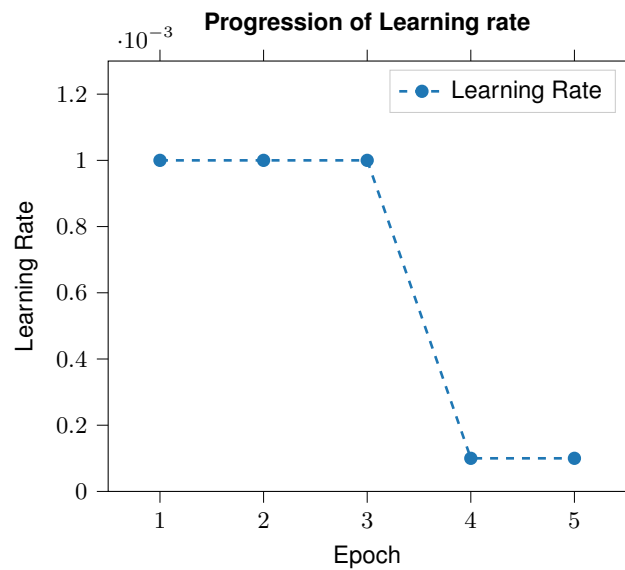
Training history See Figure 6.



(a) Accuracy learning process for model 6.



(b) Loss learning process for model 6.



(c) Learning rate per epoch for model 6.

Figure 6: Training and evaluation metrics for model 6.

Dataset

Name MNIST

Train-Test-Dev split: *Training set:* 60000, *Test set:* 10000, *Dev set:* 0,

Image size [28, 28]

Training

Number of epochs 5

Optimizer Adam (Kingma et al., 2015)

Learning Rate	0.000100000000474974513
Beta 1	0.89999999761581421
Beta 2	0.9990000128746033
Decay	0.0
Epsilon	1e-07
Amsgrad	False

Loss Categorical crossentropy

Batch size 128

Shuffle Yes

Training time 18 sec

Platform

Weights exported to path weights\MLP2layers_5ep_MNIST.h5

Device used GPU (GeForce GTX 1060 6GB)

CPU Intel(R) Xeon(R) CPU E3-1245 v5 @ 3.50GHz, X86_64

Python Version 3.7.2.final.0 (64 bit)

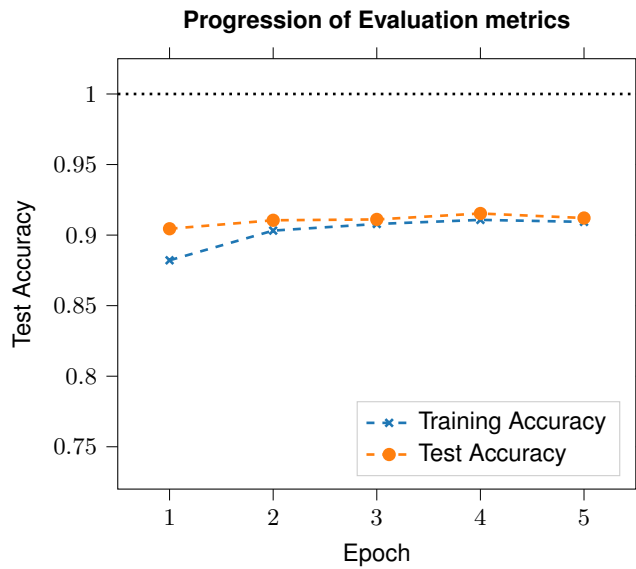
Keras Version 2.2.5 (Backend: tensorflow)

Tensorflow Version 1.14.0

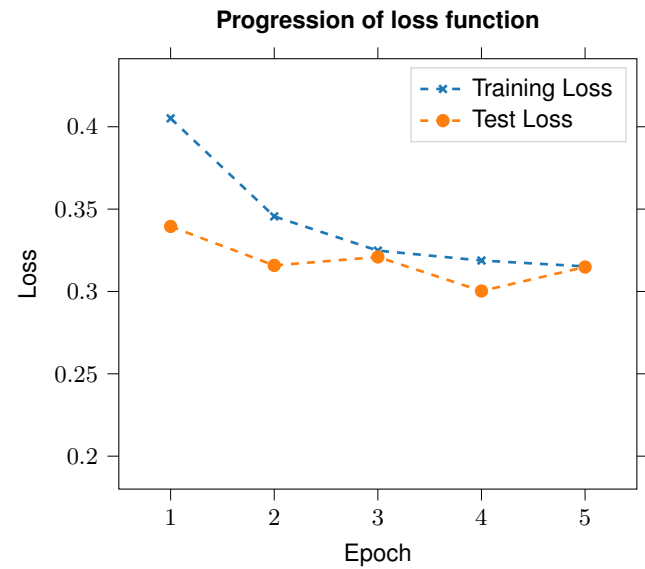
Timestamp 25.09.2019 at 16:02

2.7 Model 7: MLP2layers

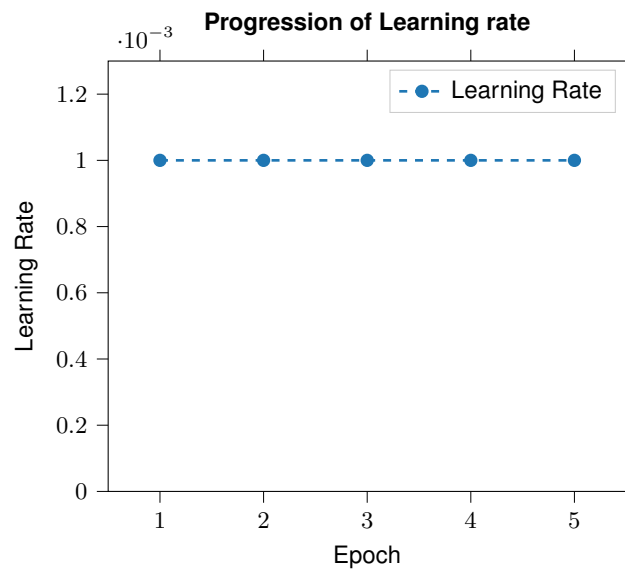
Training history See Figure 7.



(a) Accuracy learning process for model 7.



(b) Loss learning process for model 7.



(c) Learning rate per epoch for model 7.

Figure 7: Training and evaluation metrics for model 7.

Dataset

Name MNIST

Train-Test-Dev split: *Training set:* 60000, *Test set:* 10000, *Dev set:* 0,

Image size [28, 28]

Training

Number of epochs 5

Optimizer Adam (Kingma et al., 2015)

Learning Rate	0.0010000000474974513
Beta 1	0.89999999761581421
Beta 2	0.9990000128746033
Decay	0.0
Epsilon	1e-07
Amsgrad	False

Loss Categorical crossentropy

Batch size 128

Shuffle Yes

Training time 18 sec

Platform

Weights exported to path weights\MLP2layers_5ep_MNIST.h5

Device used GPU (GeForce GTX 1060 6GB)

CPU Intel(R) Xeon(R) CPU E3-1245 v5 @ 3.50GHz, X86_64

Python Version 3.7.2.final.0 (64 bit)

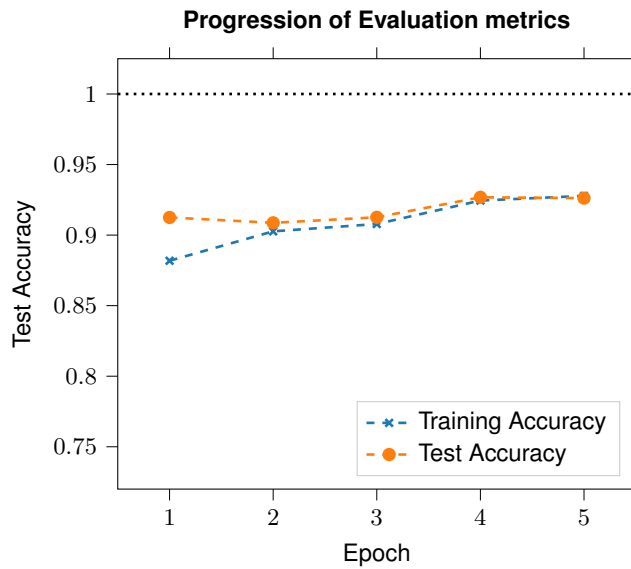
Keras Version 2.2.5 (Backend: tensorflow)

Tensorflow Version 1.14.0

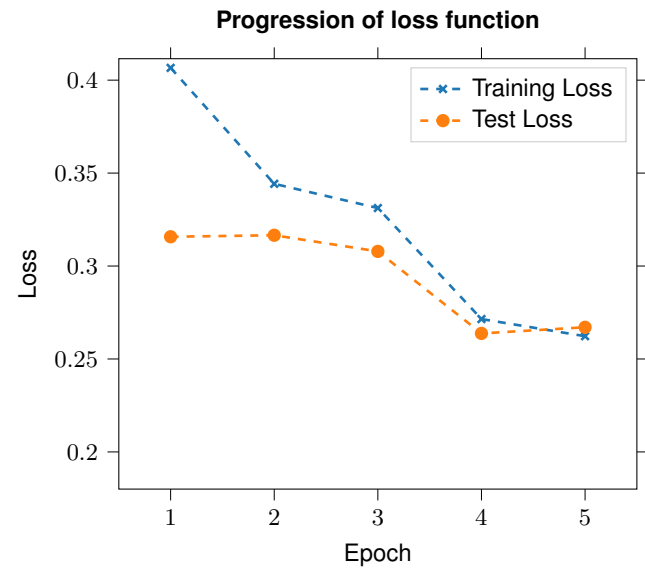
Timestamp 25.09.2019 at 16:05

2.8 Model 8: MLP2layers

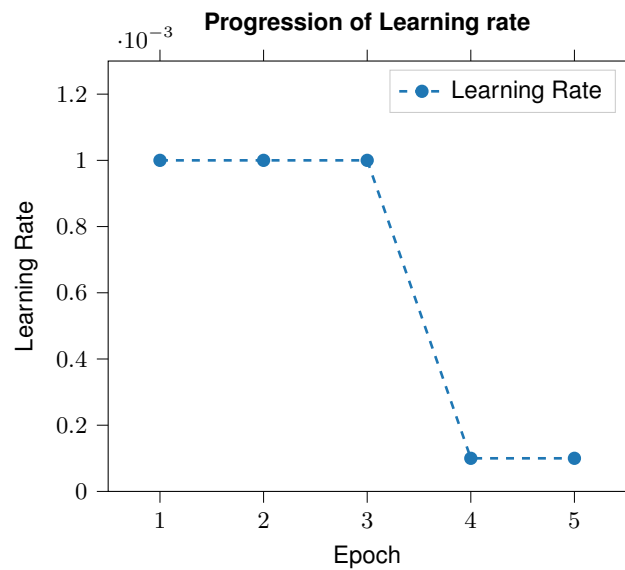
Training history See Figure 8.



(a) Accuracy learning process for model 8.



(b) Loss learning process for model 8.



(c) Learning rate per epoch for model 8.

Figure 8: Training and evaluation metrics for model 8.

Dataset

Name MNIST

Train-Test-Dev split: *Training set:* 60000, *Test set:* 10000, *Dev set:* 0,

Image size [28, 28]

Training

Number of epochs 5

Optimizer Adam (Kingma et al., 2015)

Learning Rate	0.000100000000474974513
Beta 1	0.89999999761581421
Beta 2	0.9990000128746033
Decay	0.0
Epsilon	1e-07
Amsgrad	False

Loss Categorical crossentropy

Batch size 128

Shuffle Yes

Training time 19 sec

Platform

Weights exported to path weights\MLP2layers_5ep_MNIST.h5

Device used GPU (GeForce GTX 1060 6GB)

CPU Intel(R) Xeon(R) CPU E3-1245 v5 @ 3.50GHz, X86_64

Python Version 3.7.2.final.0 (64 bit)

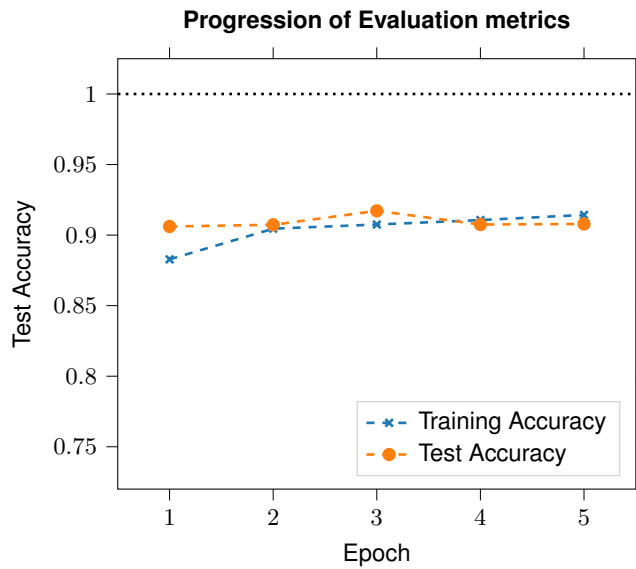
Keras Version 2.2.5 (Backend: tensorflow)

Tensorflow Version 1.14.0

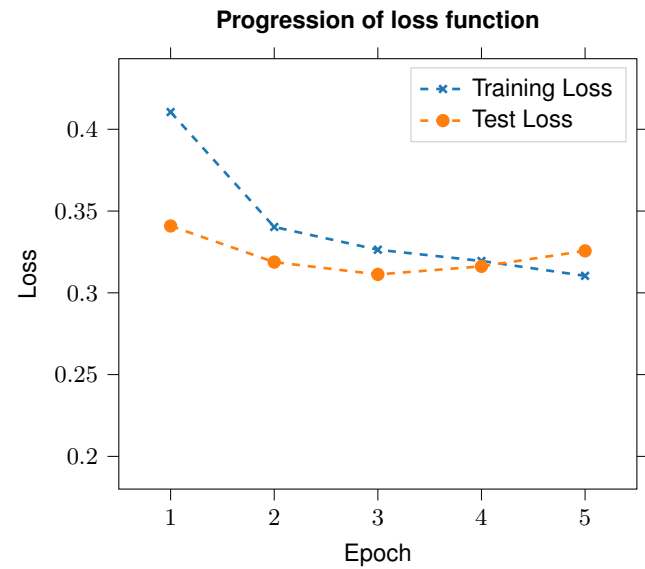
Timestamp 25.09.2019 at 16:08

2.9 Model 9: MLP2layers

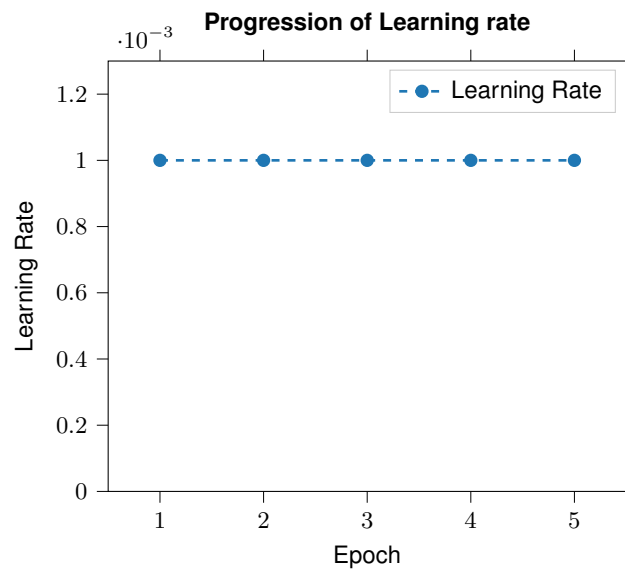
Training history See Figure 9.



(a) Accuracy learning process for model 9.



(b) Loss learning process for model 9.



(c) Learning rate per epoch for model 9.

Figure 9: Training and evaluation metrics for model 9.

Dataset

Name MNIST

Train-Test-Dev split: *Training set:* 60000, *Test set:* 10000, *Dev set:* 0,

Image size [28, 28]

Training

Number of epochs 5

Optimizer Adam (Kingma et al., 2015)

Learning Rate	0.000100000000474974513
Beta 1	0.89999999761581421
Beta 2	0.9990000128746033
Decay	0.0
Epsilon	1e-07
Amsgrad	False

Loss Categorical crossentropy

Batch size 128

Shuffle Yes

Training time 18 sec

Platform

Weights exported to path weights\MLP2layers_5ep_MNIST.h5

Device used GPU (GeForce GTX 1060 6GB)

CPU Intel(R) Xeon(R) CPU E3-1245 v5 @ 3.50GHz, X86_64

Python Version 3.7.2.final.0 (64 bit)

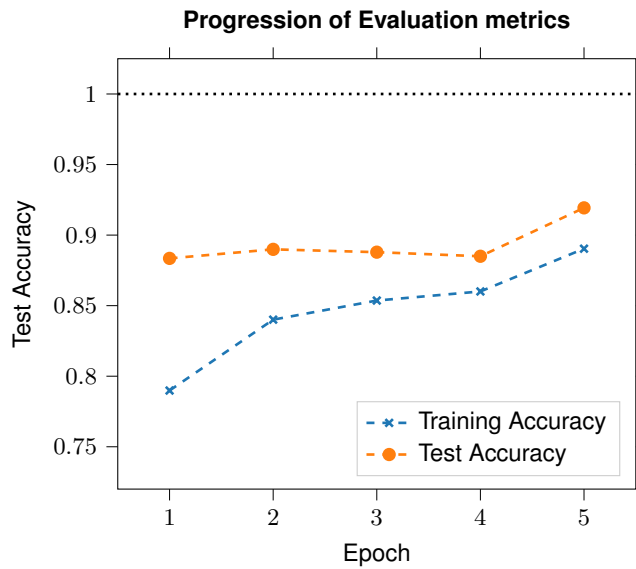
Keras Version 2.2.5 (Backend: tensorflow)

Tensorflow Version 1.14.0

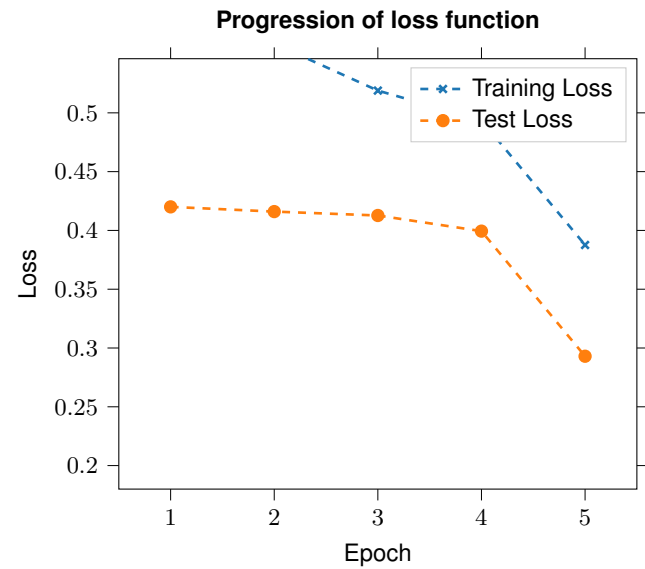
Timestamp 25.09.2019 at 16:10

2.10 Model 10: MLP5layers

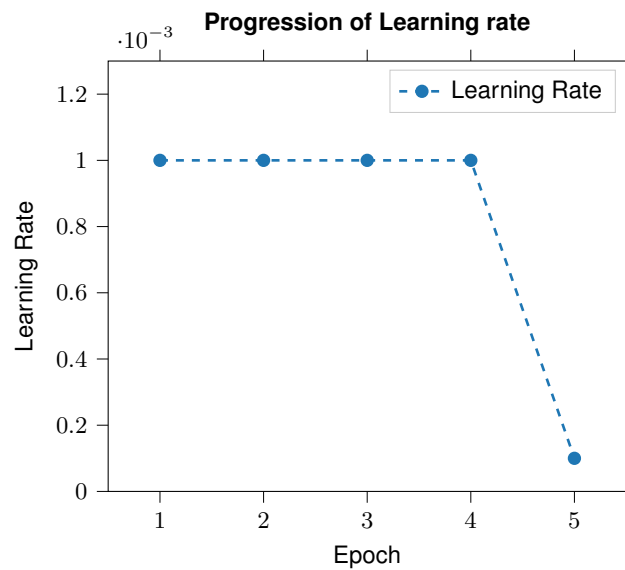
Training history See Figure 10.



(a) Accuracy learning process for model 10.



(b) Loss learning process for model 10.



(c) Learning rate per epoch for model 10.

Figure 10: Training and evaluation metrics for model 10.

Dataset

Name MNIST

Train-Test-Dev split: *Training set:* 60000, *Test set:* 10000, *Dev set:* 0,

Image size [28, 28]

Training

Number of epochs 5

Optimizer RMSProp

Learning Rate 0.000100000000474974513

Rho 0.89999999761581421

Decay 0.0

Epsilon 1e-07

Loss Categorical crossentropy

Batch size 128

Shuffle Yes

Training time 23 sec

Platform

Weights exported to path weights\MLP5layers_5ep_MNIST.h5

Device used GPU (GeForce GTX 1060 6GB)

CPU Intel(R) Xeon(R) CPU E3-1245 v5 @ 3.50GHz, X86_64

Python Version 3.7.2.final.0 (64 bit)

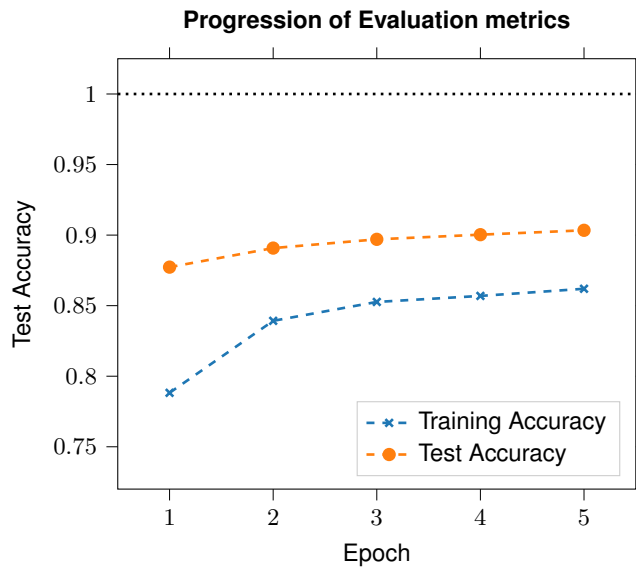
Keras Version 2.2.5 (Backend: tensorflow)

Tensorflow Version 1.14.0

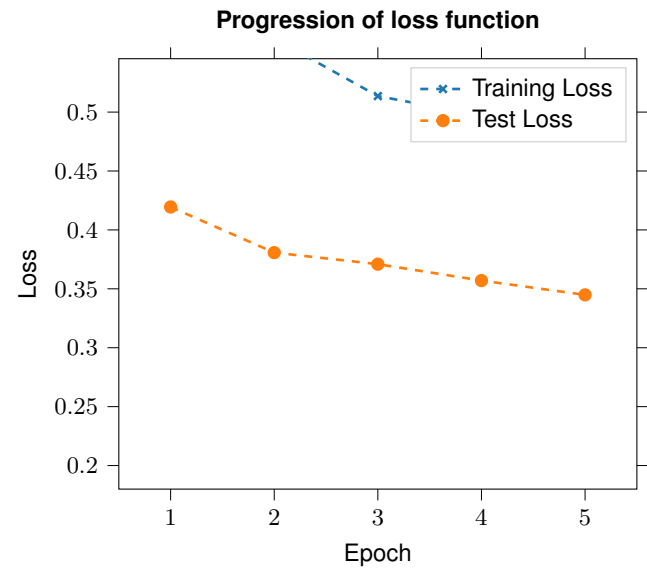
Timestamp 25.09.2019 at 16:08

2.11 Model 11: MLP5layers

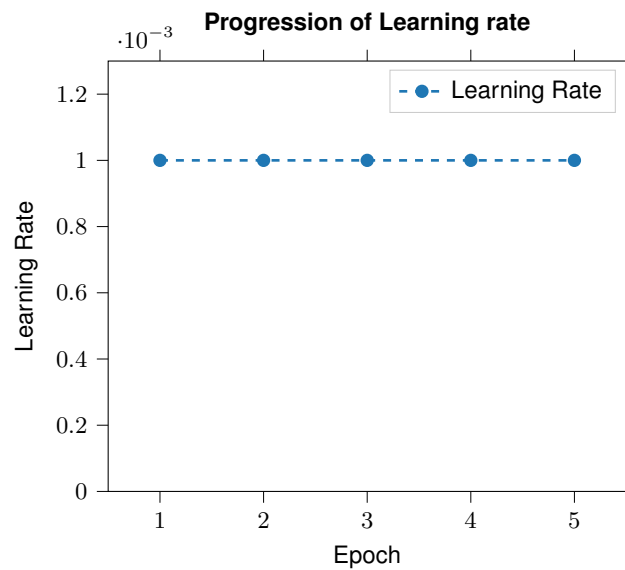
Training history See Figure 11.



(a) Accuracy learning process for model 11.



(b) Loss learning process for model 11.



(c) Learning rate per epoch for model 11.

Figure 11: Training and evaluation metrics for model 11.

Dataset

Name MNIST

Train-Test-Dev split: *Training set:* 60000, *Test set:* 10000, *Dev set:* 0,

Image size [28, 28]

Training

Number of epochs 5

Optimizer RMSProp

Learning Rate 0.0010000000474974513

Rho 0.8999999761581421

Decay 0.0

Epsilon 1e-07

Loss Categorical crossentropy

Batch size 128

Shuffle Yes

Training time 21 sec

Platform

Weights exported to path weights\MLP5layers_5ep_MNIST.h5

Device used GPU (GeForce GTX 1060 6GB)

CPU Intel(R) Xeon(R) CPU E3-1245 v5 @ 3.50GHz, X86_64

Python Version 3.7.2.final.0 (64 bit)

Keras Version 2.2.5 (Backend: tensorflow)

Tensorflow Version 1.14.0

Timestamp 25.09.2019 at 16:11

3 Model Architectures

3.1 MLP2layers

Used in №: 5, 6, 7, 8, 11

Model summary:

Nº	Layer (Type)	Output shape	Config	#Parameters	Inbound layers
0	input_1 (InputLayer)	(28, 28, 1)		0	
1	flatten_1 (Flatten)	(784,)	Parameters of layers of type Flatten not implemented.	0	input_1
2	dense_1 (Dense)	(512,)	Parameters of layers of type Dense not implemented.	401 920	flatten_1
3	dropout_1 (Dropout)	(512,)	Parameters of layers of type Dropout not implemented.	0	dense_1
4	dense_2 (Dense)	(512,)	Parameters of layers of type Dense not implemented.	262 656	dropout_1
5	dropout_2 (Dropout)	(512,)	Parameters of layers of type Dropout not implemented.	0	dense_2
6	dense_3 (Dense)	(10,)	Parameters of layers of type Dense not implemented.	5130	dropout_2

3.2 MLP5layers

Used in №: 10, 11

Model summary:

Nº	Layer (Type)	Output shape	Config	#Parameters	Inbound layers
0	input_5 (InputLayer)	(28, 28, 1)		0	
1	flatten_5 (Flatten)	(784,)	Parameters of layers of type Flatten not implemented.	0	input_5
2	dense_11 (Dense)	(128,)	Parameters of layers of type Dense not implemented.	100 480	flatten_5
3	dropout_9 (Dropout)	(128,)	Parameters of layers of type Dropout not implemented.	0	dense_11
4	dense_12 (Dense)	(256,)	Parameters of layers of type Dense not implemented.	33 024	dropout_9
5	dropout_10 (Dropout)	(256,)	Parameters of layers of type Dropout not implemented.	0	dense_12
6	dense_13 (Dense)	(512,)	Parameters of layers of type Dense not implemented.	131 584	dropout_10
7	dropout_11 (Dropout)	(512,)	Parameters of layers of type Dropout not implemented.	0	dense_13
8	dense_14 (Dense)	(256,)	Parameters of layers of type Dense not implemented.	131 328	dropout_11
9	dropout_12 (Dropout)	(256,)	Parameters of layers of type Dropout not implemented.	0	dense_14
10	dense_15 (Dense)	(128,)	Parameters of layers of type Dense not implemented.	32 896	dropout_12
11	dropout_13 (Dropout)	(128,)	Parameters of layers of type Dropout not implemented.	0	dense_15

Nº	Layer (Type)	Output shape	Config	#Parameters	Inbound layers
12	dense_16 (Dense)	(10,)	Parameters of layers of type Dense not implemented.	1290	dropout_13

3.3 ConvNet2layers

Used in №: 1, 2, 3, 11

Model summary:

Nº	Layer (Type)	Output shape	Config	#Parameters	Inbound layers
0	input_4 (InputLayer)	(28, 28, 1)		0	
1	conv2d_7 (Conv2D)	(26, 26, 32)	Activation: relu Kernel Size: [3, 3] Stride: [1, 1] Dilation: [1, 1] Padding: valid	320	input_4
2	conv2d_8 (Conv2D)	(24, 24, 64)	Activation: relu Kernel Size: [3, 3] Stride: [1, 1] Dilation: [1, 1] Padding: valid	18 496	conv2d_7
3	max_pooling2d_4 (MaxPooling2D)	(12, 12, 64)	Pool size: [2, 2] Strides: [2, 2] Padding: valid	0	conv2d_8
4	dropout_7 (Dropout)	(12, 12, 64)	Parameters of layers of type Dropout not implemented.	0	max_pooling2d_4
5	flatten_4 (Flatten)	(9216,)	Parameters of layers of type Flatten not implemented.	0	dropout_7
6	dense_7 (Dense)	(128,)	Parameters of layers of type Dense not implemented.	1 179 776	flatten_4
7	dropout_8 (Dropout)	(128,)	Parameters of layers of type Dropout not implemented.	0	dense_7
8	dense_8 (Dense)	(10,)	Parameters of layers of type Dense not implemented.	1290	dropout_8