Course: Artificial Neural Networks and Deep Learning

Exercises Unit 2: Deep Learning Methods

Exercise 1. Performance evaluation

Train error	Test error	Time (s)
16.9	21.3	94
17.3	20.8	94
16.5	21.9	95

Exercise 2: Changing basic hyperparameters

Experiments with batch size

N: nº de veces que los pesos se actualizan

Batch size	N	Train error	Test error	Time (s)
16,342	1000	28.4%	29.5%	19
8192	1,998	27.4%	27.2%	22
4096	3,990	26.5%	25.5%	26
2048	7,980	24.6%	24.6%	35
1024	15,959	21.9%	22.1%	58
512	**31,918**	**17.3%**	**20.8%**	**94**
256	63,836	11.4%	21.0%	166
128	127,672	7.1%	21.4%	346
64	255,344	5.7%	26.5%	679

N: Number of times that the weights are updated

Experiments with network structure (keeping batch size with 512)

La primera produce overfitting

Structure	Train error	Test error	Time (s)
[1000, 500, 250, 75, 25]	15.4	22.9	101
[500, 250, 75, 25]	**16.9**	**20.9**	**97**
[250, 75, 25]	19.7	24.3	93
[75, 25]	22.1	21.7	90
[25]	24.4	24.9	88

Exercise 3: Changing activation functions

Function	Train error	Test error	Seconds
relu	17.7%	20.8%	96
tanh	**17.5%**	**20.5%**	**108**
elu	20.6%	23.3%	98
linear	28.2%	28.1%	106

Exercise 4: Changing initializers

Initializer	Train eror	Test error	Time
None	17.7%	20.8%	96
Uniform (-0.1,0.1)	19.2%	20.9%	98
Uniform (-0.3,0.3)	15.5%	22.6%	107
Uniform (-0.05,0.05)	20.2%	21.6%	96
Normal	18.7%	21.8%	96
He (Normal)	**14.1%**	**20.7%**	**98**
He (Uniform)	15.1%	23.1%	99

Exercise 5: Using batch normalization

Batch normalization	Train error	Test error	Time
Without normalization	17.7%	20.8%	96
Batch normalization (with elu) [after]	**11.8%**	**20.2%**	**184**
Batch normalization (with elu) [before]	14.3%	20.5%	178

Exercise 6: Using regularization

Epochs=512

Regularizer	Train error	Test error	Time
Without regularization	17.5%	22.5%	100
Regularizer L2 (lambda=0.001)	27.5%	25.4%	117
Regularizer L2 (lambda=0.0001)	19.2%	21.0%	116
Regularizer L2 (lambda=0.00005)	18.4%	21.6%	116
Regularizer L2 (lambda=0.00001)	18.5%	23.0%	115
Regularizer L1(lambda=0.0001)	21.8%	22.3%	116
Dropout (rate=0.2)	24.1%	23.3%	108
Dropout (rate=0.1)	22.0%	21.8%	108
Dropout (rate=0.01)	**17.9%**	**19.9%**	**109**
Dropout (rate=0.001)	17.0%	21.3%	110

Epochs=256

Regularizer	Train error	Test error	Time
Without regularization	11.4%	21.0%	166
Regularizer L2 (lambda=0.001)	26.3%	27.4%	184
Regularizer L2 (lambda=0.0001)	16.7%	19.3%	187
Regularizer L2 (lambda=0.00001)	13.3%	20.7%	185
Regularizer L1(lambda=0.0001)	20.9%	21.0%	189
Dropout (rate=0.2)	22.2%	20.8%	178
Dropout (rate=0.1)	20.3%	21.3%	176
Dropout (rate=0.01)	**15.7%**	**19.3%**	**177**
Dropout (rate=0.001)	14.6%	20.5%	177

Exercise 7: Changing learning rate and epochs

Learning rate	Train error	Test error	Time
1	11.1%	26.5%	102
0.7	9.8%	24.6%	99
0.5	12.2%	23.1%	105
0.1	**17.9%**	**22.0%**	**104**
0.05	19..9%	23.3%	103
0.01	23.0%	23.8%	100
0.001	27.9%	27.8%	97

Epochs	Train error	Test error	Time
100	26.1%	26.7%	11
500	22.0%	23.0%	50
1000	17.9%	22.0%	104
2000	**11.3%**	**21.1%**	**212**
4000	0.8%	21.5.0%	409

Exercise 8: Changing optimizers

Optimizer	Learning rate	Train error	Test error	Time
SGD	0.1	17.9%	22.0%	104
Momentum	0.1	6.0%	23.6%	109
Nesterov	0.1	4.9%	22.4%	108
RSMprop	0.1	66.6%	66.9%	125
RSMprop	0.01	21.7%	24.0%	123
RSMprop	0.001	6.1%	23.8%	128
RSMprop	0.0001	18.4%	19.8%	126
Adam	0.001	4.4%	20.7%	106
Adam	**0.0001**	**16.7%**	**18.9%**	**107**

Exercise 9: Multiple changes

L. rate	Epochs	Initializer	Activation	Optimizer	Dropout.	Train error	Test error	Time
0.001	1000	He normal	Relu	Adam	0.01	10.0%	19.1%	109
0.0005	1000	He normal	Relu	Adam	0.01	12.7%	19.1%	117
0.001	2000	He normal	Relu	Adam	0.01	6.0%	19.4%	215
0.001	2000	He normal	Relu	Adam	0.05	10.2%	18.9%	216
0.001	3000	He normal	Relu	Adam	0.05	9.3%	19.4%	320
0.001	4000	He normal	Relu	Adam	0.05	8.6%	18.9%	457
0.001	4000	He normal	Relu	Adam	0.07	10.2%	18.8%	424
0.001	4000	He normal	Relu	Adam	0.1	11.7%	18.0%	428
0.001	**5000**	**He normal**	**Relu**	**Adam**	**0.15**	**13.1%**	**16.3%**	**541**