Course: Artificial Neural Networks and Deep Learning

Exercises Unit 2: Deep Learning Methods

Results

Exercise 1. Performance evaluation

Train error	Test error	Bias	Variance	Time (s)
16.9	21.3	11.9	4.4	94
17.3	20.8	12.3	3.5	94
16.5	21.9	11.5	5.4	95

Exercise 2: Changing basic hyperparameters

Experiments with batch size

Batch size	N	Train error	Test error	Bias	Variance	Time (s)
16,342	1000	28.4	29.5	23.4	1.1	19
8192	1,998	27.4	27.2	22.4	-0.2	22
4096	3,990	26.5	25.5	21.5	-1.0	26
2048	7,980	24.6	24.6	19.6	0.0	35
1024	15,959	21.9	22.1	19.9	0.2	58
512	31,918	17.3	20.8	12.3	3.5	94
256	63,836	11.4	21.0	6.4	9.6	166
128	127,672	7.1	21.4	2.1	14.3	346
64	255,344	5.7	26.5	0.7	20.8	679

N: Number of times that the weights are updated (iterations)

Experiments with network structure (keeping batch size with 512)

Structure	Train error	Test error	Bias	Variance	Time (s)
[1000, 500, 250, 75, 25]	15.4	22.9	10.4	7.5	101
[500, 250, 75, 25]	16.9	20.9	11.9	4.0	97
[250, 75, 25]	19.7	24.3	14.7	4.6	93
[75, 25]	22.1	21.7	17.1	-0.4	90
[25]	24.4	24.9	19.4	0.5	88

Exercise 3: Changing activation functions

Function	Train error	Test error	Bias	Variance	Time (s)
relu	17.7	20.8	12.7	3.1	96
tanh	17.5	20.5	12.5	3.0	108
elu	20.6	23.3	15.6	2.7	98
linear	28.2	28.1	23.2	-0.1	106

Exercise 4: Changing initializers

Initializer	Train eror	Test error	Bias	Variance	Time
None	17.7	20.8	12.7	3.1	96
Uniform (-0.1,0.1)	19.2	20.9	14.2	1.7	98
Uniform (-0.3,0.3)	15.5	22.6	10.5	7.1	107
Uniform (-0.05,0.05)	20.2	21.6	15.2	1.4	96
Normal	18.7	21.8	13.7	3.1	96
He (Normal)	14.1	20.7	9.1	6.6	98
He (Uniform)	15.1	23.1	10.1	8.0	99

Exercise 5: Using batch normalization

Batch normalization	Train error	Test error	Bias	Variance	Time
Without normalization	17.7	20.8	12.7	3.1	96
Batch normalization (with elu) [after]	11.8	20.2	6.8	8.4	184
Batch normalization (with elu) [before]	14.3	20.5	9.3	6.2	178

Exercise 6: Using regularization

Batch=512

Regularizer	Train error	Test error	Bias	Variance	Time
Without regularization	17.5	22.5	12.5	5.0	100
Regularizer L2 (lambda=0.001)	27.5	25.4	22.5	-2.1	117
Regularizer L2 (lambda=0.0001)	19.2	21.0	14.2	1.8	116
Regularizer L2 (lambda=0.00005)	18.4	21.6	13.4	3.2	116
Regularizer L2 (lambda=0.00001)	18.5	23.0	13.5	4.5	115
Regularizer L1(lambda=0.0001)	21.8	22.3	16.8	0.5	116
Dropout (rate=0.2)	24.1	23.3	19.1	-0.8	108
Dropout (rate=0.1)	22.0	21.8	17.0	-0.2	108
Dropout (rate=0.01)	17.9	19.9	12.9	2.0	109
Dropout (rate=0.001)	17.0	21.3	12.0	4.3	110

Batch=256

Regularizer	Train error	Test error	Bias	Variance	Time
Without regularization	11.4	21.0	6.4	9.6	166
Regularizer L2 (lambda=0.001)	26.3	27.4	21.3	1.1	184
Regularizer L2 (lambda=0.0001)	16.7	19.3	11.7	2.6	187
Regularizer L2 (lambda=0.00001)	13.3	20.7	8.3	7.4	185
Regularizer L1(lambda=0.0001)	20.9	21.0	15.9	0.1	189
Dropout (rate=0.2)	22.2	20.8	17.2	-1.4	178
Dropout (rate=0.1)	20.3	21.3	15.3	1.0	176
Dropout (rate=0.01)	15.7	19.3	10.7	3.6	177
Dropout (rate=0.001)	14.6	20.5	9.6	5.9	177

Exercise 7: Changing learning rate and epochs

Learning rate	Train error	Test error	Bias	Variance	Time
1	11.1	26.5	6.1	15.4	102
0.7	9.8	24.6	4.8	14.8	99
0.5	12.2	23.1	7.2	10.9	105
0.1	17.9	22.0	12.9	4.1	104
0.05	19.9	23.3	14.9	3.4	103
0.01	23.0	23.8	18	0.8	100
0.001	27.9	27.8	22.9	-0.1	97

Epochs	Train error	Test error	Bias	Variance	Time
100	26.1	26.7	21.1	0.6	11
500	22.0	23.0	17	1	50
1000	17.9	22.0	12.9	4.1	104
2000	11.3	21.1	6.3	9.8	212
4000	0.8	21.5	-4.2	20.7	409

Exercise 8: Changing optimizers

Optimizer	Learning rate	Train error	Test error	Bias	Variance	Time
SGD	0.1	17.9	22.0	12.9	4.1	104
Momentum	0.1	6.0	23.6	1.0	17.6	109
Nesterov	0.1	4.9	22.4	-0.1	17.5	108
RSMprop	0.1	66.6	66.9	61.6	0.3	125
RSMprop	0.01	21.7	24.0	16.7	2.3	123
RSMprop	0.001	6.1	23.8	1.1	17.7	128
RSMprop	0.0001	18.4	19.8	13.4	1.4	126
Adam	0.001	4.4	20.7	-0.6	16.3	106
Adam	0.0001	16.7	18.9	11.7	2.2	107

Exercise 9: Multiple changes

L. rate	Epochs	Initializer	Activation	Optimizer	Dropout.	Train error	Test error	Bias	Variance	Time
0.001	1000	He normal	Relu	Adam	0.01	10.0	19.1	5.0	9.1	109
0.0005	1000	He normal	Relu	Adam	0.01	12.7	19.1	7.7	6.4	117
0.001	2000	He normal	Relu	Adam	0.01	6.0	19.4	1.0	13.4	215
0.001	2000	He normal	Relu	Adam	0.05	10.2	18.9	5.2	8.7	216
0.001	3000	He normal	Relu	Adam	0.05	9.3	19.4	4.3	10.1	320
0.001	4000	He normal	Relu	Adam	0.05	8.6	18.9	3.6	10.3	457
0.001	4000	He normal	Relu	Adam	0.07	10.2	18.8	5.2	8.6	424
0.001	4000	He normal	Relu	Adam	0.1	11.7	18.0	6.7	6.3	428
0.001	5000	He normal	Relu	Adam	0.15	13.1	16.3	8.1	3.2	541