

Contents

1	Artificial neural networks	1
1.1	A family of models	1
1.2	Feed forward neural networks	5
1.2.1	Learning with FFNNs	6
1.2.2	Gradient	7
1.3	Recurrent neural networks	10
1.3.1	Learning with RNNs	12
1.3.2	Gradient	13
1.4	Activation functions and gradient	16
1.5	Stochastic gradient descent: a common framework	19
1.6	The vanishing and exploding gradient problem	22
1.7	On expressiveness	27
2	Literature review	31
2.1	Architectural driven methods	32
2.1.1	Long short-term memory	32
2.1.2	Gated recurrent units	34
2.1.3	Structurally constrained recurrent network	35
2.1.4	Gated feedback recurrent neural networks	37
2.2	Learning driven methods	38
2.2.1	Preserve norm by regularization and gradient clipping	38
2.2.2	Hessian-free optimization	39
2.2.3	Reservoir computing	43
2.2.4	Nesterov's accelerated gradient and proper initialization	44
2.2.5	Dropout	45

3	A new SGD approach for training RNNs	49
3.1	Preliminaries	49
3.2	Initialization	51
3.3	Descent direction	53
3.4	Learning rate	55
3.5	Putting all together	56
3.6	Proof of convergence	57
4	Numerical experiments	61
4.1	Experiments on synthetic datasets	61
4.1.1	The effect of the spectral initialization	61
4.1.2	The effect of using the simplex direction	63
4.2	Polyphonic music prediction	66
4.3	Lupus disease prediction	68
5	Conclusion	75
A	Notation	77
B	Details of the synthetic tasks	79
	Bibliography	81