

Redes Neurais Recorrentes

Conteúdo

1	Long Short-Term Memory (LSTM)	2
1.1	Proposta 1	2
1.1.1	Simple (uma camada)	2
1.1.2	Empilhadas (duas camadas)	3
1.1.3	Bidirecionais	5
1.2	Proposta 2	6
1.2.1	Quantidade de dados	6
2	Echo State Network (ESN)	15

1 Long Short-Term Memory (LSTM)

1.1 Proposta 1

Proposta: Aproximar o valor de x_{n+10}

Características do dataset: 1000 sequências de 10 valores (x_{n-9} a x_n) pertencentes a órbita gerada por $r=4$ e $x_0=0.1$. 70% do dataset é usado no treinamento e 30% no teste.

As redes apresentadas usam Adam como otimizador, erro quadrático médio como função de perda e a função de ativação relu.

1.1.1 Simples (uma camada)

n	mse treinamento	mse teste
5.0	0.12752956926822662	0.12976440479358037
10.0	0.12700041792222433	0.12999617666006089
15.0	0.12498688229492733	0.13303570727507275
20.0	0.1242253057871546	0.13372381647427878
25.0	0.12676900156906673	0.1307072361310323
30.0	0.126308528312615	0.1285240704814593
35.0	0.12478361798184259	0.13203097403049469
40.0	0.12712442900453294	0.12959943701823554
45.0	0.1207269572360175	0.13288110514481863
50.0	0.1236609634757042	0.13331147531668344
60.0	0.12094981014728547	0.13296869138876596
70.0	0.12357653311320714	0.1315018755197525
80.0	0.12197189680167607	0.1395740169286728
90.0	0.1236955246755055	0.13072559108336768
100.0	0.1226590205516134	0.13086054543654124
200.0	0.09644957610539028	0.15437853117783865
300.0	0.07831125391381127	0.16667920688788096
400.0	0.05610576182603836	0.1873800132671992
500.0	0.06436575561761856	0.16751160621643066
600.0	0.0649359027828489	0.18402166227499644
700.0	0.0506585342330592	0.2133268755674362
800.0	0.04789594426751137	0.2151570830742518
900.0	0.0772180581518582	0.19286570072174072
1000.0	0.055200338789394926	0.18593117574850718

Tabela 1: Erro quadrático para diferentes números de neurônios

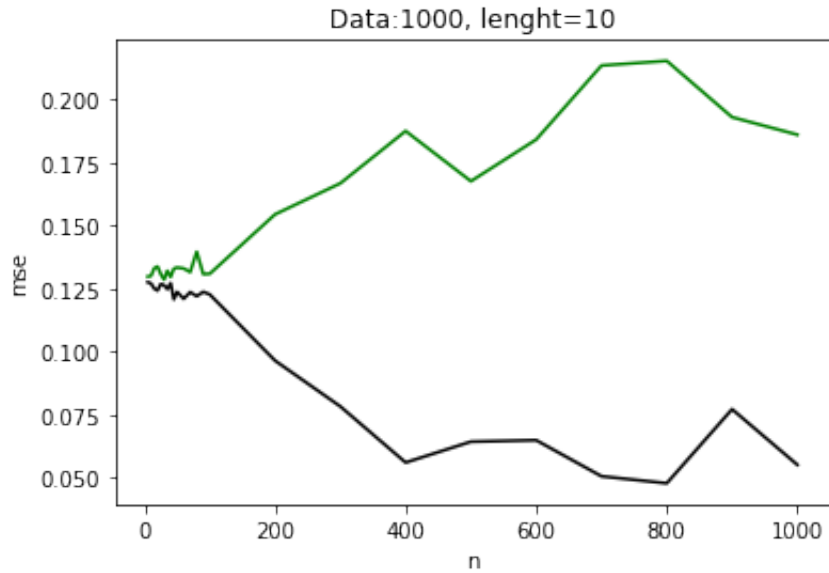


Figura 1: Erro quadrático para diferentes números de neurônios. Treinamento é representado pela linha preta e teste, pela verde

1.1.2 Empilhadas (duas camadas)

n_2	mse treinamento	mse teste
5.0	0.12578769160168513	0.13052331835031508
10.0	0.1266549672824996	0.13046442240476608
15.0	0.12314735680818557	0.13270289242267608
20.0	0.12191820102078574	0.1308850242694219
25.0	0.12193840103490013	0.13366544942061107
30.0	0.1232076757294791	0.13315442820390067
35.0	0.1241009744150298	0.12974269131819408
40.0	0.12496084119592395	0.13730596701304118
45.0	0.12346259163958685	0.13122701744238535
50.0	0.12194925376347134	0.13149945676326752
60.0	0.12231958419084549	0.13203735570112865
70.0	0.12495423951319286	0.13549541294574738
80.0	0.12511790411812918	0.13199622948964437
90.0	0.12225966027804784	0.13527747491995493
100.0	0.12469103429998671	0.1330508969227473
200.0	0.11988218814134598	0.13309228519598643
300.0	0.11750247171946934	0.14030180553595226
400.0	0.12287631550005504	0.12896898488203684
500.0	0.10922849433762687	0.14592193524042765
600.0	0.12133346651281629	0.1330738455057144
700.0	0.12296523711511068	0.1346294246117274
800.0	0.09259222094501768	0.16388486683368683
900.0	0.10387803609882082	0.14979551215966544
1000.0	0.11028124336685453	0.15249838491280873

Tabela 2: Erro quadrático para diferentes números de neurônios na segunda "camada"(com 30 neurônios na primeira)

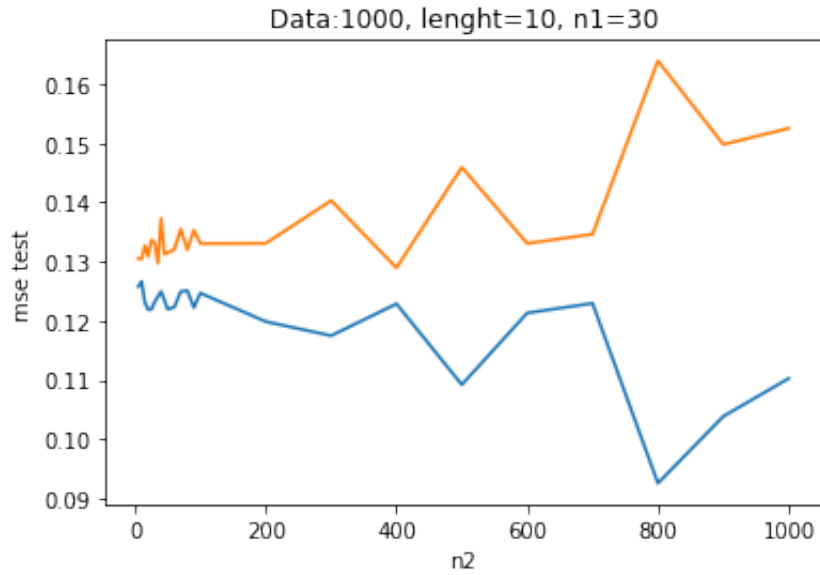


Figura 2: Erro quadrático para $n_1=30$. Treinamento é representado pela linha azul e teste pela laranja

n_1	mse treinamento	mse teste
5.0	0.12409425961119788	0.13052495062351227
10.0	0.11805936481271471	0.13953982690970104
15.0	0.11951923736504146	0.1352690060933431
20.0	0.11864756788526262	0.13356782535711925
25.0	0.12416058097566877	0.13430430014928182
30.0	0.11072808299745832	0.15191193521022797
35.0	0.1111688134925706	0.14355381667613984
40.0	0.11380587539502553	0.1457968274752299
45.0	0.10593291499785014	0.1427371456225713
50.0	0.11962868588311332	0.1352065465847651
60.0	0.11279875644615718	0.13855741441249847
70.0	0.10123310361589705	0.14760475873947143
80.0	0.10130848526954651	0.1605283941825231
90.0	0.09294509764228548	0.16244505882263183
100.0	0.10219578917537417	0.1431446240345637
200.0	0.07894093602895737	0.18927446266015371
300.0	0.057171483572040285	0.18798593600591024
400.0	0.0742920930470739	0.1676635464032491
500.0	0.051204987104449956	0.20422551155090332
600.0	0.03639684151325907	0.2173178166151047
700.0	0.051894556752273016	0.1971631807088852
800.0	0.042095037634883606	0.21776255190372468
900.0	0.04267935297318867	0.2044983277718226
1000.0	0.05629267347710473	0.19369044880072275

Tabela 3: Erro quadrático para diferentes números de neurônios na primeira "camada"(com 400 neurônios na segunda)

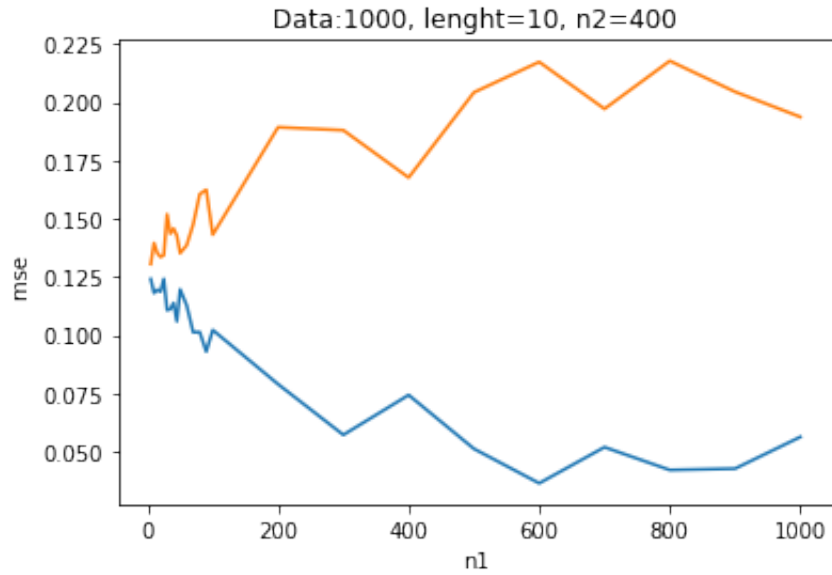


Figura 3: Erro quadrático para $n_2=400$. Treinamento é representado pela linha azul e teste pela laranja

1.1.3 Bidirecionais

n	mse treinamento	mse teste
5.0	0.12711727235998427	0.12850983788569767
10.0	0.12718840373413903	0.12936487992604573
15.0	0.126139215358666	0.13103445500135422
20.0	0.1246476262807846	0.13308628062407177
25.0	0.12635477815355572	0.12866477141777674
30.0	0.12508738292115076	0.13232315748929976
35.0	0.12343874480043139	0.1302104044953982
40.0	0.12308916194098336	0.13026496589183809
45.0	0.12132145132337298	0.13244426588217417
50.0	0.12201906876904624	0.13092372020085652
60.0	0.12231770098209381	0.14390106737613678
70.0	0.11769472356353487	0.1329864635070165
80.0	0.11210337221622467	0.1464717358350754
90.0	0.11068495448146548	0.13575749218463898
100.0	0.12089557647705078	0.13804165263970694
200.0	0.06397577328341347	0.18042319476604463
300.0	0.057051175151552475	0.1949439255396525
400.0	0.04939404809049198	0.2008100938796997
500.0	0.04670631040419851	0.19996937215328217
600.0	0.023117578189287866	0.22445662458737692
700.0	0.023514054790139197	0.20906487256288528
800.0	0.02845611168869904	0.21127357304096223
900.0	0.03580752161996705	0.20502053697903952
1000.0	0.02344024583697319	0.22923269927501677

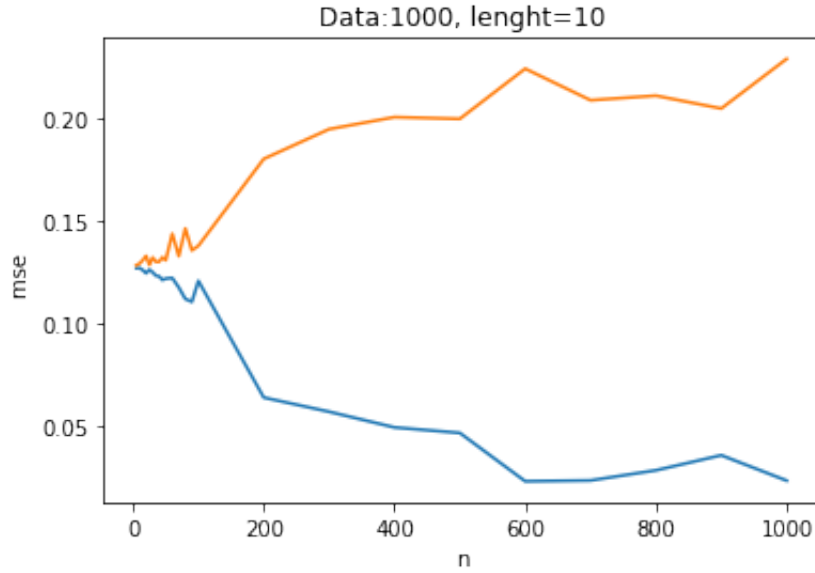


Figura 4: Erro quadrático para diferentes números de neurônios. Treinamento é representado pela linha azul e teste pela laranja

1.2 Proposta 2

Proposta: Aproximar o valor dos 10 valores futuros (x_{n+1} a x_{n+10}) para diferentes quantidades de dados

Características do dataset: sequências de 10 valores (x_{n-9} a x_n) pertencentes a órbita gerada por $r=4$ e $x_0=0.1$. 70% do dataset é usado no treinamento e 30% no teste.

As redes apresentadas têm uma camada, usam Adam como otimizador, erro quadrático médio como função de perda e a função de ativação relu.

1.2.1 Quantidade de dados

Os gráficos abaixo mostram o erro quadrático médio em função do número de neurônios para diferentes quantidades de dados. Os datasets usados tem 100, 1000 e 5000 sequências de dez números, respectivamente.

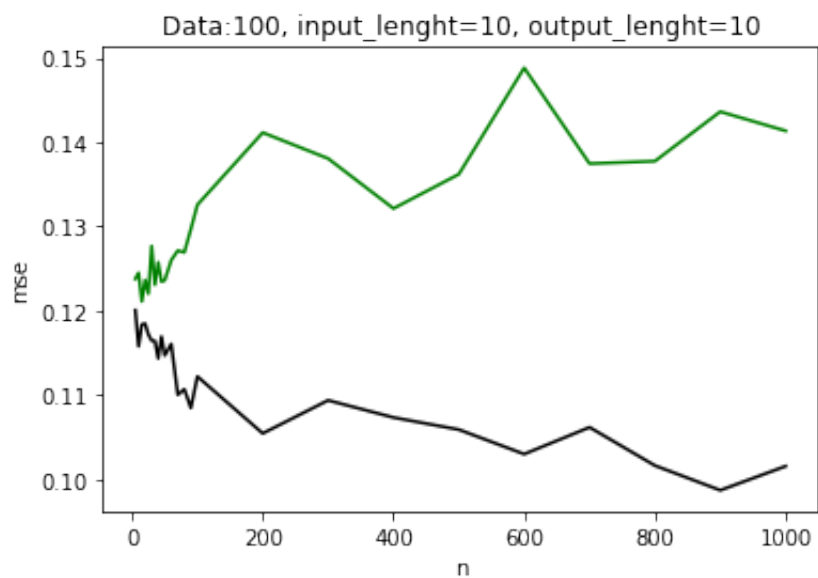


Figura 5: Erro quadrático para diferentes números de neurônios para 100 dados de entrada. Treinamento é representado pela linha preta e teste, pela verde

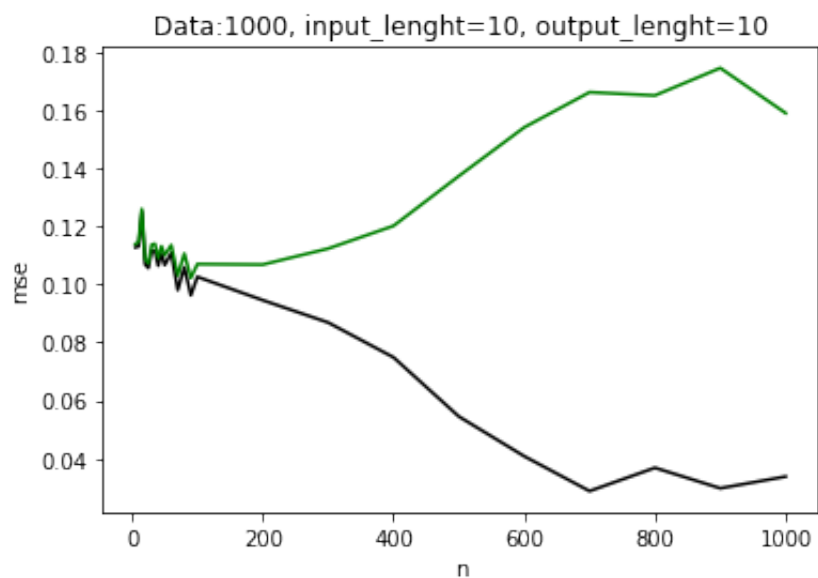


Figura 6: Erro quadrático para diferentes números de neurônios para 1000 dados de entrada. Treinamento é representado pela linha preta e teste, pela verde

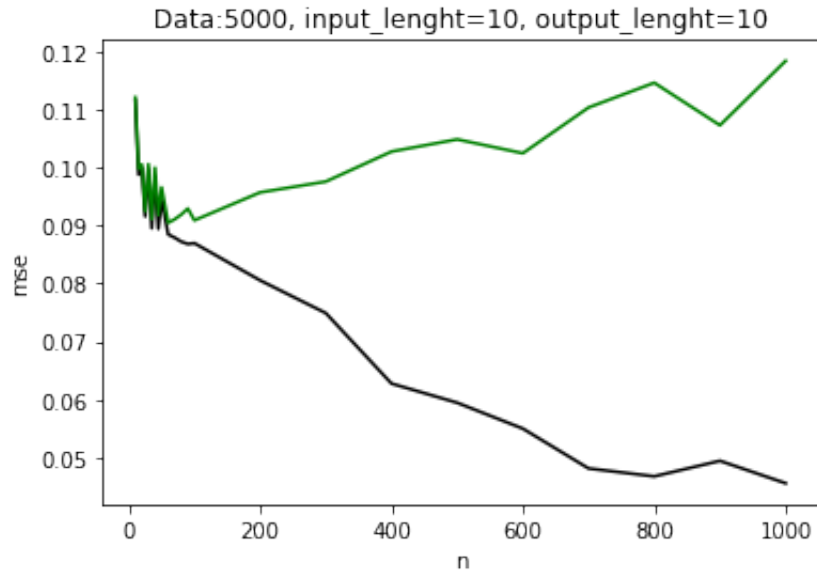


Figura 7: Erro quadrático para diferentes números de neurônios para 5000 dados de entrada. Treinamento é representado pela linha preta e teste, pela verde

Abaixo, o erro quadrático médio mínimo obtido por cada um dos datasets.

n input	mse
100	0.09045861671368281
1000	0.10217117955287297
5000	0.12108870595693588

Tabela 4: Erro quadrático mínimo obtido para diferentes quantidades de sequências de entradas

Observando apenas esses dados, cria-se a impressão que aumentar o número de dados de entrada não contribui com a melhora do erro. No entanto, quando essas redes predizem o futuro de uma entrada (de teste) específica, notamos outras características. As próximas figuras mostram algumas delas (com diferentes números de neurônios) predizendo o futuro de uma das sequências do dataset de teste treinada com dataset de tamanho 100:

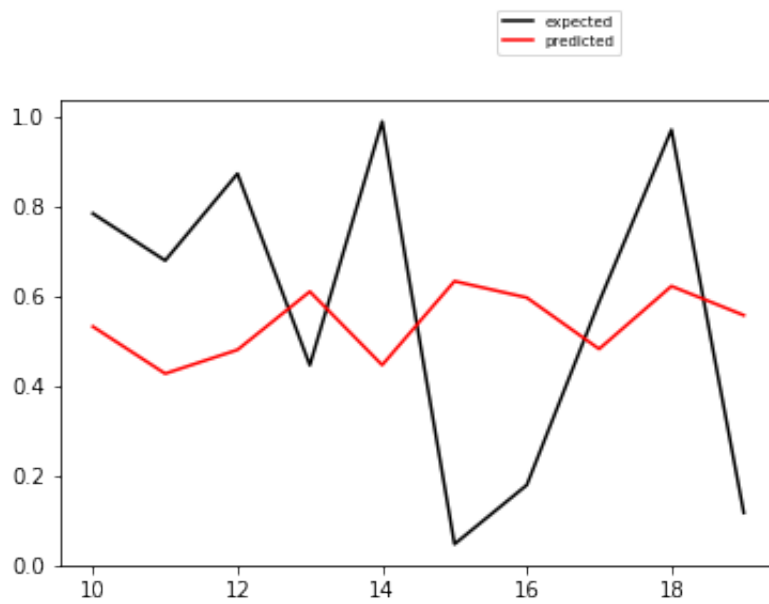


Figura 8: Futuro esperado em preto, previsto pela rede em vermelho. LSTM com 30 neurônios

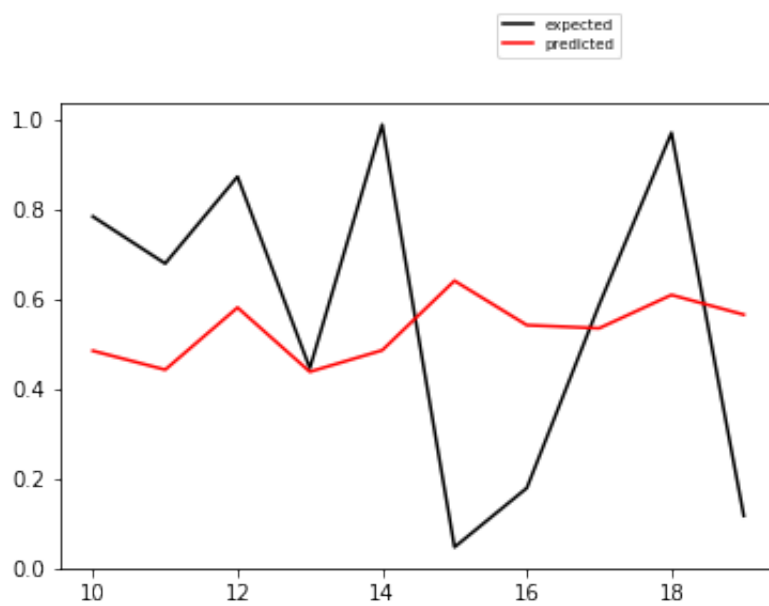


Figura 9: Futuro esperado em preto, previsto pela rede em vermelho. LSTM com 400 neurônios

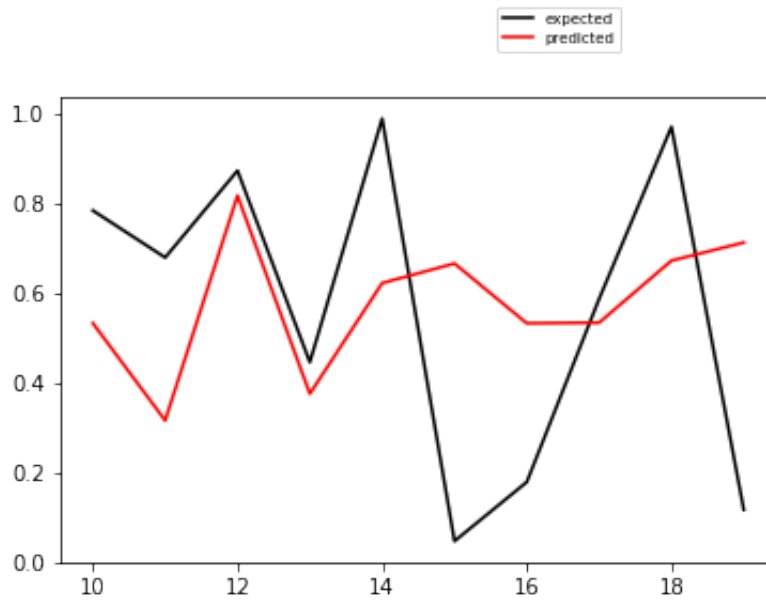


Figura 10: Futuro esperado em preto, previsto pela rede em vermelho. LSTM com 1000 neurônios

Agora analisemos as redes (com diferentes números de neurônios) predizendo o futuro de uma das sequências do dataset de teste treinada com o dataset de tamanho 5000.

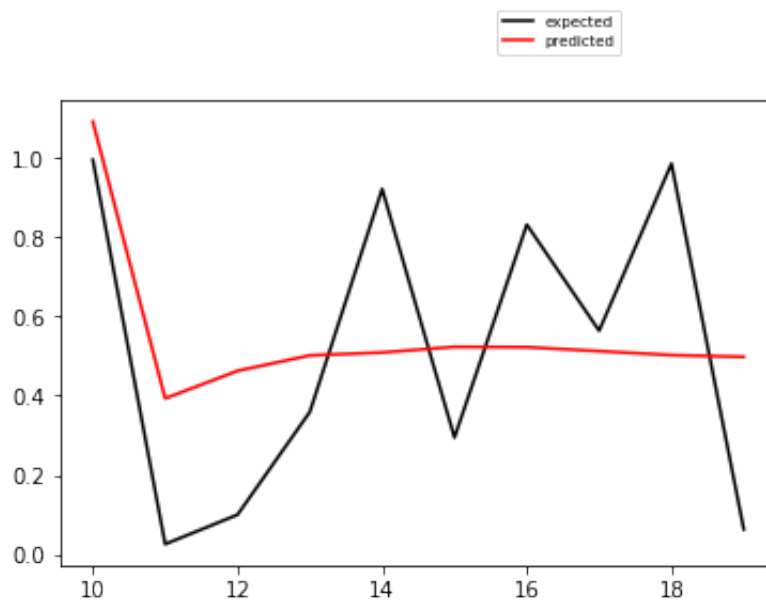


Figura 11: Futuro esperado em preto, previsto pela rede em vermelho. LSTM com 10 neurônios

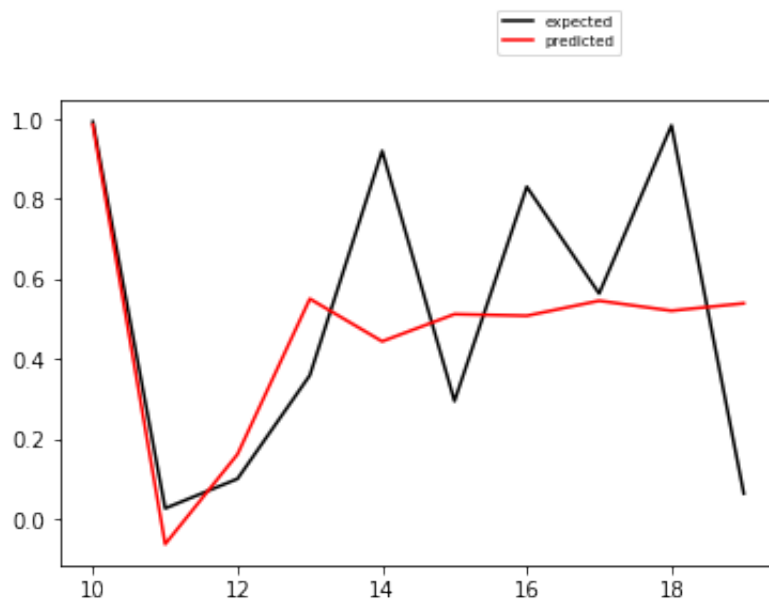


Figura 12: Futuro esperado em preto, previsto pela rede em vermelho. LSTM com 30 neurônios

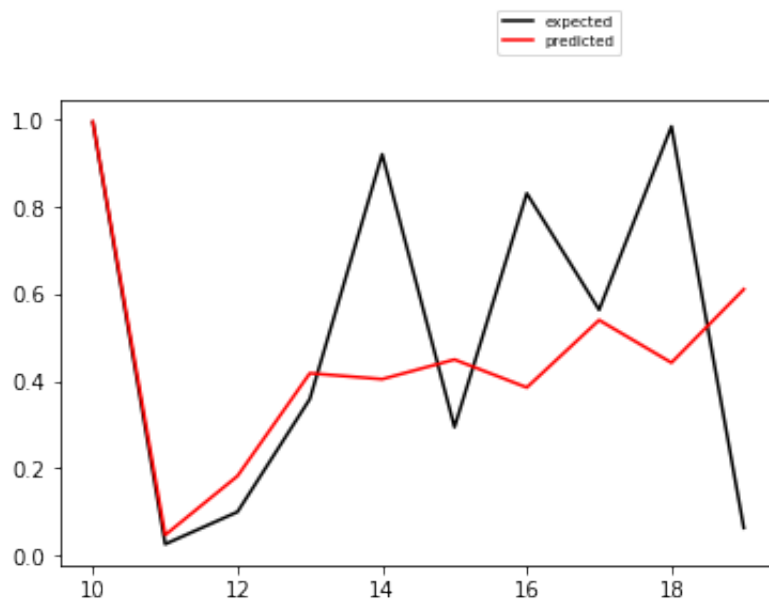


Figura 13: Futuro esperado em preto, previsto pela rede em vermelho. LSTM com 300 neurônios

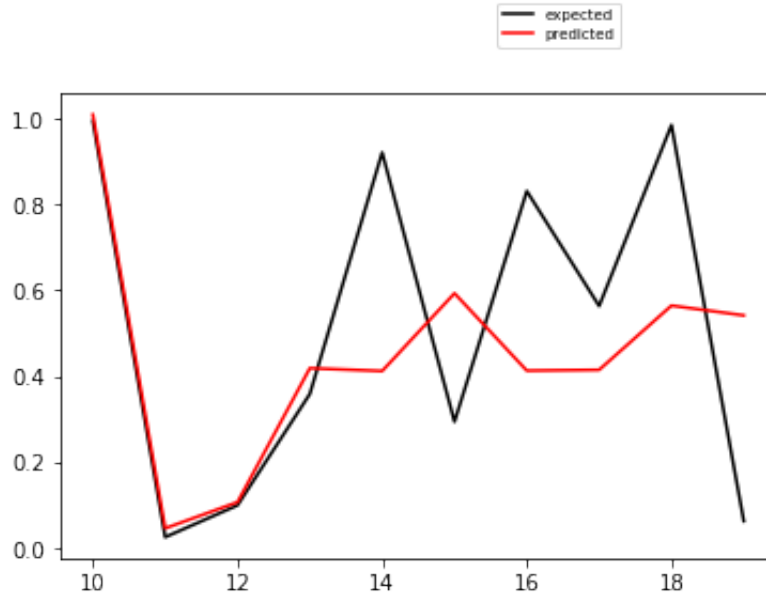


Figura 14: Futuro esperado em preto, previsto pela rede em vermelho. LSTM com 1000 neurônios

Temos a impressão que, no caso de mais dados de entrada, a rede toma mais riscos (não prevê apenas valores próximos a 0.5), e para os primeiros valores previstos (x_{n+1} até x_{n+4} ela parece fazer uma aproximação melhor do que no caso de um dataset menor. Para analisar essa hipótese melhor, calculamos o erro quadrático médio de todo dataset teste, mas tomando apenas um ponto futuro por vez:

futuro	x_{n+1}	x_{n+2}	x_{n+3}	x_{n+4}	x_{n+5}
mse	0.10941787	0.12989859	0.10176053	0.14183318	0.117045727
futuro	x_{n+6}	x_{n+7}	x_{n+8}	x_{n+9}	x_{n+10}
mse	0.12023696	0.13535324	0.12331151	0.13099482	0.11983481

Tabela 5: mse de cada ponto futuro para a rede treinada com o dataset de 100 sequências e com 15 neurônios

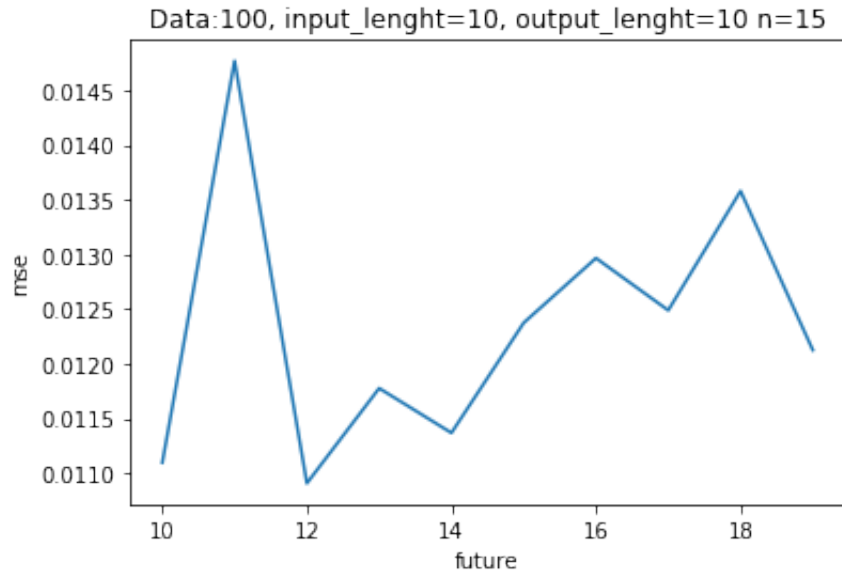


Figura 15: mse de cada ponto futuro para a rede treinada com o dataset de 100 sequências e com 15 neurônios

futuro	X_{n+1}	X_{n+2}	X_{n+3}	X_{n+4}	X_{n+5}
mse	0.13302645	0.12633655	0.11965824	0.15809337	0.1330427
futuro	X_{n+6}	X_{n+7}	X_{n+8}	X_{n+9}	X_{n+10}
mse	0.11444852 0.15798001	0.1298909	0.19976392	0.18928267	

Tabela 6: mse de cada ponto futuro para a rede treinada com o dataset de 100 sequências e com 400 neurônios

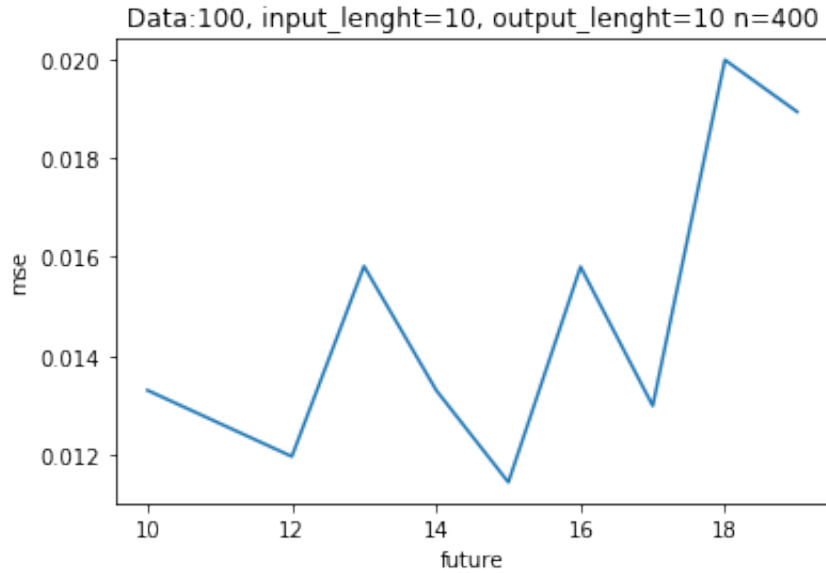


Figura 16: mse de cada ponto futuro para a rede treinada com o dataset de 100 sequências e com 400 neurônios

futuro	X_{n+1}	X_{n+2}	X_{n+3}	X_{n+4}	X_{n+5}
mse	0.00166003	0.01130036	0.12529952	0.12850603	0.12006436
futuro	X_{n+6}	X_{n+7}	X_{n+8}	X_{n+9}	X_{n+10}
mse	0.12482693	0.12574728	0.12424491	0.12733529	0.12463401

Tabela 7: mse de cada ponto futuro para a rede treinada com o dataset de 5000 sequências e com 30 neurônios

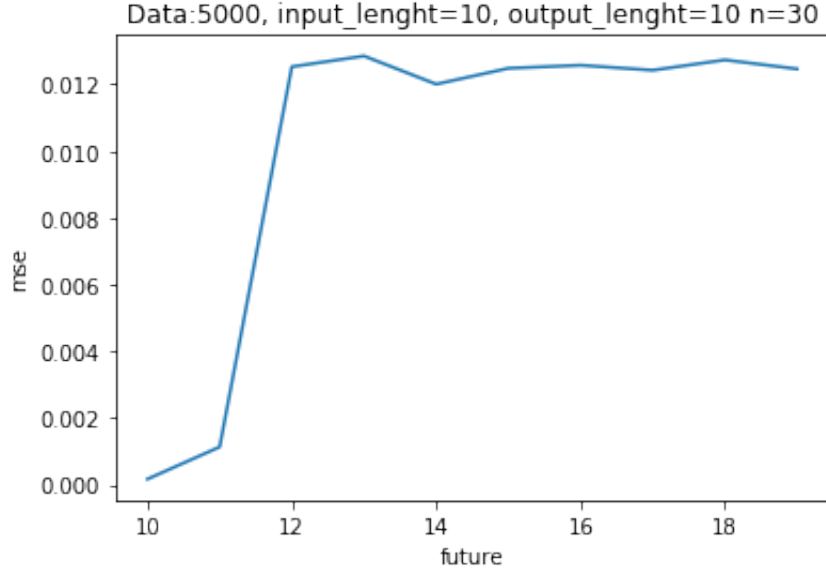


Figura 17: mse de cada ponto futuro para a rede treinada com o dataset de 5000 sequências e com 30 neurônios

futuro	X_{n+1}	X_{n+2}	X_{n+3}	X_{n+4}	X_{n+5}
mse	0.00048422	0.00201185	0.02190834	0.10124728	0.18111158
futuro	X_{n+6}	X_{n+7}	X_{n+8}	X_{n+9}	X_{n+10}
mse	0.18738036	0.19312233	0.18957402	0.19226032	0.18134764

Tabela 8: mse de cada ponto futuro para a rede treinada com o dataset de 5000 sequências e com 1000 neurônios

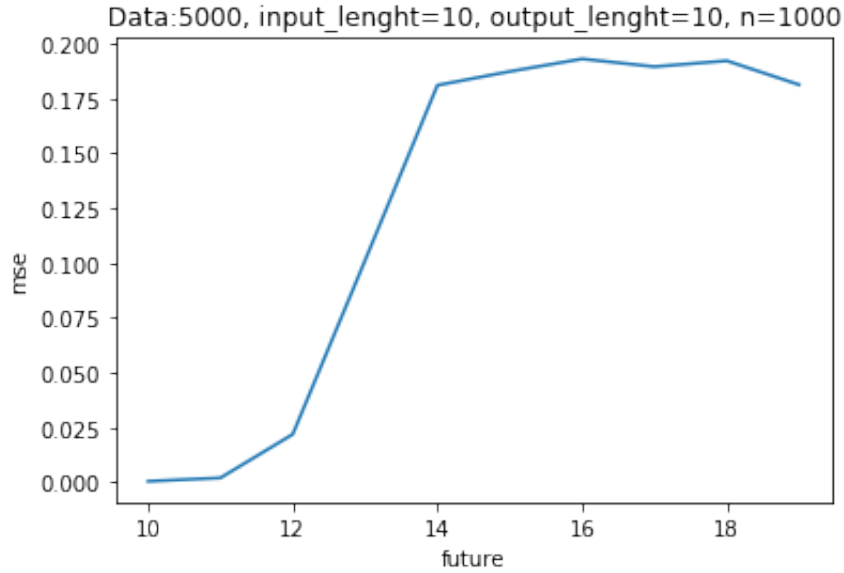


Figura 18: mse de cada ponto futuro para a rede treinada com o dataset de 5000 sequências e com 1000 neurônios

2 Echo State Network (ESN)

Proposta: Aproximar o valor dos 10 valores futuros (x_{n+1} a x_{n+10}) para diferentes quantidades de sequência de entrada

Características do dataset: sequências de M valores consecutivos ($x_{n-(M-1)}$ a x_n) pertencentes a órbita gerada por $r=4$ e $x_0=0.1$.

As tabelas e gráficos seguintes mostram os erros quadráticos mínimos obtidos por um número de neurônios no reservatório (n) variando-se o raio espectral(sr). O raio foi variado de 0.001 a 0.2. Foi usado o módulo pyESN (<https://github.com/cknd/pyESN>).

n	sr	mse	n	sr	mse
100	0.03	0.43398252599522941	3600	0.001	0.320423360991
200	0.08	0.11862201405224304	3700	0.16	0.328333542096
300	0.002	0.0022983421712321843	3800	0.06	0.0971327664837
400	0.005	0.094408261923561582	3900	0.09	0.329997877921
500	0.02	0.32836104880046713	4000	0.001	0.109783353718
600	0.15	0.002298594899986068	4100	0.001	0.121651620244
700	0.16	0.107852317984	4200	0.03	0.165122900229
800	0.19	0.265486310656	4300	0.005	0.328323558314
900	0.04	0.0954280050007	4400	0.001	0.192503561444
1000	0.001	0.235041117229	4500	0.009	0.0945803059117
1100	0.001	0.103175127698	4600	0.04	0.328314798544
1200	0.001	0.343481520139	4700	0.001	0.182507365728
1300	0.14	0.00766885759314	4800	0.02	0.095204062473
1400	0.003	0.0948143810334	4900	0.03	0.124116318189
1500	0.06	0.105813308913	5000	0.05	0.196197806083
1600	0.15	0.329277710032	5100	0.15	0.0948389273316
1700	0.07	0.102816210174	5200	0.004	0.094944975595
1800	0.001	0.150698976229	5300	0.001	0.128193565729
1900	0.001	0.198972889254	5400	0.04	0.0954393599923
2000	0.001	0.0989632950258	5500	0.06	0.0040416733294
2100	0.06	0.328343017522	5600	0.001	0.122860527267
2200	0.04	0.0949026385854	5700	0.001	0.0948082665117
2300	0.002	0.180065962436	5800	0.001	0.269915619211
2400	0.08	0.0948721592754	5900	0.04	0.185085465309
2500	0.12	0.362443437285	6000	0.002	0.0958586501495
2600	0.08	0.0979412843074	6100	0.17	0.0949004599736
2700	0.05	0.328163985196	6200	0.001	0.150429632212
2800	0.001	0.243351940076	6300	0.06	0.0962559057886
2900	0.007	0.0957251326833	6400	0.07	0.10010537505
3000	0.001	0.137452702678	6500	0.001	0.186479021
3100	0.002	0.095218761802	6600	0.001	0.228261392
3200	0.04	0.0984885354158	6700	0.001	0.185011402
3300	0.001	0.281044213905	6800	0.001	0.103611742
3400	0.08	0.0961234760552	6900	0.001	0.198722130394
3500	0.001	0.169121370285	7000	0.001	0.0670561491856

Tabela 9: mse para diferentes números de neurônios com uma sequência de entrada com 100 valores

n	sr	mse	n	sr	mse
100	0.071	0.2843035445129255	5100	0.181	0.079767343223108023
200	0.011	0.20524295576885257	5200	0.071	0.12168035646934899
300	0.181	0.33509755018580206	5300	0.006	0.10278476377192879
400	0.001	0.10314343485420328	5400	0.001	0.12385725293958237
500	0.071	0.10252921257538548	5500	0.16	0.0087862963337808162
600	0.001	0.3805048125142394	5600	0.16	0.048080690117555634
700	0.071	0.26347268232111715	5700	0.06	0.0015706908365577315
800	0.02	0.043377632795039638	5800	0.001	0.089590647443319016
900	0.001	0.11172738168912477	5900	0.001	0.15808642033168091
1000	0.131	0.23880510115369746	6000	0.181	0.25481758332143589
1100	0.131	0.25108595088352509	6100	0.001	0.089021631874496895
1200	0.001	0.22382242816422143	6200	0.16	0.040056023898263279
1300	0.006	0.1136229405357387	6300	0.011	0.22419819605568911
1400	0.007	0.025266133287820319	6400	0.171	0.11524825047180777
1500	0.071	0.23914467622762625	6500	0.02	0.0040496368491188895
1600	0.181	0.053973794917799849	6600	0.001	0.087188870169552582
1700	0.006	0.23654051200025761	6700	0.15	0.001127061122794575
1800	0.011	0.15900727697121106	6800	0.011	0.11926607084230825
1900	0.011	0.23279700560779834	6900	0.171	0.11947077898209776
2000	0.011	0.30440797291271182	7000	0.001	0.12674626195437655
2100	0.181	0.0015899706843913472	7100	0.171	0.14167768148759721
2200	0.181	0.020650684612472096	7200	0.071	0.12022511206289543
2300	0.011	0.2139086586349804	7300	0.001	0.061462818214048802
2400	0.181	0.15890277745005035	7400	0.03	0.040097734897689079
2500	0.071	0.12086343464038098	7500	0.171	0.12281637754450518
2600	0.006	0.31877206496400773	7600	0.02	0.013967567657654729
2700	0.151	0.017313264752816953	7700	0.001	0.002555691822930346
2800	0.011	0.12206284200393298	7800	0.171	0.074708964298869876
2900	0.151	0.067803861765779017	7900	0.171	0.11104950731528526
3000	0.006	0.32441381668197583	8000	0.10	0.0084645258665232009
3100	0.181	0.12671829517382149	8100	0.121	0.13283757989409961
3200	0.16	0.027025311532532424	8200	0.011	0.084996586778774558
3300	0.002	0.00078616607991842749	8300	0.121	0.10968750333459008
3400	0.001	0.12119529223681641	8400	0.171	0.12188193904242708
3500	0.006	0.10543870895455151	8500	0.171	0.056148302168541775
3600	0.071	0.10316900776752307	8600	0.121	0.18579514214374371
3700	0.16	0.071266018071877754	8700	0.011	0.099295472942046228
3800	0.181	0.10216392928335417	8800	0.006	0.11805801813120097
3900	0.181	0.089870917895201782	8900	0.171	0.12202542711688281
4000	0.131	0.10571886613053032	9000	0.02	0.00051100032034565317
4100	0.151	0.021386612881819454	9100	0.006	0.12171419509275468
4200	0.16	0.034293828881393734	9200	0.171	0.052850406581879596
4300	0.181	0.08203686828546225	9300	0.171	0.085937267711215834
4400	0.001	0.15816396362927598	9400	0.011	0.071883265001436311
4500	0.001	0.005528742628383947	9500	0.001	0.091246818387240178
4600	0.001	0.10364066959244558	9600	0.03	0.034837951023912496
4700	0.001	0.09354525485073209	9700	0.001	0.093614545322647486
4800	0.181	0.12211755916751853	9800	0.011	0.09934379141591565
4900	0.05	0.00083935855960130118	9900	0.171	0.0094482750076326366
5000	0.006	0.23658536244604969	17		

Tabela 10: erros com uma sequência de entrada de 1000 valores

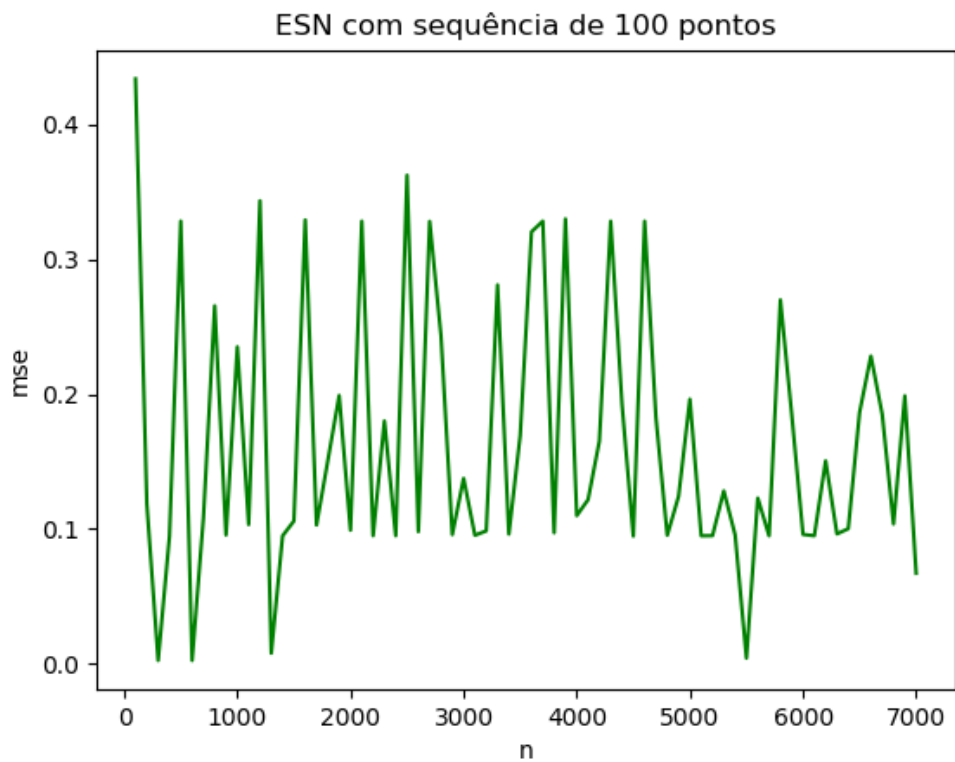


Figura 19: mse para diferentes números de neurônios com uma sequência de entrada com 100 valores

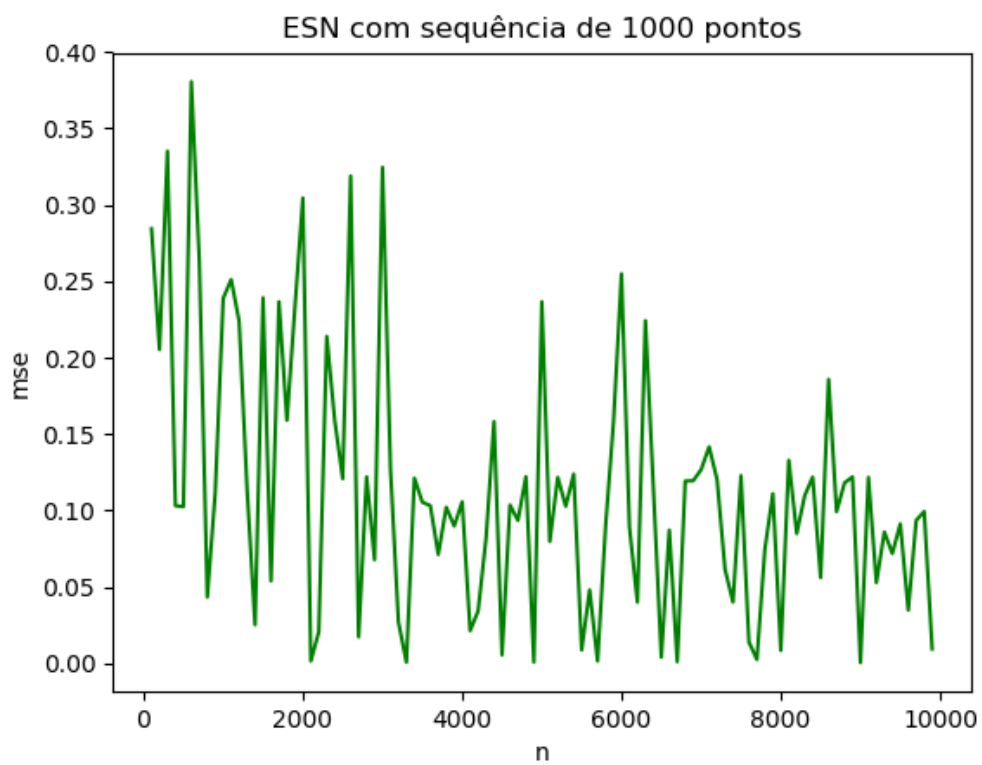


Figura 20: mse para diferentes números de neurônios com uma sequência de entrada com 1000 valores

Abaixo, algumas previsões feitas para 100 dados e $n=600$

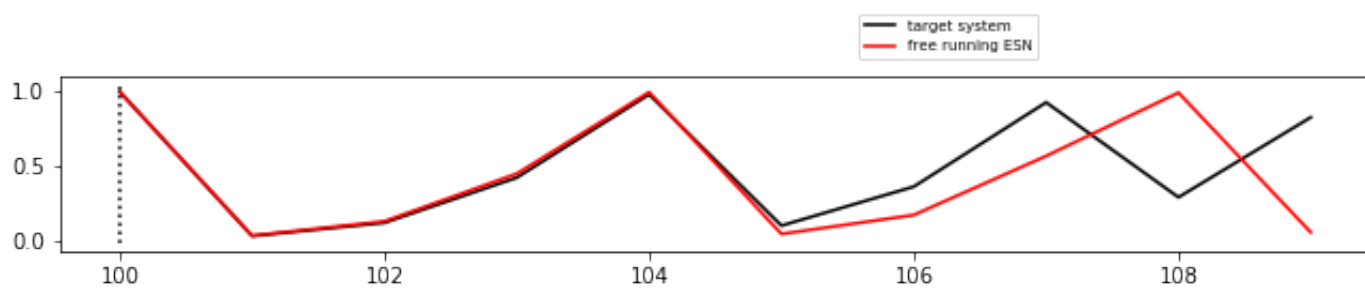


Figura 21: previsão de 10 pontos para $sr=0.01$

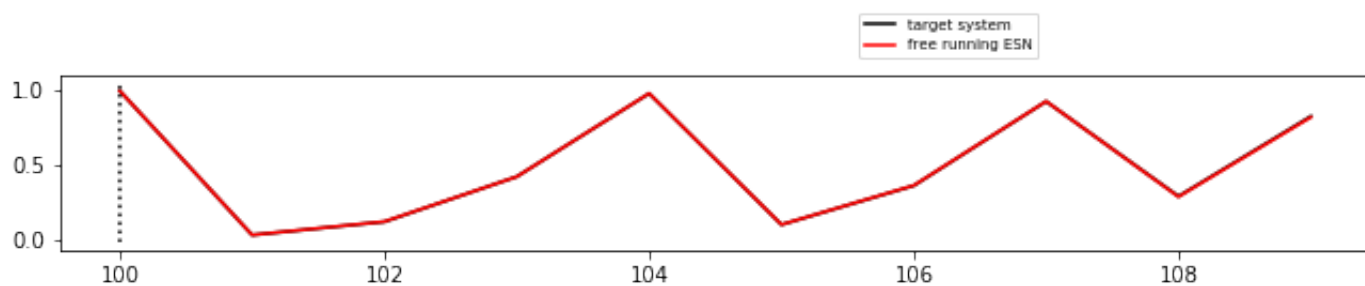


Figura 22: previsão de 10 pontos para $sr=0.15$

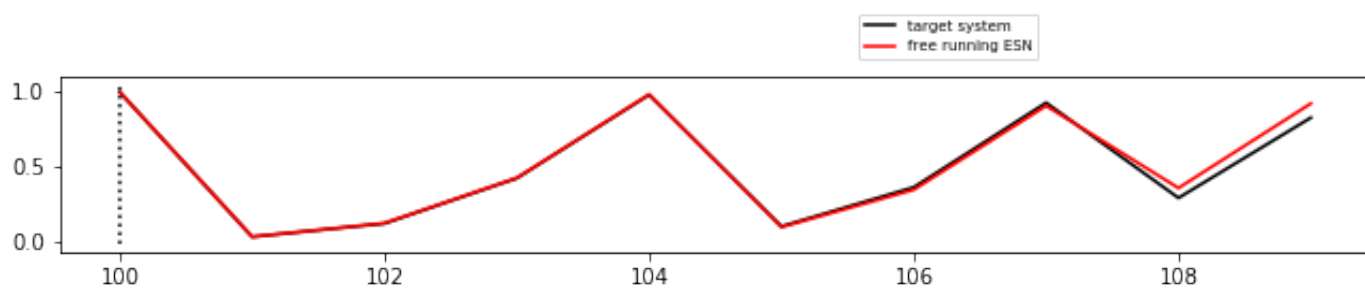


Figura 23: previsão de 10 pontos para $sr=0.2$