

Assignment 3: Transition Parsing with Neural Networks

Niranjana Balsubramanian and Jun Kang
CSE 538 Fall 2018

Name: Rohit Bhal

Solar ID: 112073893

Email: rbhal@cs.stonybrook.edu

Experiment Results:

Single Layer:

max_iter = 1001, batch_size = 10000, hidden_size = 200, embedding_size = 50, learning_rate = 0.1, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8

	Loss	UAS	UASnoP unc	LAS	LASnoP unc	UEM	UEMnoP unc	Root
Cubic	0.35859 606951 47514	61.7194 705486	65.2970 10117	56.7988 633248	60.1141 694455	4.94117 647059	5.58823 529412	37.11764 70588
Tanh	0.56102 603167 29545	59.7527 232844	63.0673 147572	53.6455 86659	56.7145 198666	4.47058 823529	4.58823 529412	39.764705 4
RELU	0.56343 606203 79448	63.4344 542214	66.6760 865879	57.6339 207817	60.4730 684451	5.64705 882353	5.94117 647059	50.7058 823529
Sigmoid	1.20552 554130 5542	42.3162 250418	45.9786 356186	30.6528 404417	33.8184 592777	1.47058 823529	1.47058 823529	8.47058 823529

Two Layers:

max_iter = 1001, batch_size = 10000, hidden_size = 200, embedding_size = 50, learning_rate = 0.1, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8

	Loss	UAS	UASnoP unc	LAS	LASnoP unc	UEM	UEMno Punc	Root
Cubic Cubic	0.40016 873568 296435	67.9686 915771	71.1156 954728	63.4743 375626	66.4217 487142	8.11764 705882	8.58823 529412	51.2941 176471
Cubic Tanh	0.35546 691298 4848	69.6612 408705	72.7773 695812	65.2665 95209	68.0099 474368	9.52941 176471	10.0	55.4117 647059
Tanh	0.44701	65.2491	68.5044	59.9745	62.8864	7.11764	7.41176	48.1176

Tanh	568156 480787	462472	933024	7437	522693	705882	470588	470588
Cubic Relu	0.33594 674438 23814	68.4772 041778	71.6074 15362	64.2969 314754	67.2328 039338	8.47058 823529	8.82352 941176	51.2352 941176

Three Layers:

max_iter = 1001, batch_size = 10000, hidden_size = 200, embedding_size = 50, learning_rate = 0.1, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8

	Loss	UAS	UASnoP unc	LAS	LASnoP unc	UEM	UEMno Punc	Root
Cubic Cubic Cubic	0.80969 228446 48361	55.9712 839943	59.5320 183123	47.7154 323603	50.5397 614876	2.58823 529412	2.88235 294118	30.8823 529412
Cubic Cubic Tanh	0.34698 296636 343	73.1884 238602	75.9424 63121	68.7090 260987	71.0309 161815	11.9411 764706	12.9411 764706	64.7058 823529
Tanh Tanh Tanh	0.38287 674725 05569	70.6433 681482	73.5432 091788	66.0418 276541	68.5553 608772	9.76470 588235	10.5294 117647	58.2352 941176
Cubic Relu	0.36913 743078 70865	69.0081 511579	72.1810 885661	64.5561 731934	67.4419 261855	8.41176 470588	9.29411 764706	52.8235 294118

Single Layer Tanh:

max_iter = 1001, batch_size = 10000, **hidden_size = 3000**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=4

learnin g_rate	Loss	UAS	UASnoP unc	LAS	LASnoP unc	UEM	UEMno Punc	Root
0.05	0.40043 379992 24663	67.3928 758382	70.4007 2345	63.1178 802004	65.9158 989431	8.41176 470588	8.76470 588235	52.0
0.1	0.31074 268221 85516	70.6658 025276	73.8314 58769	67.2607 622704	70.1379 076471	9.35294 117647	10.4705 882353	55.3529 411765
0.2	0.25863 624721 765516	75.0130 867213	77.7256 542135	72.0542 413441	74.4305 657605	14.0	14.8823 529412	64.0588 235294

0.3	0.23530 183091 759682	76.6208 839145	79.1979 879048	73.7118 927138	75.9424 63121	15.3529 411765	16.7058 823529	67.7647 058824
0.5	0.21607 981026 172637	80.1555 450308	82.0522 240434	77.3412 767655	78.8956 084327	21.1176 470588	22.5294 117647	78.9411 764706

Single Layer Tanh:

batch_size = 10000, **hidden_size = 3000**, embedding_size = 50, display_step = 100,
validation_step = 1000, n_Tokens = 48, lam = 1e-8, **learning_rate=0.5**,func=4

max_iter	Loss	UAS	UASnoP unc	LAS	LASnoP unc	UEM	UEMno Punc	Root
1000	0.21607 981026 172637	80.1555 450308	82.0522 240434	77.3412 767655	78.8956 084327	21.1176 470588	22.5294 117647	78.9411 764706
5000	0.14106 894508 00419	86.3249 993768	87.9924 263833	83.9270 134856	85.2823 150398	30.8235 294118	33.0588 235294	87.5882 352941
10000	0.11274 907030 165196	88.0973 153526	89.5156 276494	85.8439 065733	86.9637 709829	34.2352 941176	36.9411 764706	89.5294 117647
15000	0.09598 513782 024383	87.8630 007229	89.3601 989487	85.6719 096642	86.8789 916916	33.3529 411765	36.0	88.4117 647059
20000	0.08233 070775 866508	88.8700 550889	90.2786 412706	86.6166 463095	87.7239 586277	37.2352 941176	40.0588 235294	91.0588 235294

Analysis:

From the experiment, I have observed the following:

1. Increasing the number of hidden layer size for RELU & Tanh function with single layer forward pass significantly improved their performance on the dev set.
2. Keeping the hidden layer size around 200 makes Cubic function as the best function for single layer forward pass and is significantly higher than other function models.
3. Increasing the number of iterations upto 20000 improves the performance of the best model approximately by 10% although significant improvement is done within 10,000 iterations.
4. Increasing the number of hidden layer from 1,2 to 3 doesn't improve the performance of the model for the cubic function whereas it is improved for tanh or RELU function model.
5. Increasing the hidden layer size upto 1000 increases the cubic function model performance whereas Tanh & RELU have significant performance improvement in accuracy with hidden size layer = 3000. Tanh beats the best cubic function model for a single layer forward pass with high learning rate(0.5) & hidden size layer(3000).
6. Without clip gradient, the best model performance has deteriorated significantly approx. by 7% comparing to the one where gradient clipping is used.
7. For small number of hidden layer size, cubic functions beats all other function whereas for high hidden layer size, tanh beats all other functions.
8. Increasing learning rate beyond 0.3 for cubic function decreases the accuracy of the model whereas RELU & Tanh models are significantly improved at high learning rate evident by the performance of best model at learning rate = 0.5
9. No Significant improvement in accuracy is seen in sigmoid function.

Best Model:

Configuration:

max_iter = 20001
batch_size = 10000
hidden_size = 3000
embedding_size = 50
learning_rate = 0.5
display_step = 100
validation_step = 1000
n_Tokens = 48
lam = 1e-8
func=4

Average loss at step 20000 : 0.08233070775866508

Testing on dev set at step 20000

UAS: 88.8700550889

UASnoPunc: 90.2786412706

LAS: 86.6166463095

LASnoPunc: 87.7239586277

UEM: 37.2352941176

UEMnoPunc: 40.0588235294

ROOT: 91.0588235294

Without Gradient Clipping:

Average loss at step **1000** : 311.92092422485354

UAS: 21.3600219358,UASnoPunc: 23.695811903,LAS: 3.88114764314,LASnoPunc:

4.26439834963,UEM: 1.05882352941,UEMnoPunc: 1.05882352941,ROOT: 2.76470588235

Average loss at step **20000** : 0.7964006686210632

UAS: 85.4949273375

UASnoPunc: 86.9920307466

LAS: 78.871799985

LASnoPunc: 79.4862374951

UEM: 29.1764705882

UEMnoPunc: 31.2352941176

ROOT: 89.2941176471

Gradient Clipping:

In neural networks, when gradients are being propagated back in previous layer, they can vanish if they are being multiplied by number less than one or they can grow to extremely large value causing an overflow (Nan) when multiplied by number greater than one. The overflow problem can be easily detectable during runtime/training. The model starts overshooting past the minima. If the gradient becomes large, it is called Gradient Explosion Problem. Similarly, for gradient becoming too small it is called vanishing gradient problem.

Gradient Clipping will clip the gradient or cap them to a certain threshold value to prevent the gradient getting too large or small. It will clip the gradients between two numbers to prevent them from getting too large.

Additional Experimentation Details:

Single Layer:

Cubic Function:

max_iter = 1001, batch_size = 10000, **hidden_size = 200**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=1

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	0.46245 632469 654085	62.2504 175287	65.6954 727858	56.8412 393748	60.1028 6554	5.64705 882353	6.05882 352941	39.1764 705882
0.1	0.35859 606951 47514	70.0127 12815	73.0825 750297	65.8548 744921	68.6853 557904	9.41176 470588	9.94117 647059	54.2941 176471
0.2	0.28599 835634 23157	73.0737 592542	76.0470 242469	69.6662 262881	72.2856 49692	11.5294 117647	12.2352 941176	58.8235 294118
0.3	0.27551 470085 97851	76.8402 422913	79.3703 724637	73.4352 020341	75.5383 484994	15.9411 764706	16.7058 823529	71.2941 176471
0.5	997.890 171680 4504	28.8431 338335	30.7353 190527	18.0920 806641	19.4342 395298	1.0	1.05882 352941	22.4705 882353

max_iter = 1001, batch_size = 10000, **hidden_size = 400**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=1

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	0.43474 920123 815536	60.7448 213974	64.4803 029447	55.6148 266321	59.2861 583677	3.94117 647059	4.11764 705882	34.5294 117647
0.1	0.34417 895287 275313	62.6866 415734	66.1928 446278	58.1274 771294	61.7164 980501	6.47058 823529	6.94117 647059	36.5882 352941

0.2	0.27464 144811 0342	77.0247 027445	79.4918 894478	73.5847 645637	75.6711 69389	16.0	17.4117 647059	72.0
0.3	0.25661 812812 08992	74.3924 022235	77.1830 667496	71.1743 151282	73.6336 404228	13.2941 176471	14.4705 882353	62.7647 058824
0.5	853.619 166660 3088	26.3230 052098	29.3703 724637	1.27377 421043	1.44407 392754	0.47058 823529 4	0.47058 823529 4	0.47058 823529 4

max_iter = 1001, batch_size = 10000, **hidden_size = 600**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=1

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	0.43278 534889 221193	62.1656 654286	65.6474 311875	57.0979 88384	60.4137 229413	5.23529 411765	5.58823 529412	36.7058 823529
0.1	0.33450 608909 1301	68.3351 197747	71.5819 815746	64.3692 200314	67.4927 937602	8.88235 294118	9.47058 823529	49.7647 058824
0.2	0.27615 789830 68466	73.6346 187402	76.4172 27152	70.3666 774684	72.8480 189906	13.1176 470588	14.0	62.5882 352941
0.3	0.25336 048394 441607	76.5560 734851	78.9549 539366	73.3978 114017	75.4253 094444	15.7058 823529	17.0588 235294	69.7647 058824
0.5	678.480 613718 0328	31.2834 957749	33.3408 692703	18.2815 265349	18.7164 415305	1.35294 117647	1.47058 823529	12.2941 176471

max_iter = 1001, batch_size = 10000, **hidden_size = 1000**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=1

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	0.41620 476007 46155	62.4672 831966	65.8876 391793	57.6065 009846	60.8517 492794	6.41176 470588	6.70588 235294	42.0588 235294

0.1	0.32674 153417 34886	69.5216 491762	72.7236 760301	65.6155 744447	68.6853 557904	9.47058 823529	10.0588 235294	51.9411 764706
0.2	0.0.271 673283 576965 35	73.3055 811751	76.1431 074436	70.0501 034474	72.5795 51235	13.4705 882353	14.2941 176471	60.5882 352941
0.3	0.24602 573066 949845	79.1260 562854	81.3118 182332	75.9378 81696	77.7200 022608	18.4705 882353	20.0	77.1764 705882
0.5	767.894 207649 2309	27.3848 991699	29.3308 087944	15.4822 145225	16.1532 809586	0.88235 294117 6	1.58823 529412	19.9411 764706

max_iter = 1001, batch_size = 10000, **hidden_size = 3000**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=1

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	0.39868 636667 728424	65.4360 994092	68.7814 389872	60.7473 141062	63.8783 699768	7.29411 764706	7.82352 941176	45.9411 764706
0.1	0.31588 085830 21164	69.3870 428995	72.5004 238965	65.6155 744447	68.4649 296332	9.35294 117647	9.88235 294118	53.7058 823529
0.2	0.26021 209359 169006	74.7139 616621	77.3921 890013	71.6055 537553	73.9275 419657	14.0	15.1176 470588	64.9411 764706
0.3	0.24374 319270 25318	76.2345 140464	78.8673 48669	73.1036 717601	75.4111 795625	14.5882 352941	15.9411 764706	68.1176 470588
0.5	1050.63 931057 45315	26.3230 052098	29.3703 724637	4.33980 606725	4.92002 486859	0.47058 823529 4	0.47058 823529 4	0.47058 823529 4

RELU Function:

max_iter = 1001, batch_size = 10000, **hidden_size = 200**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=3

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	0.80913 655102 25296	53.1844 355261	56.1436 726389	44.0337 014233	46.1538 461538	3.41176 470588	3.47058 823529	39.6470 588235
0.1	0.56343 606203 79448	63.4344 542214	66.6760 865879	57.6339 207817	60.4730 684451	5.64705 882353	5.94117 647059	50.7058 823529
0.2	0.38802 938759 32693	62.7589 301294	66.5517 436274	58.4241 094798	61.9340 982309	5.35294 117647	5.64705 882353	34.1176 470588
0.3	0.32396 116077 899934	66.9292 319964	70.5476 742214	62.9309 270384	66.2324 082971	7.0	7.35294 117647	46.9411 764706
0.5	0.27176 390558 481217	72.8469 22751	75.7700 785621	69.5266 345938	72.0906 573221	11.5294 117647	2.23529 41176	58.7647 058824

max_iter = 1001, batch_size = 10000, **hidden_size = 400**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=3

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	0.70820 845663 54752	57.9829 000174	61.0580 455547	49.8541 765336	52.2636 070762	4.11764 705882	4.23529 411765	44.7058 823529
0.1	0.48916 308403 015135	64.6783 159259	67.8969 083818	59.4187 003016	62.3919 064037	6.23529 411765	6.58823 529412	49.1176 470588
0.2	0.35677 580416 20255	68.4747 11469	71.7204 54417	64.6110 127876	67.6256 146499	9.0	9.47058 823529	51.4117 647059
0.3	0.30566 906362 77199	68.6616 64631	71.9295 766687	65.1170 326794	68.1879 839485	8.47058 823529	8.88235 294118	50.1764 705882

0.5	0.25925 891473 88935	73.2931 176309	76.1939 750184	70.1847 097241	72.7491 098174	11.7058 823529	12.3529 411765	60.1764 705882
-----	----------------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

max_iter = 1001, batch_size = 10000, **hidden_size = 600**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=3

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	0.64637 573838 23395	53.4885 460029	56.7625 61465	46.0503 028641	49.0815 576782	3.23529 411765	3.23529 411765	27.5294 117647
0.1	0.44393 912643 1942	64.9674 701498	68.3490 646018	60.0219 358377	63.0644 887809	6.88235 294118	7.11764 705882	45.7058 823529
0.2	0.33620 834052 562715	69.9005 409178	73.1419 205335	66.2512 151956	69.2166 393489	9.94117 647059	0.58823 52941	55.3529 411765
0.3	0.29131 427869 200704	68.6492 010868	72.0426 157237	65.1494 378942	68.2784 151924	8.70588 235294	9.11764 705882	48.8823 529412
0.5	0.25481 068700 551984	75.4842 086896	77.9771 661109	72.3334 247327	74.4644 77477	14.4117 647059	15.2352 941176	66.7647 058824

max_iter = 1001, batch_size = 10000, **hidden_size = 1000**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=3

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	0.55437 196850 77667	62.6691 926116	65.9808 963997	57.1652 915223	60.1085 174928	5.58823 529412	5.88235 294118	49.1176 470588
0.1	0.41112 640023 231506	66.5353 840018	69.7959 645057	62.0858 987462	65.1104 956763	7.17647 058824	7.47058 823529	50.7058 823529
0.2	0.31478 930026 2928	68.7264 750604	72.0341 377946	65.2765 660443	68.3631 944837	8.41176 470588	8.82352 941176	49.8235 294118
0.3	0.28051	69.0604	72.3732	65.7800	68.8520	9.05882	9.47058	49.2941

	915287 971496	980432	549596	932273	883965	352941	823529	176471
0.5	0.24627 889275 550843	74.0509 011142	76.8637 314192	70.9748 984221	73.4329 961002	12.9411 764706	13.7058 823529	61.2352 941176

max_iter = 1001, batch_size = 10000, **hidden_size = 3000**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=3

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	0.45201 096236 70578	61.6247 476132	65.2433 165659	56.6567 789217	59.9926 524614	5.41176 470588	5.52941 176471	37.6470 588235
0.1	0.34441 839039 325717	68.1232 395244	71.3559 034646	64.4140 887903	67.4080 14469	8.29411 764706	8.58823 529412	49.2941 176471
0.2	0.28107 148438 69209	70.1672 607623	73.3227 830215	66.6525 413166	69.4794 551518	10.2941 176471	10.9411 764706	53.5882 352941
0.3	0.26132 851704 9551	69.8805 992472	73.0797 490533	66.6026 871401	69.5359 746793	9.64705 882353	10.0588 235294	51.2941 176471
0.5	0.23249 011784 791945	77.2191 34033	79.6812 298649	74.3151 282499	76.4737 466795	16.0	17.1176 470588	69.3529 411765

Tanh Function:

max_iter = 1001, batch_size = 10000, **hidden_size = 200**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=4

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	0.84082 640647 88818	49.2908 243388	52.4699 033516	40.1026 996037	42.6015 938507	2.29411 764706	2.29411 764706	26.1176 470588
0.1	0.56102 603167 29545	59.7527 232844	63.0673 147572	53.6455 86659	56.7145 198666	4.47058 823529	4.58823 529412	39.7647 058824
0.2	0.38927 578449 249267	67.3330 508263	70.6946 249929	62.9408 978737	66.0119 821398	8.05882 352941	8.52941 176471	51.0588 235294
0.3	0.32526 051968 336106	71.8548 246379	74.9138 077206	68.0983 124361	70.8048 380716	10.6470 588235	11.7647 058824	60.4117 647059
0.5	0.27133 956179 02279	74.9009 148241	77.6267 450404	71.6404 516788	73.9812 355169	14.1176 470588	15.1764 705882	65.0

max_iter = 1001, batch_size = 10000, **hidden_size = 600**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=4

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	0.60226 954638 95797	59.2915 721515	62.3975 583564	52.9800 334023	55.7112 982535	4.94117 647059	4.94117 647059	44.4117 647059
0.1	0.41758 341252 803804	66.6874 392402	70.0192 166393	62.0709 424932	65.0681 060306	7.52941 176471	8.05882 352941	50.8823 529412
0.2	0.31427 413344 383237	71.0721 140664	74.0999 265246	67.5972 77962	70.3329 00017	11.0588 235294	11.4705 882353	57.9411 764706
0.3	0.27761 517420 41111	73.9237 729641	76.6489 572147	70.6433 681482	73.0062 736676	12.3529 411765	13.1764 705882	63.2352 941176
0.5	0.24914 021357 893945	75.2374 305157	77.7680 438592	72.1240 371912	74.3457 864692	13.8823 529412	14.8823 529412	66.1764 705882

max_iter = 1001, batch_size = 10000, **hidden_size = 1000**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=4

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	0.51735 515385 86616	63.0779 968592	66.3737 071158	57.6289 353641	60.5748 035946	6.23529 411765	6.58823 529412	48.7058 823529
0.1	0.36964 027196 168897	68.3675 249894	71.6837 167241	64.3243 512725	67.3401 91036	8.88235 294118	9.35294 117647	53.5294 117647
0.2	0.29044 159054 756163	72.2935 413914	75.2953 145312	69.0555 126256	71.7204 54417	11.9411 764706	12.8823 529412	59.1176 470588
0.3	0.25959 276363 25359	75.2498 940599	78.0336 856384	72.1265 299	74.5633 866501	13.8823 529412	15.0588 235294	65.4117 647059
0.5	0.23896 628767 251968	76.0650 098462	78.4688 860001	73.1659 894808	75.2444 469564	14.2941 176471	15.4117 647059	67.9411 764706

max_iter = 1001, batch_size = 10000, **hidden_size = 3000**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=4

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	0.40043 379992 24663	67.3928 758382	70.4007 2345	63.1178 802004	65.9158 989431	8.41176 470588	8.76470 588235	52.0
0.1	0.31074 268221 85516	70.6658 025276	73.8314 58769	67.2607 622704	70.1379 076471	9.35294 117647	10.4705 882353	55.3529 411765
0.2	0.25863 624721 765516	75.0130 867213	77.7256 542135	72.0542 413441	74.4305 657605	14.0	14.8823 529412	64.0588 235294
0.3	0.23530 183091 759682	76.6208 839145	79.1979 879048	73.7118 927138	75.9424 63121	15.3529 411765	16.7058 823529	67.7647 058824
0.5	0.21607 981026 172637	80.1555 450308	82.0522 240434	77.3412 767655	78.8956 084327	21.1176 470588	22.5294 117647	78.9411 764706

Sigmoid Function:

max_iter = 1001, batch_size = 10000, **hidden_size = 200**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=5

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	1.585429952144623	21.8211730688	22.7067201718	12.2815763891	13.7935906856	0.941176470588	0.941176470588	2.88235294118
0.1	1.205525541305542	42.3162250418	45.9786356186	30.6528404417	33.8184592777	1.47058823529	1.47058823529	8.47058823529
0.2	0.8374202483892441	52.2820749308	55.1461029786	42.6801605304	44.6504267224	3.29411764706	3.58823529412	39.2941176471
0.3	0.6939410281181335	52.8504125433	55.9995478438	45.3498516838	47.9426891991	3.0	3.05882352941	30.0588235294
0.5	0.5365118047595024	58.075130244	61.6062849715	51.7935040008	54.9058949867	4.11764705882	4.29411764706	34.2941176471

max_iter = 1001, batch_size = 10000, **hidden_size = 600**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=5

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	1.2976011884212495	37.0391604557	39.9932176567	26.2557020714	29.3958062511	1.29411764706	1.29411764706	4.76470588235
0.1	0.9149715566635132	43.8442555525	47.1514158142	34.1027494578	37.311366077	1.94117647059	1.94117647059	12.2352941176
0.2	0.7147728562355041	48.1566418227	51.4073362347	41.36401027	44.3141355338	2.35294117647	2.52941176471	19.1176470588

0.3	0.62437 959253 78799	49.8641 473689	53.2866 105239	43.8666 899319	47.0609 845702	2.35294 117647	2.35294 117647	17.9411 764706
0.5	0.50591 830044 98482	51.3547 872473	54.8465 494828	45.8857 840816	49.0589 498672	2.76470 588235	2.76470 588235	17.0588 235294

max_iter = 1001, batch_size = 10000, **hidden_size = 1000**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=5

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	1.11237 917780 87615	42.6278 136451	45.9899 395241	32.3304 334821	35.5112 191262	1.52941 176471	1.52941 176471	10.4705 882353
0.1	.813142 467737 1979	45.0158 28701	48.1885 491437	37.0391 604557	40.0186 514441	2.35294 117647	2.47058 823529	14.0588 235294
0.2	0.67797 795414 92462	48.7424 283969	52.0827 445883	42.4259 042301	45.4756 118239	2.23529 411765	2.23529 411765	16.4117 647059
0.3	0.60011 277019 97757	50.8363 038114	54.2107 047985	45.2601 141661	48.4852 766631	2.35294 117647	2.35294 117647	18.6470 588235
0.5	0.50117 521017 78985	52.5313 458135	56.2312 779065	47.0473 863948	50.5171 536766	2.94117 647059	3.0	18.7058 823529

max_iter = 1001, batch_size = 10000, **hidden_size = 3000**, embedding_size = 50, display_step = 100, validation_step = 1000, n_Tokens = 48, lam = 1e-8, func=5

learning_rate	Loss	UAS	UASno Punc	LAS	LASno Punc	UEM	UEMno Punc	Root
0.05	1.25672 141432 76215	44.4076 077473	47.0638 105465	35.1571 652915	36.9298 592664	2.35294 117647	2.47058 823529	32.0588 235294
0.1	0.72980 051100 25406	48.7573 846499	51.9668 795569	41.9398 260089	45.0178 036512	2.47058 823529	2.47058 823529	17.2941 176471

0.2	0.60985 366046 42868	50.0610 713663	53.3657 378624	44.4649 400504	47.6996 552309	2.52941 176471	2.52941 176471	17.1176 470588
0.3	0.52953 373640 77568	45.2476 506219	48.5502 741197	37.9614 627215	40.8297 066637	1.41176 470588	1.52941 176471	5.41176 470588
0.5	0.48925 978034 734724	50.5471 495875	54.2107 047985	45.0158 28701	48.3354 999152	2.70588 235294	2.76470 588235	13.2941 176471

Quadratic Function:

Using Quadratic Function in Single Layer for Forward Pass instead of cubic function.

Some Experiment **Results:**

=====

hidden size: 200

learning_rate: 0.05

func: quad : 2

Average loss at step 1000 : 0.4290625768899918

Testing on dev set at step 1000

UAS: 61.7194705486

UASnoPunc: 65.297010117

LAS: 56.7988633248

LASnoPunc: 60.1141694455

UEM: 4.94117647059

UEMnoPunc: 5.58823529412

ROOT: 37.1176470588

Train Finished.

=====

hidden size: 200

learning_rate: 0.1

func: quad : 2

Average loss at step 1000 : 0.3524330887198448

Testing on dev set at step 1000

UAS: 67.0114913877

UASnoPunc: 70.6578873001

LAS: 62.3949946407

LASnoPunc: 65.7180805968

UEM: 7.05882352941

UEMnoPunc: 7.70588235294

ROOT: 44.7058823529

Train Finished.

=====

hidden size: 200

learning_rate: 0.2

func: quad : 2

Average loss at step 1000 : 0.2895881976187229

Testing on dev set at step 1000

UAS: 73.2382780367

UASnoPunc: 76.0470242469

LAS: 69.6363137822

LASnoPunc: 72.0765274402

UEM: 11.5882352941

UEMnoPunc: 12.2941176471

ROOT: 60.2941176471

Train Finished.

=====

hidden size: 200

learning_rate: 0.3

func: quad : 2

Average loss at step 1000 : 80.41801336050034

Testing on dev set at step 1000

UAS: 38.1957773512

UASnoPunc: 40.4990674278

LAS: 24.0745818481

LASnoPunc: 24.6933815633

UEM: 2.11764705882

UEMnoPunc: 2.35294117647

ROOT: 29.4705882353

Train Finished.

=====

hidden size: 200

learning_rate: 0.5

func: quad : 2

Average loss at step 1000 : 20781.557530727387

Testing on dev set at step 1000

UAS: 4.44948525563

UASnoPunc: 5.01045611259
LAS: 1.48814716953
LASnoPunc: 1.68710789578

UEM: 0.470588235294
UEMnoPunc: 0.470588235294
ROOT: 0.470588235294

Train Finished.

=====

hidden size: 400
learning_rate: 0.05
func: quad : 2
Average loss at step 1000 : 0.4304360654950142

Testing on dev set at step 1000
UAS: 62.8461749383
UASnoPunc: 66.4952240999
LAS: 57.5516613904
LASnoPunc: 60.9450064997

UEM: 5.94117647059
UEMnoPunc: 6.29411764706
ROOT: 40.4117647059

Train Finished.

=====

hidden size: 400
learning_rate: 0.1
func: quad : 2
Average loss at step 1000 : 0.33442181259393694

Testing on dev set at step 1000
UAS: 69.0455417903
UASnoPunc: 72.217826259
LAS: 65.0447441234
LASnoPunc: 67.9110382637

UEM: 9.23529411765
UEMnoPunc: 9.70588235294
ROOT: 52.7058823529

Train Finished.

=====

hidden size: 400

learning_rate: 0.2

func: quad : 2

Average loss at step 1000 : 0.2824294178187847

Testing on dev set at step 1000

UAS: 72.9790363188

UASnoPunc: 75.9509410501

LAS: 69.4742877085

LASnoPunc: 72.0850053694

UEM: 12.3529411765

UEMnoPunc: 13.1764705882

ROOT: 61.0588235294

Train Finished.

=====

hidden size: 400

learning_rate: 0.3

func: quad : 2

Average loss at step 1000 : 63.015637335777285

Testing on dev set at step 1000

UAS: 13.9990527706

UASnoPunc: 14.0648844176

LAS: 2.41543485305

LASnoPunc: 1.32820889617

UEM: 0.647058823529

UEMnoPunc: 1.11764705882

ROOT: 8.29411764706

Train Finished.

=====

hidden size: 400

learning_rate: 0.5

func: quad : 2

Average loss at step 1000 : 15534.146905083657

Testing on dev set at step 1000

UAS: 13.7423037615

UASnoPunc: 14.4266093935

LAS: 2.26587232345

LASnoPunc: 2.23817328887

UEM: 1.35294117647

UEMnoPunc: 1.41176470588

ROOT: 11.6470588235

Train Finished.

=====

hidden size: 600

learning_rate: 0.05

func: quad : 2

Average loss at step 1000 : 0.41330215334892273

Testing on dev set at step 1000

UAS: 61.3929256923

UASnoPunc: 64.822246086

LAS: 56.3227559389

LASnoPunc: 59.6478833437

UEM: 5.29411764706

UEMnoPunc: 5.76470588235

ROOT: 37.8823529412

Train Finished.

=====

hidden size: 600

learning_rate: 0.1

func: quad : 2

Average loss at step 1000 : 0.33328216314315795

Testing on dev set at step 1000

UAS: 70.9026098661

UASnoPunc: 73.8286327926

LAS: 66.9691153376

LASnoPunc: 69.6659695925

UEM: 10.3529411765

UEMnoPunc: 10.9411764706
ROOT: 56.1764705882

Train Finished.

=====

hidden size: 600
learning_rate: 0.2
func: quad : 2
Average loss at step 1000 : 0.2786798539757729

Testing on dev set at step 1000
UAS: 75.0554627714
UASnoPunc: 77.5278358673
LAS: 71.4734401875
LASnoPunc: 73.5658169898

UEM: 13.5882352941
UEMnoPunc: 14.8823529412
ROOT: 67.3529411765

Train Finished.

=====

hidden size: 600
learning_rate: 0.3
func: quad : 2
Average loss at step 1000 : 0.3080630245804787

Testing on dev set at step 1000
UAS: 75.4293690954
UASnoPunc: 77.8867348669
LAS: 72.0043871675
LASnoPunc: 74.0631888317

UEM: 15.4117647059
UEMnoPunc: 16.1764705882
ROOT: 71.3529411765

Train Finished.

=====

hidden size: 600

learning_rate: 0.5
func: quad : 2
Average loss at step 1000 : 19905.040354309083

Testing on dev set at step 1000
UAS: 15.0385123514
UASnoPunc: 15.0115865031
LAS: 1.18902211033
LASnoPunc: 1.3479907308

UEM: 0.588235294118
UEMnoPunc: 0.588235294118
ROOT: 2.58823529412

Train Finished.

=====

hidden size: 1000
learning_rate: 0.05
func: quad : 2
Average loss at step 1000 : 0.40350531101226805

Testing on dev set at step 1000
UAS: 65.697833836
UASnoPunc: 68.953823546
LAS: 61.0215120772
LASnoPunc: 64.0055389137

UEM: 6.76470588235
UEMnoPunc: 7.05882352941
ROOT: 47.4117647059

Train Finished.

=====

hidden size: 1000
learning_rate: 0.1
func: quad : 2
Average loss at step 1000 : 0.3263826879858971

Testing on dev set at step 1000
UAS: 71.8149412967
UASnoPunc: 74.8375063584
LAS: 67.9462571977

LASnoPunc: 70.6324535127

UEM: 10.7058823529

UEMnoPunc: 11.4705882353

ROOT: 60.1764705882

Train Finished.

=====

hidden size: 1000

learning_rate: 0.2

func: quad : 2

Average loss at step 1000 : 0.2751336260139942

Testing on dev set at step 1000

UAS: 74.7513522945

UASnoPunc: 77.4402305997

LAS: 71.264052646

LASnoPunc: 73.6449443283

UEM: 13.7647058824

UEMnoPunc: 15.1764705882

ROOT: 65.2941176471

Train Finished.

=====

hidden size: 1000

learning_rate: 0.3

func: quad : 2

Average loss at step 1000 : 0.35085672289133074

Testing on dev set at step 1000

UAS: 71.7127402348

UASnoPunc: 74.5407788391

LAS: 67.7518259092

LASnoPunc: 70.1718193636

UEM: 10.4705882353

UEMnoPunc: 11.1764705882

ROOT: 60.2352941176

Train Finished.

=====

hidden size: 1000

learning_rate: 0.5

func: quad : 2

Average loss at step 1000 : 18229.873020272255

Testing on dev set at step 1000

UAS: 2.40795672657

UASnoPunc: 2.39077601311

LAS: 0.588279283097

LASnoPunc: 0.4380263381

UEM: 0.470588235294

UEMnoPunc: 0.470588235294

ROOT: 5.0

Train Finished.

=====

hidden size: 3000

learning_rate: 0.05

func: quad : 2

Average loss at step 1000 : 0.3994730448722839

Testing on dev set at step 1000

UAS: 68.2453822569

UASnoPunc: 71.471768496

LAS: 63.6961886482

LASnoPunc: 66.6082631549

UEM: 9.35294117647

UEMnoPunc: 9.64705882353

ROOT: 52.3529411765

Train Finished.

=====

hidden size: 3000

learning_rate: 0.1

func: quad : 2

Average loss at step 1000 : 0.3186617481708527

Testing on dev set at step 1000

UAS: 72.630057083
UASnoPunc: 75.3744418697
LAS: 68.7788219458
LASnoPunc: 71.1693890239

UEM: 12.0588235294
UEMnoPunc: 12.5882352941
ROOT: 60.5882352941

Train Finished.

=====

hidden size: 3000
learning_rate: 0.2
func: quad : 2
Average loss at step 1000 : 0.2652495774626732

Testing on dev set at step 1000
UAS: 75.7883191664
UASnoPunc: 78.4010625671
LAS: 72.4680310093
LASnoPunc: 74.6990335161

UEM: 14.9411764706
UEMnoPunc: 16.1176470588
ROOT: 68.2941176471

Train Finished.

=====

hidden size: 3000
learning_rate: 0.3
func: quad : 2

Average loss at step 1000 : 0.2860678420960903

Testing on dev set at step 1000
UAS: 77.2216267418
UASnoPunc: 79.4466738258
LAS: 73.5673156019
LASnoPunc: 75.3433561295

UEM: 17.0
UEMnoPunc: 18.4117647059

ROOT: 77.0588235294

=====

hidden size: 3000

learning_rate: 0.5

func: quad : 2

Average loss at step 1000 : 23564.90342576027

Testing on dev set at step 1000

UAS: 1.35354089289

UASnoPunc: 1.22929972305

LAS: 0.378891741656

LASnoPunc: 0.124342960493

UEM: 0.647058823529

UEMnoPunc: 0.647058823529

ROOT: 2.58823529412