Appendix

A. Language Modelling

A.1. Model details

For WikiText-103 we swept over LSTM hidden sizes $\{1024, 2048, 4096\}$, no. LSTM layers $\{1, 2\}$, embedding dropout $\{0, 0.1, 0.2, 0.3\}$, use of layer norm (Ba et al., 2016b) $\{\mathit{True}, \mathit{False}\}$, and whether to share the input/output embedding parameters $\{\mathit{True}, \mathit{False}\}$ totalling 96 parameters.

A single-layer LSTM with 2048 hidden units with tied embedding parameters and an input dropout rate of 0.3 was selected, and we used this same model configuration for the other language corpora. We trained the models on 8 P100 Nvidia GPUs by splitting the batch size into 8 sub-batches, sending them to each GPU and summing the resulting gradients. The total batch size used was 512 and a sequence length of 100 was chosen. Gradients were clipped to a maximum norm value of 0.1. We did not pass the state of the LSTM between sequences during training, however the state is passed during evaluation.

A.2. Dynamic Evaluation Parameters

For the Neural Cache, we swept over the hyper-parameters:

- Softmax inverse temperature: $\theta_{cache} \in$ {0.1, 0.2, 0.3}
- Cache output interpolation: $\lambda_{cache} \in$ {0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35}
- Cache size $n_{cache} \in$ {1000, 5000, 8000, 9000, 10000}

and chose $\theta_{cache} = 0.3$, $\lambda_{cache} = 0.1$, $n_{cache} = 10000$ by sweeping over the validation set.

For the mixture of Neural Cache and MbPA we swept over the same cache parameters, alongside:

- MbPA output interpolation: $\lambda_{mbpa} \in$ {0.02, 0.04, 0.06, 0.08, 0.10},
- Number of neighbours retrieved from memory: $K \in \{512, 1024\}$,
- Number of MbPA steps: $n_{mbpa} \in \{1, 2\}$

and selected $\lambda_{mbpa} = 0.04, \lambda_{cache} = 0.1, \theta_{cache} = 0.3, K = 1024, n_{mbpa} = 1, n_{cache} = 10000$. We also selected the MbPA learning rate $\alpha_{lr} = 0.3$, and the L2-regularization $\beta_{mbpa} = 0.5$ on the MbPA-modified parameters. The memory size for MbPA was chosen to be equal to the cache size.

A.3. Gutenberg

A.3.1. Splits

We downloaded a subset of books (listed below) from Project Gutenberg on January 2, 2018 from a mirror site (https://www.gutenberg.org/MIRRORS.ALL). We selected 2042 English-language books under the /1 subdirectory. Each book has a unique id, we shuffled the books and split out a reasonably sized train, validation and test set. The book ids are listed below for these splits.

Test (13 books, $526, 646$ tokens):

11959, 12211, 10912, 11015, 12585, 10827, 10268, 11670, 126, 1064, 11774, 12505, 11931

Validation (12 books, $609, 545$ tokens):

11399, 10003, 1202, 12213, 11177, 12856, 10516, 11635, 12315, 11804, 11249, 11163

Train (2017 books, $175, 181, 505$ tokens)

10000, 10064, 1019, 10358, 10482, 10598, 10675, 10787, 10864, 10929, 10, 11112, 11190, 11257, 11363, 11449, 11526, 11612, 11715, 11825, 11920, 11987, 1204, 12118, 12184, 12249, 12326, 12405, 12478, 12582, 12691, 12806, 12895, 10001, 10065, 101, 1035, 10483, 10599, 10676,

```
10788, 10865, 10930, 11007, 11113, 11191, 11258, 11364, 11451, 11527, 11613, 11716, 11826, 11921, 11988, 12050, 1211, 12185, 12252,
12327, 12406, 1247, 12583, 12692, 12807, 12896, 10002, 10066, 10201, 10363, 10489, 1059, 1067, 10789, 10867, 10931, 11008, 11114, 11192,
11259, 11365, 11452, 11528, 11614, 1171, 11827, 11922, 11989, 12051, 12121, 12186, 12253, 12328, 12409, 12486, 12584, 12696, 12808,
12897, 10067, 10202, 10365, 1048, 105, 10684, 1078, 10868, 10932, 11009, 11115, 11193, 11260, 11366, 11454, 1152, 11615, 1172, 11828,
11923, 1198, 12052, 12122, 12187, 12254, 12329, 1240, 1248, 12697, 12809, 12898, 10004, 10068, 1020, 10366, 10490, 10600, 10687, 10790,
10869, 10933, 11010, 11119, 11194, 11263, 11367, 11455, 11530, 11623, 11734, 11829, 11924, 11990, 12054, 12123, 12188, 12256, 1232,
12412, 12490, 12589, 12699, 1280, 12899, 10005, 10069, 10210, 10367, 10491, 10601, 1068, 10791, 10870, 10934, 11012, 11120, 11195,
11264, 11368, 11456, 11531, 11624, 11735, 1182, 11926, 11991, 12055, 12124, 12189, 12257, 12330, 12413, 12491, 1258, 1269, 12810, 128,
10006, 1006, 10211, 10368, 10493, 10602, 10690, 10792, 10871, 10935, 11013, 11121, 11196, 11265, 11369, 11459, 11533, 11625, 11736,
11830, 11929, 11992, 12056, 12125, 1218, 12259, 12333, 12414, 12498, 12590, 12811, 12900, 10007, 10070, 10212, 10369, 1049, 10603,
10691, 10793, 10872, 10936, 11014, 11122, 11197, 11266, 11370, 1145, 11534, 11626, 11737, 11831, 11930, 11993, 12057, 12126, 12190,
1225, 12336, 12415, 124, 12591, 12700, 12813, 12901, 10008, 10071, 10213, 1036, 104, 10605, 10692, 10794, 10873, 10937, 11123, 11198,
11267, 11371, 11460, 11537, 11632, 11738, 11832, 11994, 12058, 12127, 12191, 12261, 12337, 12416, 12504, 12592, 1270, 12814, 12902,
10009, 10072, 10214, 10370, 1050, 10606, 10693, 10795, 10874, 10938, 11016, 11124, 111, 11268, 11372, 11461, 11538, 11633, 1173, 11833,
11932, 11995, 12059, 12128, 12192, 12262, 12340, 12417, 12593, 1271, 12815, 12904, 10010, 10073, 10216, 10371, 10510, 10607, 10694,
10796, 10875, 10939, 11017, 11125, 11200, 11269, 11373, 11462, 11539, 11634, 11740, 11834, 11933, 11996, 12060, 12129, 12193, 12263,
12341, 12418, 12506, 12594, 12732, 12816, 12905, 10011, 1007, 10217, 10372, 10609, 10698, 10797, 10876, 10940, 11018, 11127, 11201,
11270, 11376, 11464, 1153, 11741, 11835, 11934, 11997, 12061, 1212, 12194, 12264, 12342, 12419, 12507, 12595, 12736, 12817, 1290, 10012,
10084, 10219, 10373, 10517, 10610, 10699, 10798, 10877, 10942, 11019, 11128, 11202, 11271, 11377, 11468, 11540, 11636, 11742, 11836,
11935, 11998, 12062, 12130, 12195, 12265, 12343, 1241, 1250, 12596, 12737, 12819, 12915, 10013, 10085, 1021, 10374, 10518, 10611, 1069,
10799, 10878, 10943, 11020, 1112, 11203, 11272, 11378, 11469, 11541, 11637, 11743, 11837, 11936, 11999, 12063, 12131, 12196, 12269,
12344, 12420, 12511, 1259, 12738, 1281, 12916, 10014, 10090, 10222, 10375, 10519, 10612, 106, 1079, 10879, 10944, 11021, 11130, 11204,
11273, 11379, 1146, 11542, 11638, 11745, 11838, 11937, 119, 12064, 12132, 12197, 12270, 12345, 12421, 12512, 125, 12739, 12821, 12917,
10015, 10091, 10224, 10376, 1051, 10613, 10700, 10800, 1087, 10945, 11028, 11136, 11210, 11274, 11382, 11470, 11543, 11647, 11746,
11839, 11938, 11, 12066, 12133, 12198, 12272, 12346, 12422, 12513, 12600, 12740, 12823, 1291, 10016, 10092, 10225, 10377, 10520, 10615,
10707, 10801, 10880, 10946, 11029, 11137, 11211, 11275, 11383, 11471, 11544, 11648, 11749, 1183, 11939, 12001, 12067, 12134, 12199,
12277, 12349, 12423, 12514, 12601, 12741, 12825, 12922, 10017, 10095, 10226, 10378, 10523, 10616, 10708, 10803, 10881, 10947, 11030,
11138, 11212, 11276, 11385, 11472, 11545, 11649, 1174, 11840, 11941, 12002, 12068, 12135, 121, 12278, 12350, 12424, 12515, 12611, 12742,
12826, 12923, 10018, 10096, 1022, 10379, 1052, 10617, 10709, 10804, 10882, 10948, 11031, 11140, 11213, 11277, 11386, 11473, 11546,
1164, 11753, 11841, 11942, 12004, 12069, 12136, 12200, 12279, 12351, 12425, 12516, 12614, 12743, 12827, 12924, 10019, 10097, 10234,
10380, 10538, 10618, 10712, 10805, 10883, 10949, 11032, 11141, 11214, 11278, 11387, 11474, 11548, 11651, 11754, 11842, 11943, 12006,
1206, 12137, 12201, 1227, 12352, 12426, 12517, 12617, 12744, 12828, 12925, 1001, 10098, 1024, 10381, 10543, 10619, 10714, 10806, 10884,
1094, 11033, 11142, 11215, 11279, 11388, 11475, 11549, 11652, 1175, 11843, 11944, 12007, 12071, 12138, 12202, 12280, 12353, 12427,
1251, 12618, 12745, 1282, 12926, 10020, 10099, 10266, 10382, 10544, 1061, 10715, 10807, 10885, 10950, 11034, 11143, 11216, 11280, 11389,
11476, 1154, 11653, 11761, 11844, 11945, 1200, 12073, 12139, 12203, 12281, 12354, 12428, 12521, 12619, 12746, 12830, 12928, 10022,
100, 10267, 10383, 10545, 10620, 10716, 10808, 10886, 10954, 1103, 11144, 11217, 11281, 1138, 11477, 11550, 11654, 11762, 11845, 11946,
12010, 12074, 12140, 12204, 12282, 12357, 12429, 12522, 1261, 12747, 12832, 12929, 10023, 10100, 10386, 10546, 10621, 10717, 1080,
10887, 10955, 11045, 11145, 11218, 11282, 11390, 11478, 11551, 11655, 11763, 11846, 11947, 12013, 12077, 12141, 12205, 12283, 12358,
1242, 12523, 12622, 1274, 12833, 12933, 10024, 10101, 1027, 10388, 10550, 10622, 10720, 10811, 10888, 10956, 11047, 11146, 11219, 11283,
11391, 11479, 11552, 11656, 11764, 11847, 11948, 12014, 12078, 12142, 12206, 12285, 12359, 12430, 12524, 12628, 12750, 12834, 12934,
10025, 10102, 1028, 10389, 10551, 10623, 10721, 10812, 10889, 10957, 11050, 11147, 1121, 11284, 11392, 1147, 11553, 11658, 11765, 11848,
11949, 12015, 12079, 12143, 12207, 12286, 1235, 12431, 12525, 12629, 12753, 12835, 12935, 10029, 10103, 10291, 10392, 10554, 10624,
10722, 10813, 1088, 10958, 11051, 11148, 11221, 11289, 11395, 11480, 11554, 11659, 11768, 11849, 11950, 12016, 1207, 12144, 12208,
12287, 12360, 12433, 1252, 1262, 12754, 12836, 12936, 10030, 10104, 10292, 10393, 10555, 10625, 10734, 10814, 10890, 10959, 11052,
11149, 11222, 112, 11397, 11481, 11555, 11660, 1176, 1184, 11951, 12017, 12081, 12145, 12209, 12288, 12361, 12434, 12532, 12630, 12755,
12839, 12937, 10031, 10105, 10294, 10394, 10556, 10628, 10737, 10815, 10891, 1095, 11053, 11150, 11223, 11308, 11398, 11482, 11556,
11661, 11771, 11850, 11952, 12018, 12083, 12146, 1220, 1228, 12362, 12436, 12535, 12631, 12758, 1283, 12938, 10032, 10106, 1029, 10395,
10557, 10629, 10738, 10816, 10892, 10960, 11054, 11153, 11224, 11309, 11483, 11557, 11662, 11772, 11851, 11953, 12019, 12084, 12147,
12210, 12291, 12363, 12438, 12536, 12632, 12759, 12841, 12939, 10033, 10107, 102, 10396, 10560, 1062, 10739, 10817, 10893, 10961, 11055,
11156, 11225, 11310, 1139, 11485, 11558, 11664, 11852, 11954, 12020, 12085, 1214, 12292, 12366, 12439, 12537, 12633, 12760, 12843,
12940, 10034, 10108, 10314, 10399, 10561, 10630, 10740, 10818, 10894, 10962, 11056, 11157, 11226, 11311, 113, 11488, 11559, 11665,
1177, 11853, 11955, 12021, 12086, 12150, 12212, 12294, 1236, 12440, 12538, 12634, 12761, 12845, 12941, 10035, 10112, 10317, 1039, 10562,
10631, 10741, 1081, 10895, 10965, 11059, 11158, 11227, 11312, 11400, 11489, 1155, 11666, 1178, 11854, 11956, 12022, 12087, 12151, 12296,
12370, 12441, 12539, 12635, 12762, 12846, 12942, 10036, 10118, 10318, 103, 10563, 10632, 10743, 10826, 10896, 10966, 11060, 11159,
11228, 11313, 11401, 11490, 11565, 11667, 1179, 11855, 11957, 12023, 12088, 12152, 12214, 12297, 12371, 12442, 1253, 12638, 12763,
12847, 12943, 10037, 10119, 10319, 10401, 10564, 10633, 10744, 10897, 10967, 11067, 11160, 11229, 11314, 11402, 11491, 11566, 11668,
117, 11856, 11958, 12024, 12089, 12153, 12215, 12298, 12372, 12443, 12540, 12639, 12764, 12849, 12944, 10038, 10120, 10320, 10402,
10565, 10635, 10747, 10828, 10898, 10968, 11068, 11161, 11230, 11315, 11403, 11493, 11567, 11800, 1185, 12025, 1208, 12154, 12217,
12299, 12373, 12444, 12541, 1263, 12765, 1284, 12945, 10039, 10121, 10321, 10409, 10566, 10636, 10748, 10830, 10899, 10969, 11069,
11162, 11231, 11321, 11408, 11496, 11568, 11671, 11801, 11878, 11960, 12026, 12090, 12155, 12218, 122, 12374, 12445, 12542, 12645,
12766, 12851, 12946, 10040, 10125, 10322, 1040, 10567, 10637, 10760, 10831, 1089, 10970, 11074, 11232, 11322, 11409, 11498, 11569,
11672, 11802, 11880, 11961, 12027, 12091, 12156, 12219, 12300, 12375, 12450, 12545, 1264, 12767, 12852, 12947, 10041, 10127, 10323,
10410, 10568, 10638, 10761, 10832, 108, 10973, 11078, 11164, 11233, 11323, 1140, 11499, 1156, 11673, 11803, 11881, 11962, 12028, 12092,
12157, 12220, 12302, 12376, 12452, 12548, 12652, 12768, 12853, 12948, 10042, 10128, 10324, 10417, 10569, 10639, 10762, 10835, 10900,
10974, 11079, 11165, 11234, 11327, 11410, 114, 11570, 11674, 11882, 11963, 12029, 12093, 12158, 12221, 12304, 12378, 12453, 12549,
12653, 12769, 12854, 1294, 10043, 10129, 10327, 10418, 10570, 1063, 10763, 10837, 10901, 10976, 11080, 11166, 11235, 11328, 11411,
11503, 11571, 11675, 11805, 11883, 11965, 12094, 12159, 12222, 12305, 1237, 12454, 1254, 12654, 12770, 12855, 12951, 10044, 10130,
10328, 1041, 10571, 10642, 10765, 10840, 10902, 10979, 11082, 11167, 11236, 11329, 11416, 11504, 11575, 11676, 11806, 11885, 11966,
12030, 12096, 1215, 12223, 12306, 12380, 12455, 12550, 12655, 12779, 12955, 10045, 10131, 10329, 10422, 10572, 10643, 10766, 10842,
10904, 1097, 11083, 11168, 11237, 11330, 11417, 11505, 11576, 1167, 11807, 11886, 11969, 12031, 12097, 12160, 12224, 12307, 12381,
12456, 12551, 12658, 1277, 1285, 12956, 10046, 10132, 10330, 1043, 10573, 10767, 10843, 10905, 10981, 11084, 11169, 11238, 11331, 11418,
11506, 11577, 1168, 11808, 11887, 1196, 12032, 12098, 12161, 12225, 12308, 12383, 1245, 12552, 12659, 12781, 12860, 12957, 10047, 10133,
10331, 10451, 10576, 10655, 10769, 10844, 10908, 10983, 11085, 11170, 11239, 11332, 11419, 11507, 11578, 11690, 11809, 11888, 11970,
12033, 12099, 12162, 12226, 12309, 12384, 12460, 12553, 1265, 12784, 12863, 12958, 10048, 10134, 10332, 10452, 10577, 10656, 1076,
10847, 10909, 10984, 11088, 11171, 11240, 11333, 11420, 11508, 11579, 11691, 1180, 11889, 11971, 12034, 1209, 12163, 12227, 1230, 12387,
12461, 12554, 12664, 12785, 12864, 1297, 10049, 1013, 10333, 10453, 10578, 10657, 10770, 10848, 1090, 10985, 11089, 11172, 11241, 11334,
11421, 11509, 1157, 11692, 11810, 11890, 11972, 12035, 120, 12164, 12228, 12310, 12388, 12462, 12563, 12667, 12786, 12866, 1298, 10050,
10140, 10335, 10454, 10579, 10658, 10771, 10849, 10910, 10986, 11090, 11173, 11242, 11335, 11422, 1150, 11580, 11693, 11811, 11892,
11973, 12036, 12100, 12166, 12229, 12311, 12389, 12463, 12564, 12668, 12787, 12867, 1299, 10051, 10142, 10338, 10455, 10580, 10659,
10772, 10850, 10911, 10987, 11091, 11174, 11243, 11336, 11424, 11510, 11581, 11694, 11812, 11894, 11974, 12037, 12101, 12169, 1222,
12312, 1238, 12464, 12565, 12669, 12788, 12868, 129, 10052, 10143, 10339, 10458, 10581, 1065, 10773, 10851, 10988, 11092, 11244, 11339,
11426, 11512, 11582, 11695, 11813, 11895, 11975, 12038, 12102, 1216, 12231, 12313, 12390, 12465, 12567, 1266, 1278, 12870, 12, 10053,
10144, 1033, 10459, 10582, 10660, 10776, 10852, 10913, 10989, 11093, 11179, 11245, 11343, 11427, 11513, 1158, 11696, 11814, 11897,
11976, 12039, 12103, 12170, 12232, 12314, 12391, 12466, 12568, 12675, 12792, 12871, 13, 10054, 10147, 10340, 1045, 10583, 10661, 10777,
10853, 10916, 1098, 11095, 11180, 11246, 11344, 1142, 11514, 11599, 11697, 11815, 1189, 11977, 1203, 12104, 12171, 12233, 12392, 12467,
12569, 12676, 12793, 12872, 14, 10055, 10148, 10341, 10460, 10585, 10662, 10778, 10854, 10918, 10991, 11096, 11181, 11247, 11345, 11431,
11515, 1159, 11698, 11816, 118, 11978, 12040, 12106, 12172, 12235, 12317, 12393, 12468, 12570, 12677, 12794, 12874, 15, 10056, 1014,
10342, 10461, 10586, 10665, 10779, 10855, 10919, 10993, 11097, 11182, 11248, 11347, 11433, 11516, 115, 1169, 11817, 11901, 11979, 12041,
12107, 12173, 12236, 12318, 12394, 12469, 12571, 12678, 12797, 12880, 16, 10057, 10159, 10345, 10462, 10587, 10666, 1077, 10856, 1091,
10994, 11100, 11183, 11349, 11435, 11517, 11604, 11707, 11818, 11902, 1197, 12042, 12109, 12175, 12239, 12319, 12395, 1246, 12572, 1267,
12798, 12881, 17, 10058, 1015, 1034, 10463, 10588, 10667, 10780, 10857, 10920, 10995, 11101, 11184, 11250, 11350, 11436, 11518, 11605,
11708, 11819, 11904, 11980, 12043, 1210, 12176, 1223, 1231, 12396, 12470, 12573, 12684, 12799, 12882, 18, 10059, 10161, 10350, 10464,
10589, 10668, 10781, 10858, 10921, 10996, 11102, 11185, 11251, 11351, 11437, 11519, 11606, 11709, 1181, 11906, 11981, 12044, 12110,
12177, 12240, 12320, 12397, 12472, 12574, 12685, 1279, 12885, 19, 1005, 10162, 10351, 1046, 10590, 10670, 10782, 10859, 10922, 10997,
11105, 11186, 11252, 11352, 11438, 1151, 11607, 1170, 11820, 11912, 11982, 12045, 12111, 12179, 12241, 12321, 12398, 12473, 12575,
12686, 127, 12886, 2609, 10060, 10163, 10352, 10472, 10592, 10671, 10783, 1085, 10924, 10998, 11106, 11187, 11253, 11353, 11440, 11520,
11608, 11710, 11821, 11913, 11983, 12046, 12112, 1217, 12242, 12322, 1239, 12474, 12576, 12687, 12800, 12887, 10061, 10164, 10355,
10473, 10593, 10672, 10784, 10860, 10925, 10999, 11107, 11188, 11254, 11354, 11441, 11521, 11609, 11711, 11822, 11915, 11984, 12047,
12114, 12180, 12244, 12323, 123, 12475, 1257, 12689, 12801, 12888, 10062, 10166, 10356, 10474, 10596, 10673, 10785, 10861, 10926, 1099,
11110, 11189, 11255, 11356, 11442, 11524, 11610, 11712, 11823, 11917, 11985, 12048, 12115, 12181, 12245, 12324, 12400, 12476, 12580,
1268, 12803, 1288, 10063, 1016, 10357, 1047, 10597, 10674, 10786, 10862, 10928, 109, 11111, 1118, 11256, 11357, 11448, 11525, 11611,
11713, 11824, 11918, 11986, 12049, 12116, 12183, 12248, 12325, 12402, 12477, 12581, 12690, 12805, 12892
```

A.3.2. Data pre-processing

For Project Gutenberg and GigaWord v5 we used a very simple python script to pre-process and tokenize the data using NLTK. We post the Gutenberg script here for ease of reproduction. The GigaWord v5 script excludes the Project Gutenberg-specific selection of start / end markers to extract the text. The NLTK library[3] is used to split out sentence and word tokens, the resulting text contains lower-case text with one sentence per line.

```
# Copyright 2018 Google LLC.
#
# Licensed under the Apache License, Version 2.0 (the "License");
# you may not use this file except in compliance with the License.
# You may obtain a copy of the License at
#
# https://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.

def process_text(text):
  import nltk
  start_text = "START OF THIS PROJECT GUTENBERG EBOOK"
  start = text.find(start_text) + len(start_text)
  end = text.find("END OF THIS PROJECT GUTENBERG EBOOK")
  text = text[start:end]
  text = text.decode("utf-8", "ignore")
  text = text.replace("\r", " ")
  text = text.replace("\n", " ")
  final_text_list = []
  sent_text_tokens = nltk.sent_tokenize(text)
  for sentence in sent_text_tokens:
    final_text_list.extend(nltk.word_tokenize(sentence) + ["\n"])
  return " ".join(final_text_list).lower().encode("utf-8")
```

A.3.3. Learning Curves

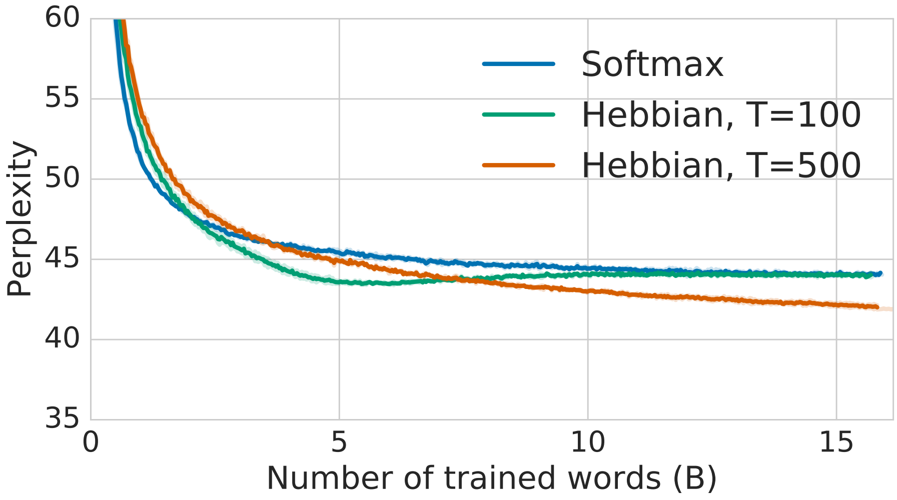

Figure 5. **Validation perplexity on Gutenberg**. All word classes have been observed after around 4B training tokens and we observe the performance of Hebbian Softmax return to that of the vanilla LSTM thereafter, as all parameters are optimized by gradient descent.

[3] https://www.nltk.org/

A.4. GigaWord

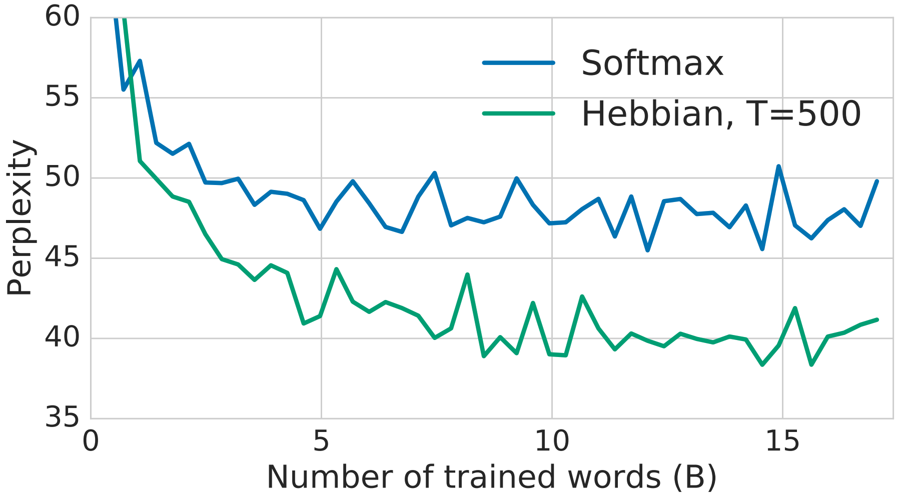

Figure 6. **Test perplexity on GigaWord v5**. Each model is trained on all articles from 200 − 2009 and tested on 2010. Because the test set is very large, a random subsample of articles are used per evaluation cycle. For this reason, the measurements are more noisy.

A.5. WikiText-103

A.5.1. AdaGrad

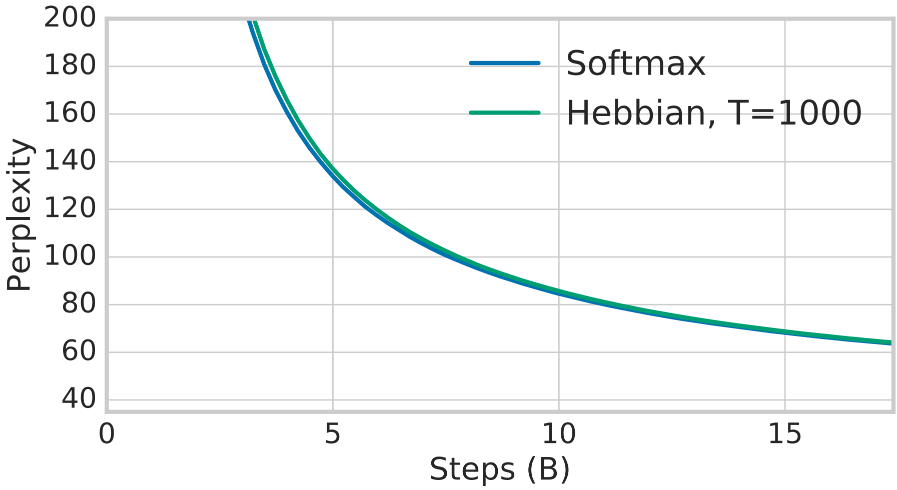

Figure 7. **Using AdaGrad**. Validation curve on WikiText-103 when using the AdaGrad optimizer with learning rate 0.2 instead of RMSProp with learning rate 0.001. The use of AdaGrad with a learning rate of 0.2 has been popular within recent language modelling literature (Grave et al., 2016b; Jozefowicz et al., 2016). After 15B steps of training both the baseline softmax and Hebbian Softmax are far from the state-of-the-art achieved with RMSProp.

A.5.2. Sampled Softmax

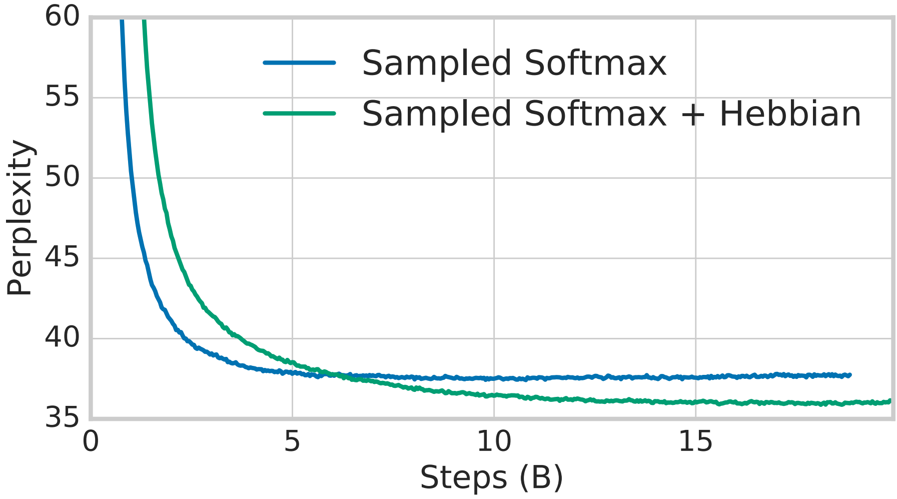

Figure 8. **Interaction with Sampled Softmax**. Validation curve on WikiText-103 when using a sampled softmax (Jean et al., 2014) with 8192 samples. Due to the smaller memory overhead, we trained with a batch size of 256 (vs 64 when using the full softmax) using 2 P100 GPUs instead of 8. The total batch size of 512 is kept the same, however training wall time is reduced from 6 days to 2. We see an improvement when using the Hebbian Softmax however both models plateau at $2 - 3$ perplexity points higher than the exact softmax.

A.5.3. Alternate Objective Functions

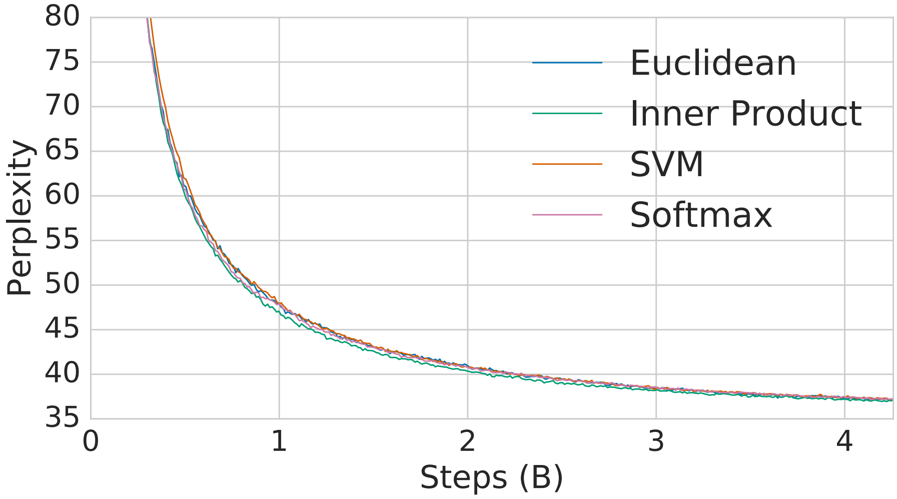

Figure 9. **Objective Function Comparison**. Validation learning curves for WikiText-103 comparing different overfitting objectives as illustrated in (7). We observe there is not a significant improvement in performance by choosing inner objectives which relate to the overall training objective, e.g. Softmax, vs $L2$.

B. Omniglot

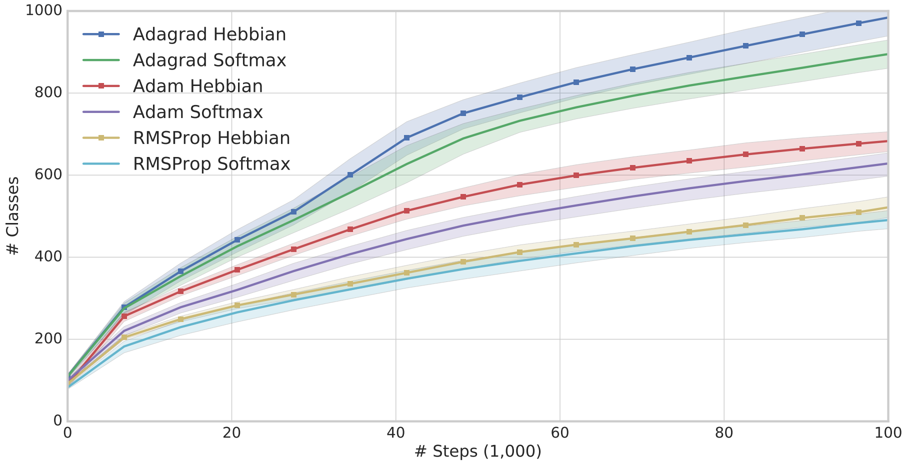

Figure 10. **Omniglot curriculum task**. Starting from 30 classes, 5 new classes are added when total test error exceeds 60%. Each line shows a 2-σ confidence band obtained from 10 independent seed runs. The Hebbian Softmax uses hyper-parameters $T = 10$ and $\gamma = 0.1$. The learning rate chosen for AdaGrad was 0.08, and 0.006 for RMSProp and Adam — these were obtained from a prior sweep with the baseline softmax model.

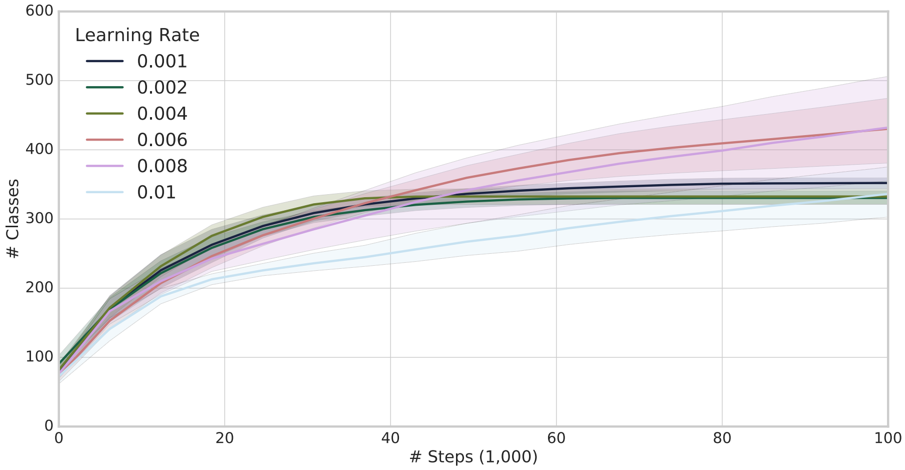

Figure 11. **Learning rate comparison**. Omniglot curriculum performance versus learning rate for a regular softmax architecture using RMSProp. Values of 0.006 to 0.008 are similarly fast to learn and are stable. Stability breaks down for larger values.