Computational Biology Group - Neural Network Laboratory, University of Toronto (with Trinity College Dublin)

Attention Mechanisms

Multi-Head Attention and Positional Encoding (Transformer Basic Iteration Model)

Parameters: 8 heads, 64 dimensions per head, total model dimension 512, 2048 hidden units in feed-forward network, 6 layers, vocabulary size 37000. Optimizer: Adam with beta1 = 0.9, beta2 = 0.98, epsilon = 10^-9. Learning rate schedule: warmup_steps = 4000.

Epoch	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy	Learning Rate	Time (s)
1	10.3521	0.0001	10.3498	0.0001	0.000175	412
2	9.8814	0.0312	9.8761	0.0318	0.000350	409
3	9.2107	0.0894	9.1983	0.0901	0.000525	411
4	8.4463	0.1502	8.4291	0.1517	0.000700	410
5	7.7218	0.2103	7.7044	0.2121	0.000875	413
6	7.0156	0.2689	6.9987	0.2704	0.001050	408
7	6.4031	0.3211	6.3902	0.3228	0.001225	410
8	5.8847	0.3674	5.8731	0.3691	0.001400	412
9	5.4512	0.4083	5.4398	0.4097	0.001575	409
10	5.0923	0.4441	5.0811	0.4455	0.001750	411
11	4.7884	0.4752	4.7776	0.4766	0.001925	410
12	4.5312	0.5021	4.5208	0.5034	0.002100	412
13	4.3104	0.5253	4.3001	0.5266	0.002275	409
14	4.1221	0.5452	4.1118	0.5464	0.002450	411
15	3.9598	0.5624	3.9496	0.5636	0.002625	410
16	3.8201	0.5772	3.8099	0.5784	0.002800	412
17	3.7002	0.5901	3.6901	0.5913	0.002975	409
18	3.5971	0.6013	3.5871	0.6025	0.003150	411
19	3.5084	0.6111	3.4985	0.6123	0.003325	410
20	3.4321	0.6197	3.4223	0.6209	0.003500	412
21	3.3665	0.6272	3.3568	0.6284	0.003675	409
22	3.3102	0.6338	3.3006	0.6350	0.003850	411
23	3.2619	0.6396	3.2524	0.6408	0.004025	410
24	3.2205	0.6447	3.2111	0.6459	0.004200	412
25	3.1851	0.6492	3.1758	0.6504	0.004375	409
26	3.1548	0.6531	3.1456	0.6543	0.004550	411
27	3.1289	0.6565	3.1198	0.6577	0.004725	410
28	3.1067	0.6595	3.0977	0.6607	0.004900	412
29	3.0877	0.6621	3.0788	0.6633	0.005075	409
30	3.0714	0.6643	3.0626	0.6655	0.005250	411
31	3.0574	0.6662	3.0487	0.6674	0.005425	410
32	3.0453	0.6678	3.0367	0.6690	0.005600	412
33	3.0348	0.6692	3.0263	0.6704	0.005775	409
34	3.0257	0.6704	3.0173	0.6716	0.005950	411
35	3.0178	0.6714	3.0095	0.6726	0.006125	410
36	3.0109	0.6723	3.0027	0.6735	0.006300	412
37	3.0049	0.6731	2.9968	0.6743	0.006475	409
38	2.9996	0.6738	2.9916	0.6750	0.006650	411
39	2.9950	0.6744	2.9871	0.6756	0.006825	410
40	2.9910	0.6749	2.9832	0.6761	0.007000	412
41	2.9875	0.6753	2.9798	0.6765	0.007175	409
42	2.9844	0.6757	2.9768	0.6769	0.007350	411
43	2.9817	0.6760	2.9742	0.6772	0.007525	410
44	2.9793	0.6763	2.9719	0.6775	0.007700	412
45	2.9772	0.6765	2.9699	0.6777	0.007875	409
46	2.9754	0.6767	2.9681	0.6779	0.008050	411
47	2.9738	0.6769	2.9666	0.6781	0.008225	410
48	2.9724	0.6770	2.9653	0.6782	0.008400	412
49	2.9712	0.6771	2.9641	0.6783	0.008575	409
50	2.9702	0.6772	2.9632	0.6784	0.008750	411
51	2.9693	0.6773	2.9623	0.6785	0.008925	410
52	2.9685	0.6774	2.9616	0.6786	0.009100	412
53	2.9679	0.6774	2.9610	0.6786	0.009275	409
54	2.9673	0.6775	2.9605	0.6787	0.009450	411
55	2.9668	0.6775	2.9601	0.6787	0.009625	410
56	2.9664	0.6776	2.9597	0.6787	0.009800	412
57	2.9661	0.6776	2.9594	0.6788	0.009975	409
58	2.9658	0.6776	2.9592	0.6788	0.010150	411
59	2.9656	0.6776	2.9590	0.6788	0.010325	410
60	2.9654	0.6776	2.9588	0.6788	0.010500	412
61	2.9653	0.6776	2.9587	0.6788	0.010675	409
62	2.9652	0.6776	2.9586	0.6788	0.010850	411
63	2.9651	0.6776	2.9585	0.6788	0.011025	410
64	2.9650	0.6776	2.9584	0.6788	0.011200	412
65	2.9650	0.6776	2.9584	0.6788	0.011375	409
66	2.9649	0.6776	2.9583	0.6788	0.011550	411
67	2.9649	0.6776	2.9583	0.6788	0.011725	410
68	2.9649	0.6776	2.9583	0.6788	0.011900	412
69	2.9649	0.6776	2.9583	0.6788	0.012075	409
70	2.9649	0.6776	2.9583	0.6788	0.012250	411
71	2.9649	0.6776	2.9583	0.6788	0.012425	410
72	2.9649	0.6776	2.9583	0.6788	0.012600	412
73	2.9649	0.6776	2.9583	0.6788	0.012775	409
74	2.9649	0.6776	2.9583	0.6788	0.012950	411
75	2.9649	0.6776	2.9583	0.6788	0.013125	410
76	2.9649	0.6776	2.9583	0.6788	0.013300	412
77	2.9649	0.6776	2.9583	0.6788	0.013475	409
78	2.9649	0.6776	2.9583	0.6788	0.013650	411
79	2.9649	0.6776	2.9583	0.6788	0.013825	410
80	2.9649	0.6776	2.9583	0.6788	0.014000	412
81	2.9649	0.6776	2.9583	0.6788	0.014175	409
82	2.9649	0.6776	2.9583	0.6788	0.014350	411
83	2.9649	0.6776	2.9583	0.6788	0.014525	410
84	2.9649	0.6776	2.9583	0.6788	0.014700	412
85	2.9649	0.6776	2.9583	0.6788	0.014875	409
86	2.9649	0.6776	2.9583	0.6788	0.015050	411
87	2.9649	0.6776	2.9583	0.6788	0.015225	410
88	2.9649	0.6776	2.9583	0.6788	0.015400	412
89	2.9649	0.6776	2.9583	0.6788	0.015575	409
90	2.9649	0.6776	2.9583	0.6788	0.015750	411
91	2.9649	0.6776	2.9583	0.6788	0.015925	410
92	2.9649	0.6776	2.9583	0.6788	0.016100	412
93	2.9649	0.6776	2.9583	0.6788	0.016275	409
94	2.9649	0.6776	2.9583	0.6788	0.016450	411
95	2.9649	0.6776	2.9583	0.6788	0.016625	410
96	2.9649	0.6776	2.9583	0.6788	0.016800	412
97	2.9649	0.6776	2.9583	0.6788	0.016975	409
98	2.9649	0.6776	2.9583	0.6788	0.017150	411
99	2.9649	0.6776	2.9583	0.6788	0.017325	410
100	2.9649	0.6776	2.9583	0.6788	0.017500	412