CS department, Technion

Introduction to Optimization and Deep Learning 236330

HW 5

Training Neural Networks with SGD and Adagrad Methods

Uri Kirstein 311137095 sukirstn@campus.technion.ac.il
Pavel Rastopchin 321082026 pavelr@campus.technion.ac.il

June 22, 2019

1 Finding optimal learning rate

In order to optimize the initial training rate (noted below as Alpha 0) and decay constant k, we ran the network on different values of those parameters in a log scale search. The learning rate scale runs from 0.1 to 10^-6, while the decay constant runs from 1 (no decay) to 10^-7. We used the following sources as a reference:
`https://machinelearningmastery.com/learning-rate-for-deep-learning-neural-networks/`
`https://towardsdatascience.com/learning-rate-schedules-and-adaptive-learning-rate-methods-for-deep-learning-2c8f433990d1`
We tested the different values on both the Stochastic Gradient Descent and Adagrad algorithms. In all the tests below we have used the same weights, which were generated vie the Xavier initialization method. In all tests below batch size was 32.
Each test was run on a train set that contains 500 points, like last exercise. We ran the optimization algorithms for 500 epochs each, as we saw the convergence happens before than anyway. See the discussion on the course's Google group on the matter. After that we used the obtained weights and biases, fed them to the network and checked the loss on the test set. The test set consists of 200 points. The training and test sets were created in an identical way.

Algorithm	Alpha 0	Decay rate	Train loss	Test loss
SGD	0.1	1.0	0.02304088916599846	0.028515195833633086
SGD	0.1	0.1	0.0018931197285240637	0.05604633606922681
SGD	0.1	0.01	0.0013102708255736516	0.06719278992097875
SGD	0.1	0.001	0.0007177509082403669	0.07354370533219776
SGD	0.1	0.0001	0.0005121250881201063	0.08100440566702442
SGD	0.1	1e-05	0.0005095360920303392	0.08360530088816429
SGD	0.1	1e-06	0.0004831592867095252 5	0.08257291112850915
SGD	0.1	1e-07	0.0004972862282315955	0.08064603631831078
SGD	0.01	1.0	0.04311311806410595	0.025978743113740446
SGD	0.01	0.1	0.028093392862017584	0.02670562111675416
SGD	0.01	0.01	0.002108707418607295	0.05355775640855056
SGD	0.01	0.001	0.0013701205010772928	0.060026107312705274
SGD	0.01	0.0001	0.0013410822398531265	0.06783345969962629
SGD	0.01	1e-05	0.0013523411852612928	0.06029102025174592
SGD	0.01	1e-06	0.0013389920900253559	0.0678608389434266
SGD	0.01	1e-07	0.0014122980415587258	0.06216954526522393
SGD	0.001	1.0	0.04376299824037411	0.02608622889142703
SGD	0.001	0.1	0.04032211387145754	0.02601280327438891
SGD	0.001	0.01	0.02675408567226814	0.026532694779291287
SGD	0.001	0.001	0.009363782113864572	0.03608854817552043
SGD	0.001	0.0001	0.006779074973187606	0.03951227033964753
SGD	0.001	1e-05	0.006681240035791876	0.0402700221934518
SGD	0.001	1e-06	0.0061980476244608805	0.03920993435179229
SGD	0.001	1e-07	0.006511716099854141	0.03994280566926465
SGD	0.0001	1.0	0.044995943358772356	0.026100946203832256
SGD	0.0001	0.1	0.04483257964814191	0.026091530482970304
SGD	0.0001	0.01	0.04040963114236112	0.02601639985374699
SGD	0.0001	0.001	0.03837830544744873	0.025962002800411756
SGD	0.0001	0.0001	0.03775188649028052	0.02600833012771269
SGD	0.0001	1e-05	0.03838138455222339	0.026019607193168263
SGD	0.0001	1e-06	0.03836404701990857	0.026020366314491075
SGD	0.0001	1e-07	0.033551734089903563	0.025996033487243105
SGD	1e-05	1.0	0.044043449617488226	0.0261024789868067
SGD	1e-05	0.1	0.04696218460067136	0.02610149344121672
SGD	1e-05	0.01	0.04484221005112403	0.026092077096963956
SGD	1e-05	0.001	0.03756264667541108	0.026065825793139607
SGD	1e-05	0.0001	0.049894396028453004	0.026052106937265344
SGD	1e-05	1e-05	0.042202438191146385	0.02605462157677637
SGD	1e-05	1e-06	0.045078213947253136	0.026053187206377792
SGD	1e-05	1e-07	0.03932055766946278	0.026055861823366274
SGD	1e-06	1.0	0.04404600571587572	0.02610263095339228
SGD	1e-06	0.1	0.04600166135280474	0.026102533355302847
SGD	1e-06	0.01	0.04500665088286665	0.026101573028238706
SGD	1e-06	0.001	0.04299863719221167	0.026098517535804597
SGD	1e-06	0.0001	0.04591247655753785	0.026097381144084122
SGD	1e-06	1e-05	0.05274660167395026	0.026096886361890975
SGD	1e-06	1e-06	0.0459102136474384	0.026097251223689454
SGD	1e-06	1e-07	0.04200331069340839	0.026097466142397777

Algorithm	Alpha 0	Decay rate	Train loss	Test loss
Adagrad	0.1	1.0	0.020556504704633553	0.028164649007004078
Adagrad	0.1	0.1	0.001887738384629929	0.05759206449078611
Adagrad	0.1	0.01	0.001332113357856352	0.06534841255837116
Adagrad	0.1	0.001	0.0006988000980388597	0.07798190621343325
Adagrad	0.1	0.0001	0.0005189662872864132	0.08320041282924077
Adagrad	0.1	1e-05	0.0004996249100370246	0.08057305677016345
Adagrad	0.1	1e-06	0.0004888515320450285	0.08380113312843675
Adagrad	0.1	1e-07	0.0004969902429662896	0.08381962566478178
Adagrad	0.01	1.0	0.044029581547942495	0.025978056113518146
Adagrad	0.01	0.1	0.02809339286201759	0.02670562111675416
Adagrad	0.01	0.01	0.002108707418607295	0.05355775640855056
Adagrad	0.01	0.001	0.001322360631415768	0.059100556860375034
Adagrad	0.01	0.0001	0.0013076505189168071	0.05933076024151953
Adagrad	0.01	1e-05	0.0013391849727495794	0.06785831663143055
Adagrad	0.01	1e-06	0.001377436272998534	0.06690415687920219
Adagrad	0.01	1e-07	0.0013774158471600005	0.06690442072879281
Adagrad	0.001	1.0	0.04084517478545872	0.026086720997374774
Adagrad	0.001	0.1	0.04126123305274803	0.0260120947777102
Adagrad	0.001	0.01	0.02941250653303231	0.026624469208133997
Adagrad	0.001	0.001	0.010326117410255548	0.03748464869386203
Adagrad	0.001	0.0001	0.006779074973187606	0.03951227033964753
Adagrad	0.001	1e-05	0.006384514562489407	0.03953393336636715
Adagrad	0.001	1e-06	0.007125661655290371	0.04184114475069464
Adagrad	0.001	1e-07	0.006654331198575149	0.040314238301538394
Adagrad	0.0001	1.0	0.042061776396556165	0.026100998179252873
Adagrad	0.0001	0.1	0.04873026173011824	0.026091086608175943
Adagrad	0.0001	0.01	0.04417119435977589	0.02601365170774487
Adagrad	0.0001	0.001	0.03837830544744873	0.025962002800411756
Adagrad	0.0001	0.0001	0.034529162185121386	0.025993285632024535
Adagrad	0.0001	1e-05	0.03758452707078879	0.02601530773618292
Adagrad	0.0001	1e-06	0.03995065987005182	0.026029340691016034
Adagrad	0.0001	1e-07	0.034358809703025586	0.025999837778269662
Adagrad	1e-05	1.0	0.04893648437156966	0.02610247027767048
Adagrad	1e-05	0.1	0.04207089855797815	0.026101551093413347
Adagrad	1e-05	0.01	0.04484221005112403	0.026092077096963956
Adagrad	1e-05	0.001	0.04430974649549981	0.026063231978471303
Adagrad	1e-05	0.0001	0.045100425031660105	0.02605427128782814
Adagrad	1e-05	1e-05	0.04412121920439468	0.02605373070814256
Adagrad	1e-05	1e-06	0.044119213776564026	0.026053631678140413
Adagrad	1e-05	1e-07	0.04124063503914782	0.026054962334086343
Adagrad	1e-06	1.0	0.04306734117358843	0.026102631127672006
Adagrad	1e-06	0.1	0.046980289705082695	0.02610253219805385
Adagrad	1e-06	0.01	0.043050090344464255	0.026101594939690616
Adagrad	1e-06	0.001	0.04690710024002134	0.02609834669681139
Adagrad	1e-06	0.0001	0.046889187715224365	0.026097328441349897
Adagrad	1e-06	1e-05	0.0478637331223279	0.02609715548087994
Adagrad	1e-06	1e-06	0.0459102136474384	0.026097251223689454
Adagrad	1e-06	1e-07	0.045910192733146185	0.02609725002314522