1d)

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02195529826247768 | 0.023044304535428106 | 0.02796169729737883

L\_1 Model |0.16699146530459466 | 0.15706138462087946 | 0.20572626053247667

L\_inf Model |5.00974395913784 | 4.711841538626384 | 6.1717878159743

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.0197020636144262 | 0.01976465141082499 | 0.02394708648907498

L\_1 Model |0.15908510988891536 | 0.15889173091578115 | 0.17452235311639228

L\_inf Model |159.08510988891535 | 158.89173091578115 | 174.52235311639228

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.008752817096550427 | 0.010677393902190128 | 0.012263669973475017

L\_1 Model |0.10567865679129301 | 0.09725614502158306 | 0.12456209232930808

L\_inf Model |3.1703597037387903 | 2.917684350647492 | 3.7368627698792425

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02125981235371155 | 0.021385321561594627 | 0.024019361084271395

L\_1 Model |0.16321793303009113 | 0.16269328015306836 | 0.1760734900880556

L\_inf Model |163.21793303009113 | 162.69328015306837 | 176.0734900880556

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.013499382685245327 | 0.016085967030890914 | 0.014062858114292679

L\_1 Model |0.1374416653875065 | 0.13081120494045734 | 0.1425706482457309

L\_inf Model |4.123249961625195 | 3.9243361482137202 | 4.277119447371927

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01811270124466558 | 0.018139438173941418 | 0.021578814672670442

L\_1 Model |0.15129302090387506 | 0.15115310072181126 | 0.16388770757099552

L\_inf Model |151.29302090387506 | 151.15310072181126 | 163.88770757099553

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01836494334840313 | 0.020616546465203448 | 0.023941239912593514

L\_1 Model |0.14688834208972787 | 0.14146647609786295 | 0.19300674644327778

L\_inf Model |4.406650262691836 | 4.243994282935889 | 5.790202393298333

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019954625722984832 | 0.02009395211604993 | 0.024845952561566614

L\_1 Model |0.15678943939460993 | 0.1560642563825354 | 0.17791271383917867

L\_inf Model |156.78943939460993 | 156.0642563825354 | 177.91271383917868

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01977439659798246 | 0.023341171632482092 | 0.02400850911734353

L\_1 Model |0.1491691531435136 | 0.13609773657351282 | 0.168581544897101

L\_inf Model |4.475074594305408 | 4.082932097205385 | 5.05744634691303

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020197943031174215 | 0.02021148145235356 | 0.02248442769788625

L\_1 Model |0.15856663652556244 | 0.15845711980105193 | 0.1682351076162827

L\_inf Model |158.56663652556244 | 158.45711980105193 | 168.23510761628268

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.023272401838622128 | 0.026282207199102556 | 0.037756658929304035

L\_1 Model |0.17567463117836707 | 0.1682102648589518 | 0.23991638474877602

L\_inf Model |5.270238935351012 | 5.046307945768554 | 7.197491542463281

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020592790241771863 | 0.020615236170335578 | 0.0212975800839843

L\_1 Model |0.16135105956172308 | 0.1612602919183508 | 0.16303059973182218

L\_inf Model |161.35105956172308 | 161.26029191835082 | 163.03059973182218

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.025561731373163518 | 0.02700896285585302 | 0.04095806026468084

L\_1 Model |0.16237515903538105 | 0.15329995913498928 | 0.24184609096281534

L\_inf Model |4.871254771061431 | 4.5989987740496785 | 7.25538272888446

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019664026005465223 | 0.019730992903327672 | 0.023746824457817946

L\_1 Model |0.15879953665174926 | 0.1584758139969644 | 0.1759517818375853

L\_inf Model |158.79953665174926 | 158.4758139969644 | 175.95178183758532

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018740472294329028 | 0.020933606503146115 | 0.02866589521652946

L\_1 Model |0.1557593917770396 | 0.15048554286773666 | 0.2130017324151519

L\_inf Model |4.672781753311188 | 4.5145662860321 | 6.3900519724545575

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.021154136065387248 | 0.021179284315651997 | 0.025332955417215842

L\_1 Model |0.16459562459493282 | 0.16444944258245328 | 0.1807055582238245

L\_inf Model |164.59562459493281 | 164.4494425824533 | 180.7055582238245

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.022242110070891653 | 0.02400775720521399 | 0.034131974670979184

L\_1 Model |0.15869005354073515 | 0.14739314359056352 | 0.22613158746331788

L\_inf Model |4.760701606222055 | 4.4217943077169055 | 6.783947623899536

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.0197634341675143 | 0.01982160668589218 | 0.026455292665964133

L\_1 Model |0.15845402911250941 | 0.15811057197664155 | 0.18394812172343328

L\_inf Model |158.4540291125094 | 158.11057197664155 | 183.94812172343327

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.017491947452888276 | 0.023213383084767725 | 0.02198782403340773

L\_1 Model |0.15540583284476012 | 0.1401896601199901 | 0.18767401773164544

L\_inf Model |4.662174985342803 | 4.205689803599703 | 5.630220531949363

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02055199785119886 | 0.020596709192683648 | 0.0274072573953328

L\_1 Model |0.16081213330529476 | 0.16051348235361723 | 0.1891471656638954

L\_inf Model |160.81213330529476 | 160.51348235361723 | 189.14716566389538

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01752766841269136 | 0.022876187775553662 | 0.02172128892139564

L\_1 Model |0.15198214131621757 | 0.13898286462774895 | 0.179182242341092

L\_inf Model |4.559464239486527 | 4.1694859388324685 | 5.37546727023276

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019723822233585493 | 0.01981149735224863 | 0.01994437617311476

L\_1 Model |0.15830820611167745 | 0.1579533536186054 | 0.15985032233238264

L\_inf Model |158.30820611167746 | 157.9533536186054 | 159.85032233238263

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.017237742295462686 | 0.0198913753113182 | 0.026700130610797658

L\_1 Model |0.15452045528683148 | 0.14924771737670436 | 0.19780218185001752

L\_inf Model |4.635613658604944 | 4.477431521301131 | 5.934065455500526

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019603675781241208 | 0.019644203944078037 | 0.026562389456833986

L\_1 Model |0.1588895428490465 | 0.1586832389091324 | 0.1864418865107403

L\_inf Model |158.8895428490465 | 158.6832389091324 | 186.44188651074032

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02021587831923378 | 0.02194420063035372 | 0.02839060931181086

L\_1 Model |0.16890493649783195 | 0.16127066417244648 | 0.20900082028351893

L\_inf Model |5.067148094934958 | 4.838119925173395 | 6.270024608505568

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020167929598441117 | 0.020259476682990444 | 0.022074257023379132

L\_1 Model |0.1615694442155779 | 0.161161882769433 | 0.16728410797673182

L\_inf Model |161.5694442155779 | 161.161882769433 | 167.28410797673183

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.015055036514925624 | 0.017204557930210005 | 0.01782540718920499

L\_1 Model |0.13978420950322593 | 0.1318113528920079 | 0.16488804970680995

L\_inf Model |4.193526285096778 | 3.954340586760237 | 4.946641491204298

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.0194274902233011 | 0.01947533086830826 | 0.022905709815673466

L\_1 Model |0.15791806679017686 | 0.15765322473828938 | 0.17367316793498902

L\_inf Model |157.91806679017685 | 157.6532247382894 | 173.673167934989

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.016650448363992425 | 0.020698016535881214 | 0.022569858873778586

L\_1 Model |0.14726263832033928 | 0.13669701540107532 | 0.1850684383999022

L\_inf Model |4.417879149610179 | 4.100910462032259 | 5.5520531519970655

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019581855116755528 | 0.019652725721338006 | 0.020407515313740264

L\_1 Model |0.15945325456306342 | 0.15918409779511922 | 0.1638651051527742

L\_inf Model |159.4532545630634 | 159.18409779511921 | 163.8651051527742

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.015142278335053005 | 0.016024649624925297 | 0.020676669278926123

L\_1 Model |0.1291153352449778 | 0.1214319032992137 | 0.1723113738567438

L\_inf Model |3.8734600573493347 | 3.642957098976411 | 5.169341215702314

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02037787938167375 | 0.020429517328455105 | 0.024006202983601076

L\_1 Model |0.1615470882453407 | 0.16118465269610247 | 0.1774422284380188

L\_inf Model |161.5470882453407 | 161.18465269610246 | 177.4422284380188

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.016362755048176407 | 0.017546011575480306 | 0.018918371012604247

L\_1 Model |0.14203823418901337 | 0.13593990336994316 | 0.1561397725101537

L\_inf Model |4.261147025670401 | 4.0781971010982945 | 4.684193175304611

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019432087925893977 | 0.019488063518607657 | 0.02193693112399669

L\_1 Model |0.15698298396298488 | 0.1567444097053316 | 0.16851314508138174

L\_inf Model |156.9829839629849 | 156.7444097053316 | 168.51314508138174

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.013527900533876541 | 0.016314028469915587 | 0.015367107786566015

L\_1 Model |0.13778103261865168 | 0.12985367243409293 | 0.15120134705563568

L\_inf Model |4.13343097855955 | 3.8956101730227877 | 4.53604041166907

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01974273002640424 | 0.019764742726281165 | 0.02244681796181545

L\_1 Model |0.15840756841798964 | 0.15826224782023035 | 0.16734664826496917

L\_inf Model |158.40756841798964 | 158.26224782023036 | 167.34664826496916

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.017329667349628826 | 0.01827908359260939 | 0.023331785416344456

L\_1 Model |0.1360414504604143 | 0.13141162577753873 | 0.17333293523015064

L\_inf Model |4.081243513812429 | 3.942348773326162 | 5.19998805690452

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018911580669392993 | 0.018968582440108465 | 0.021693228389465843

L\_1 Model |0.15430817193116925 | 0.15399863528142896 | 0.16541886430406674

L\_inf Model |154.30817193116926 | 153.99863528142896 | 165.41886430406674

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018065773363269675 | 0.023836068160838917 | 0.02175465803170375

L\_1 Model |0.15410306845227023 | 0.13269114814741748 | 0.17953911697787242

L\_inf Model |4.623092053568107 | 3.9807344444225246 | 5.386173509336173

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019210289324323428 | 0.019229290137182543 | 0.021620044224097343

L\_1 Model |0.1541298104902452 | 0.15403715556870456 | 0.16328457184107978

L\_inf Model |154.1298104902452 | 154.03715556870458 | 163.28457184107978

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.010955079040904132 | 0.012087418061012207 | 0.013953989689533824

L\_1 Model |0.12116968644408803 | 0.11111322812268538 | 0.13643245781459754

L\_inf Model |3.6350905933226407 | 3.3333968436805614 | 4.092973734437926

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020879712894578857 | 0.020897541168548663 | 0.021918805237831226

L\_1 Model |0.16219608515459433 | 0.16215730282258986 | 0.16638569119627766

L\_inf Model |162.19608515459433 | 162.15730282258986 | 166.38569119627766

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02413419926660108 | 0.02713641729963268 | 0.038006551331576086

L\_1 Model |0.16447772313655687 | 0.15984272310192954 | 0.2301055912439382

L\_inf Model |4.934331694096706 | 4.795281693057886 | 6.903167737318146

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01973481694573114 | 0.019789920356076646 | 0.02392646978576763

L\_1 Model |0.1593839575105546 | 0.15917578998180562 | 0.1739338521700883

L\_inf Model |159.38395751055458 | 159.17578998180562 | 173.9338521700883

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.017285262633502736 | 0.020103267701723042 | 0.019352324340646508

L\_1 Model |0.15573737342537144 | 0.14950408825509684 | 0.169306217701957

L\_inf Model |4.672121202761144 | 4.485122647652905 | 5.07918653105871

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020649829177498226 | 0.020699132840563385 | 0.022965511612375613

L\_1 Model |0.16129086984669688 | 0.16108754099253111 | 0.1717141522387046

L\_inf Model |161.2908698466969 | 161.0875409925311 | 171.7141522387046

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.011951205192693727 | 0.01332473435416138 | 0.013112337289141785

L\_1 Model |0.1267929785629277 | 0.11987938548873471 | 0.1384828467523744

L\_inf Model |3.803789356887831 | 3.596381564662041 | 4.1544854025712326

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02095718852128461 | 0.021007454703398615 | 0.02427289822921936

L\_1 Model |0.16431753955228207 | 0.1641597132773084 | 0.17651457262661066

L\_inf Model |164.31753955228206 | 164.1597132773084 | 176.51457262661066

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.012802181297566726 | 0.01565872241117478 | 0.015945268529325908

L\_1 Model |0.13227408224332096 | 0.12604004044783015 | 0.15160663047754072

L\_inf Model |3.968222467299629 | 3.7812012134349047 | 4.548198914326222

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019897018078976654 | 0.019947353125676817 | 0.02334021393730012

L\_1 Model |0.15926385096916276 | 0.1590130445504282 | 0.17220847022481703

L\_inf Model |159.26385096916277 | 159.0130445504282 | 172.20847022481703

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.011998827749818976 | 0.012655207021677857 | 0.013854346256913511

L\_1 Model |0.12267503471729185 | 0.11832123894379477 | 0.13116768073388843

L\_inf Model |3.6802510415187557 | 3.549637168313843 | 3.935030422016653

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020726168762113875 | 0.020841492290756228 | 0.02578484836591248

L\_1 Model |0.16394945872625866 | 0.16353963302973848 | 0.1836587884367355

L\_inf Model |163.94945872625865 | 163.53963302973847 | 183.6587884367355

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.014682774735602205 | 0.015305492138244475 | 0.0199056024233583

L\_1 Model |0.13122822073980817 | 0.1264320627812644 | 0.169396653519

L\_inf Model |3.9368466221942446 | 3.792961883437932 | 5.081899605569999

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01923676055495136 | 0.01932230558485573 | 0.022608050856021467

L\_1 Model |0.1591989502124915 | 0.15885934311366687 | 0.17211994379488457

L\_inf Model |159.1989502124915 | 158.85934311366688 | 172.11994379488456

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.025205021105789968 | 0.026142179461571335 | 0.02730418095417095

L\_1 Model |0.18047562955679822 | 0.17296765113552642 | 0.200585108531473

L\_inf Model |5.414268886703947 | 5.189029534065793 | 6.01755325594419

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01959579689594081 | 0.019644920476908436 | 0.022805089237441827

L\_1 Model |0.1571242257654945 | 0.15685798723142627 | 0.17012814994296294

L\_inf Model |157.12422576549451 | 156.85798723142628 | 170.12814994296295

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01513099192632439 | 0.01617780537166983 | 0.019352457996298845

L\_1 Model |0.1415798029613482 | 0.13568234291855905 | 0.17559976134126504

L\_inf Model |4.247394088840446 | 4.0704702875567715 | 5.267992840237951

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.021219509406255225 | 0.021260214296801565 | 0.026767345066218286

L\_1 Model |0.1622021353765534 | 0.1620312342289163 | 0.18336670019700405

L\_inf Model |162.2021353765534 | 162.0312342289163 | 183.36670019700404

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.009924245822362539 | 0.011738650823488307 | 0.013375915313151255

L\_1 Model |0.11628713536101044 | 0.10889983247754288 | 0.13929304298798015

L\_inf Model |3.4886140608303133 | 3.2669949743262863 | 4.178791289639404

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02136743412476164 | 0.02149674902032545 | 0.02508195487182669

L\_1 Model |0.1634146559150387 | 0.1629548866173102 | 0.1785591515348317

L\_inf Model |163.4146559150387 | 162.9548866173102 | 178.5591515348317

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.013051520924531738 | 0.01602569002487886 | 0.021229308419666706

L\_1 Model |0.12255708244003279 | 0.11094635405417116 | 0.17506688472448004

L\_inf Model |3.6767124732009835 | 3.328390621625135 | 5.252006541734401

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.021680098303547248 | 0.0217313936226797 | 0.02288702745647245

L\_1 Model |0.16839334563239403 | 0.16820721397952385 | 0.1724605472268547

L\_inf Model |168.39334563239402 | 168.20721397952386 | 172.4605472268547

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.015230585533031112 | 0.016052676130737922 | 0.03383360129465273

L\_1 Model |0.12540665770686232 | 0.11868313065667424 | 0.23412016891482187

L\_inf Model |3.7621997312058695 | 3.560493919700227 | 7.023605067444656

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020597791373749237 | 0.020660041573519007 | 0.025976212650909984

L\_1 Model |0.16164353683965044 | 0.16145117085490335 | 0.18435071653876753

L\_inf Model |161.64353683965044 | 161.45117085490335 | 184.35071653876753

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.016912914897551132 | 0.020671372217000997 | 0.019026175586283153

L\_1 Model |0.14693597072225117 | 0.13786810510919562 | 0.16648746012627302

L\_inf Model |4.4080791216675355 | 4.136043153275868 | 4.9946238037881905

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019083270038514892 | 0.01915651955807651 | 0.021404688029841402

L\_1 Model |0.1533905392845693 | 0.1530946491118694 | 0.16554520156759908

L\_inf Model |153.39053928456931 | 153.09464911186942 | 165.5452015675991

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.013723693906968672 | 0.014733303676033352 | 0.021763337603765386

L\_1 Model |0.13427436299653678 | 0.13066550686980838 | 0.18476040986730563

L\_inf Model |4.028230889896103 | 3.9199652060942514 | 5.542812296019169

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020637932343063624 | 0.020733869639916788 | 0.024349830719806082

L\_1 Model |0.16270619090524913 | 0.16243677266776463 | 0.17526312804618355

L\_inf Model |162.70619090524912 | 162.43677266776461 | 175.26312804618354

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02319895475139062 | 0.02936023098548153 | 0.02479326314737297

L\_1 Model |0.19158480808733186 | 0.18041977043144836 | 0.20032210357712238

L\_inf Model |5.747544242619956 | 5.41259311294345 | 6.0096631073136715

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019627027350002317 | 0.019683061131496277 | 0.021026553634504536

L\_1 Model |0.15860899997735142 | 0.15836980866583567 | 0.16345478162425972

L\_inf Model |158.60899997735143 | 158.36980866583568 | 163.4547816242597

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.011410710588638358 | 0.015110213818633948 | 0.01545689991408798

L\_1 Model |0.12080098786770023 | 0.10700541958516063 | 0.15395676968851682

L\_inf Model |3.6240296360310067 | 3.210162587554819 | 4.618703090655504

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018153310875722105 | 0.018203259867925634 | 0.019938233152866844

L\_1 Model |0.15291165409785656 | 0.15260801178150757 | 0.16001075254121408

L\_inf Model |152.91165409785657 | 152.60801178150757 | 160.0107525412141

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.012636556801984456 | 0.014506896179645363 | 0.01876420845278531

L\_1 Model |0.12448618488381642 | 0.11888972792721372 | 0.16478980512273483

L\_inf Model |3.734585546514493 | 3.566691837816412 | 4.943694153682045

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.022061944503824392 | 0.022164257349278686 | 0.0247253777954714

L\_1 Model |0.1688492124037648 | 0.16842098316367865 | 0.17963839325236963

L\_inf Model |168.8492124037648 | 168.42098316367864 | 179.63839325236964

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.016967733802976677 | 0.019577760567786547 | 0.023171660350710666

L\_1 Model |0.1490197832347463 | 0.13867623669275375 | 0.19298057129470897

L\_inf Model |4.470593497042389 | 4.160287100782613 | 5.789417138841269

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020115628507160893 | 0.020140084938659108 | 0.023632742327445614

L\_1 Model |0.1579125518707904 | 0.15775343314777912 | 0.17313141738948154

L\_inf Model |157.9125518707904 | 157.75343314777913 | 173.13141738948153

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018637390064418998 | 0.021042291771659748 | 0.025182362687570823

L\_1 Model |0.1590142885920043 | 0.15370590982600987 | 0.18612234807784905

L\_inf Model |4.770428657760129 | 4.611177294780296 | 5.583670442335472

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019680716133109587 | 0.019746743485291474 | 0.024701782885019487

L\_1 Model |0.1592742440304859 | 0.15897048369332661 | 0.18010134734329367

L\_inf Model |159.2742440304859 | 158.97048369332663 | 180.10134734329367

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.012988152693761542 | 0.015593325135829336 | 0.013326872353477303

L\_1 Model |0.1434568061636711 | 0.1372398314631275 | 0.14524948234389334

L\_inf Model |4.303704184910133 | 4.117194943893825 | 4.3574844703168

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020477836388927084 | 0.02063394815464747 | 0.0213478937391193

L\_1 Model |0.16227711631526145 | 0.16183485780822743 | 0.16528481530461453

L\_inf Model |162.27711631526145 | 161.83485780822744 | 165.28481530461454

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.011879801208010814 | 0.01337183175912941 | 0.01945320329340185

L\_1 Model |0.12585357568220518 | 0.11822617339245008 | 0.172036447992256

L\_inf Model |3.7756072704661556 | 3.546785201773502 | 5.16109343976768

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02013326826932213 | 0.020159561424765524 | 0.022267258961798864

L\_1 Model |0.16084066133831382 | 0.1606777529168794 | 0.16967190043696675

L\_inf Model |160.84066133831382 | 160.6777529168794 | 169.67190043696675

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018296771643987894 | 0.021317833025841025 | 0.02132971396681544

L\_1 Model |0.1547039153023414 | 0.1434282367797971 | 0.17974013231250632

L\_inf Model |4.641117459070242 | 4.302847103393913 | 5.392203969375189

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01894547959831914 | 0.01902809815558252 | 0.02212946076990061

L\_1 Model |0.15528680084316898 | 0.15511074985085 | 0.16831689049167156

L\_inf Model |155.28680084316898 | 155.11074985085 | 168.31689049167156

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02066032646986726 | 0.024334296173379715 | 0.022335301146537763

L\_1 Model |0.1671250979907497 | 0.16497685003423096 | 0.17363311687603616

L\_inf Model |5.013752939722491 | 4.949305501026929 | 5.208993506281085

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.021155537290622375 | 0.021265889334404455 | 0.024629826097550427

L\_1 Model |0.16356201648602364 | 0.16311009800957385 | 0.17679917080301347

L\_inf Model |163.56201648602365 | 163.11009800957385 | 176.79917080301345

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02051850469557437 | 0.02306682737292338 | 0.03192851601058286

L\_1 Model |0.1507086325707789 | 0.14570108202552137 | 0.21811418238640357

L\_inf Model |4.5212589771233676 | 4.371032460765641 | 6.543425471592107

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01976920943330998 | 0.019828880270717112 | 0.02264172361849981

L\_1 Model |0.16075489820430164 | 0.16046211865176413 | 0.17278662194509076

L\_inf Model |160.75489820430164 | 160.46211865176414 | 172.78662194509076

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01971487781403823 | 0.021206323173935393 | 0.023113946279782478

L\_1 Model |0.15802948804334732 | 0.14931907324644253 | 0.17308048488809796

L\_inf Model |4.740884641300419 | 4.479572197393276 | 5.192414546642939

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01933785964906126 | 0.019389728096225085 | 0.022677934182296983

L\_1 Model |0.156371763798297 | 0.15616114452747654 | 0.16996812096714123

L\_inf Model |156.371763798297 | 156.16114452747652 | 169.96812096714123

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.0145004432522264 | 0.01646433729880497 | 0.018712187720127074

L\_1 Model |0.14702517004066962 | 0.13478136200573754 | 0.17409058913425465

L\_inf Model |4.410755101220088 | 4.043440860172126 | 5.222717674027639

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019159104022239703 | 0.019285903154970675 | 0.024097402509328198

L\_1 Model |0.1576601545253367 | 0.1573227482638951 | 0.17698529835821453

L\_inf Model |157.66015452533668 | 157.3227482638951 | 176.98529835821452

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.015549001374093903 | 0.017218333812258922 | 0.024458986107272398

L\_1 Model |0.14168187813249034 | 0.13005761318228043 | 0.1819932680647362

L\_inf Model |4.25045634397471 | 3.9017283954684125 | 5.459798041942086

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01978805571349212 | 0.01982421027306172 | 0.02808710289110733

L\_1 Model |0.15862192997818914 | 0.1584895842980625 | 0.19043530962826097

L\_inf Model |158.62192997818914 | 158.4895842980625 | 190.43530962826097

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.014410545759534031 | 0.015495737351051768 | 0.020852100296210728

L\_1 Model |0.12384816807912209 | 0.11903248749243237 | 0.18165437623219116

L\_inf Model |3.7154450423736627 | 3.570974624772971 | 5.449631286965735

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01945309185050201 | 0.019515409445183044 | 0.024855681606709347

L\_1 Model |0.15518292737095737 | 0.15495519607483457 | 0.17722363725888016

L\_inf Model |155.18292737095737 | 154.95519607483456 | 177.22363725888016

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.015234116018721811 | 0.01633614457737872 | 0.016986924183464414

L\_1 Model |0.14363479807085908 | 0.1374107008373534 | 0.16429025918692877

L\_inf Model |4.309043942125772 | 4.122321025120602 | 4.928707775607863

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020000682103627227 | 0.02004324689409589 | 0.0268070836012368

L\_1 Model |0.15668775574763014 | 0.15646625530235428 | 0.1840648143232776

L\_inf Model |156.68775574763015 | 156.46625530235428 | 184.0648143232776

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019249776851256957 | 0.02167582121243036 | 0.025188801011301186

L\_1 Model |0.16556864191358972 | 0.15127446670084208 | 0.1932580959557986

L\_inf Model |4.9670592574076915 | 4.538234001025263 | 5.797742878673958

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01982829275497404 | 0.019866691423215725 | 0.02551107359836104

L\_1 Model |0.1585079989632902 | 0.15831769573726495 | 0.1797486251738897

L\_inf Model |158.50799896329022 | 158.31769573726496 | 179.7486251738897

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.015218662733003313 | 0.015999915720463928 | 0.018835981452782695

L\_1 Model |0.13428287507717032 | 0.12934973954728618 | 0.15878097153284976

L\_inf Model |4.028486252315109 | 3.8804921864185857 | 4.7634291459854925

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018607674899405492 | 0.018624462349836543 | 0.020678636803419657

L\_1 Model |0.15408312752441697 | 0.15404428706066228 | 0.162040911228006

L\_inf Model |154.08312752441697 | 154.04428706066227 | 162.04091122800602

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.021702338092502608 | 0.02522234758578412 | 0.025396990583067606

L\_1 Model |0.17743721212872085 | 0.1629406438358205 | 0.19206896537832824

L\_inf Model |5.323116363861626 | 4.888219315074615 | 5.762068961349847

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019924465484705584 | 0.019962291226136415 | 0.023853144853818634

L\_1 Model |0.15987780165587095 | 0.1597237093670153 | 0.17430520531113206

L\_inf Model |159.87780165587097 | 159.7237093670153 | 174.30520531113206

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.017892213025391103 | 0.020047791983506726 | 0.023729916993718497

L\_1 Model |0.14878906086972316 | 0.13437533159533005 | 0.18782336573996006

L\_inf Model |4.463671826091694 | 4.031259947859901 | 5.634700972198802

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019132688309802838 | 0.019302820294533566 | 0.021254737962056836

L\_1 Model |0.15599694479977202 | 0.15542263916321056 | 0.16471498652256122

L\_inf Model |155.99694479977202 | 155.42263916321056 | 164.71498652256122

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02189545216184751 | 0.023705101866113472 | 0.023506814507597472

L\_1 Model |0.1783002330026714 | 0.1730450296060498 | 0.18895614164347913

L\_inf Model |5.349006990080142 | 5.1913508881814945 | 5.668684249304373

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02022110723888837 | 0.020274556470250574 | 0.024505509442443292

L\_1 Model |0.15931087222949078 | 0.15913433910142907 | 0.17737765430340338

L\_inf Model |159.31087222949077 | 159.13433910142908 | 177.37765430340338

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.016665087011594526 | 0.01933734693959089 | 0.01930133246763229

L\_1 Model |0.1532581739575972 | 0.14125202946408452 | 0.16906410099365962

L\_inf Model |4.5977452187279155 | 4.237560883922535 | 5.071923029809788

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.0198981819563136 | 0.019922062503307768 | 0.020572788016541825

L\_1 Model |0.16090172812631803 | 0.16081184684892522 | 0.16422548696773043

L\_inf Model |160.90172812631803 | 160.81184684892523 | 164.22548696773043

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018395087415808746 | 0.02176688924929463 | 0.021597770472362863

L\_1 Model |0.15595083731971174 | 0.14376913734199662 | 0.17554148143558965

L\_inf Model |4.678525119591352 | 4.313074120259898 | 5.26624444306769

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020641845406021524 | 0.02068901263944924 | 0.02184232729886155

L\_1 Model |0.1647432227737946 | 0.16460656518880476 | 0.1685235716956425

L\_inf Model |164.7432227737946 | 164.60656518880475 | 168.5235716956425

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.012463765521589542 | 0.013742905040100707 | 0.015855807473978484

L\_1 Model |0.12403834383115561 | 0.11323301946019972 | 0.15003387953989883

L\_inf Model |3.7211503149346683 | 3.3969905838059913 | 4.501016386196965

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019078185554156083 | 0.019164912295524816 | 0.021939662808778786

L\_1 Model |0.15600726506626225 | 0.15567449256151195 | 0.16715648319861773

L\_inf Model |156.00726506626225 | 155.67449256151195 | 167.15648319861774

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.024136397551685684 | 0.02755953862413715 | 0.038575476792060207

L\_1 Model |0.16750427339777907 | 0.15249246662937938 | 0.25762900649696224

L\_inf Model |5.025128201933372 | 4.574773998881382 | 7.7288701949088665

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019668229111727022 | 0.019720745091819487 | 0.022153345921301946

L\_1 Model |0.1593903157133856 | 0.15912074048571315 | 0.16918243487799356

L\_inf Model |159.3903157133856 | 159.12074048571316 | 169.18243487799356

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.010896438991670542 | 0.01149389722358705 | 0.012140990331455123

L\_1 Model |0.11951574395819056 | 0.11303512591503971 | 0.12992829230006786

L\_inf Model |3.5854723187457167 | 3.3910537774511913 | 3.8978487690020356

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019164619922347737 | 0.019218324316147357 | 0.023444084039323437

L\_1 Model |0.15564804883121258 | 0.15544472652253388 | 0.17445301787652118

L\_inf Model |155.64804883121258 | 155.44472652253387 | 174.4530178765212

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.014999173860505633 | 0.01647441203271083 | 0.02166590352616552

L\_1 Model |0.12757990591697366 | 0.11743332040421967 | 0.17598088803740836

L\_inf Model |3.8273971775092095 | 3.52299961212659 | 5.27942664112225

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018563650161040182 | 0.018633414084809688 | 0.023152184398583495

L\_1 Model |0.15464745391819743 | 0.1542141045081248 | 0.17292185045924205

L\_inf Model |154.6474539181974 | 154.21410450812482 | 172.92185045924205

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01743070839429687 | 0.021885416646320276 | 0.02402256618055198

L\_1 Model |0.14898417864205624 | 0.1328372487116067 | 0.18801090564331238

L\_inf Model |4.469525359261687 | 3.9851174613482017 | 5.640327169299371

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019910192740239946 | 0.019995633596891667 | 0.024341469932078988

L\_1 Model |0.1577792143595792 | 0.1574672422711603 | 0.17562882302227337

L\_inf Model |157.7792143595792 | 157.4672422711603 | 175.62882302227337

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.021631972299723946 | 0.02464361204457324 | 0.02858957597052702

L\_1 Model |0.16326686864077125 | 0.1532965805895999 | 0.20046277014773603

L\_inf Model |4.898006059223137 | 4.598897417687997 | 6.013883104432081

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.017791906165024178 | 0.017816349176864737 | 0.018584144006080546

L\_1 Model |0.14720818847662584 | 0.1471002344147896 | 0.1504316314035534

L\_inf Model |147.20818847662585 | 147.10023441478958 | 150.43163140355338

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.007676487580082631 | 0.008539592707031533 | 0.011826462207424972

L\_1 Model |0.09957918129959377 | 0.09304880532731487 | 0.1315207407040602

L\_inf Model |2.987375438987813 | 2.791464159819446 | 3.9456222211218055

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.021283165150503693 | 0.021317520764528912 | 0.02346928204300176

L\_1 Model |0.16493516394278585 | 0.1648177331264314 | 0.17346788350011652

L\_inf Model |164.93516394278583 | 164.8177331264314 | 173.46788350011653

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018642587661358498 | 0.02001829309830958 | 0.024125806674729574

L\_1 Model |0.15567036086371305 | 0.1524182817878216 | 0.1927326475049339

L\_inf Model |4.670110825911392 | 4.572548453634648 | 5.781979425148017

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020180799149149487 | 0.020320438857294496 | 0.022039922824541994

L\_1 Model |0.1628152895745409 | 0.16228214592026274 | 0.17120876553548878

L\_inf Model |162.81528957454088 | 162.28214592026274 | 171.20876553548877

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.012865894449836377 | 0.014414279938540532 | 0.019452161055759195

L\_1 Model |0.13043305250188955 | 0.11795181328724133 | 0.17526897719371312

L\_inf Model |3.9129915750566866 | 3.53855439861724 | 5.258069315811394

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02037042517126739 | 0.020465622769380725 | 0.021617497097444066

L\_1 Model |0.16052964056666927 | 0.1602423983027416 | 0.1654438956850926

L\_inf Model |160.52964056666926 | 160.24239830274158 | 165.4438956850926

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.010226737593526429 | 0.011075121017153458 | 0.013902433474798088

L\_1 Model |0.11486876252837201 | 0.11104785308596775 | 0.14340314471763035

L\_inf Model |3.4460628758511604 | 3.3314355925790324 | 4.30209434152891

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018300592889112014 | 0.01835691405099676 | 0.018935504233077053

L\_1 Model |0.15358876627185897 | 0.15339675942591843 | 0.15676695656329848

L\_inf Model |153.58876627185896 | 153.39675942591845 | 156.76695656329846

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.009260912037011255 | 0.010280332864352879 | 0.015871585768407546

L\_1 Model |0.10858292627414928 | 0.10498921020935302 | 0.15822525404900772

L\_inf Model |3.2574877882244784 | 3.1496763062805906 | 4.746757621470231

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018240168775122047 | 0.0182870257431411 | 0.02290368628885629

L\_1 Model |0.1535094200814326 | 0.15326835949636258 | 0.17114399583432113

L\_inf Model |153.50942008143258 | 153.2683594963626 | 171.14399583432115

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02220277029661513 | 0.02584031645344972 | 0.0502212329603811

L\_1 Model |0.15667850641528322 | 0.15169452169877265 | 0.27185727280698824

L\_inf Model |4.700355192458496 | 4.5508356509631795 | 8.155718184209647

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01988313316528747 | 0.020027368616006104 | 0.02155888351241826

L\_1 Model |0.1575542407929551 | 0.1571006272833868 | 0.1645588145400685

L\_inf Model |157.5542407929551 | 157.1006272833868 | 164.55881454006848

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.014421472616411465 | 0.016977330338835004 | 0.019249631273132742

L\_1 Model |0.13547743877935195 | 0.12973967982026366 | 0.17153352374126118

L\_inf Model |4.064323163380559 | 3.8921903946079097 | 5.1460057122378355

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02267065548708689 | 0.022808134040632134 | 0.024365055418285028

L\_1 Model |0.17062743818948659 | 0.17016398537839075 | 0.17696758756668993

L\_inf Model |170.62743818948658 | 170.16398537839075 | 176.96758756668993

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.021215086344610765 | 0.030223260414757983 | 0.025197856811696347

L\_1 Model |0.1811874806866053 | 0.1735890516054824 | 0.1936347368094837

L\_inf Model |5.43562442059816 | 5.207671548164472 | 5.8090421042845115

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01847228128726503 | 0.018499607061679312 | 0.022132731362703963

L\_1 Model |0.15475535736425494 | 0.15458744157437543 | 0.17040967563475884

L\_inf Model |154.75535736425493 | 154.58744157437542 | 170.40967563475883

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018675642445324023 | 0.022799453976208896 | 0.021249237484277327

L\_1 Model |0.15840403778152898 | 0.14653847119027047 | 0.1783386597576561

L\_inf Model |4.752121133445869 | 4.396154135708114 | 5.3501597927296825

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01973915612602359 | 0.01978525769133186 | 0.021245832115962135

L\_1 Model |0.15868821520551887 | 0.158453491326447 | 0.16555146851689076

L\_inf Model |158.68821520551887 | 158.45349132644702 | 165.55146851689076

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02293251134122 | 0.025397594673272947 | 0.04027351889424092

L\_1 Model |0.16517167440591363 | 0.154894072032652 | 0.25290606895517576

L\_inf Model |4.955150232177409 | 4.646822160979561 | 7.587182068655272

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019262919498569283 | 0.019335053497993686 | 0.024852463405811466

L\_1 Model |0.15669056895360495 | 0.156468068616311 | 0.17941513379774102

L\_inf Model |156.69056895360495 | 156.468068616311 | 179.41513379774102

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01325503741124588 | 0.015464410132610731 | 0.019444151845486705

L\_1 Model |0.12931153046540325 | 0.1140333515913906 | 0.17703899688388977

L\_inf Model |3.8793459139620974 | 3.421000547741718 | 5.311169906516693

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02093531043895335 | 0.020979305442617516 | 0.022591387398251693

L\_1 Model |0.1628642859246258 | 0.16255642121448385 | 0.17080883344149406

L\_inf Model |162.8642859246258 | 162.55642121448386 | 170.80883344149404

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.00731219014310136 | 0.008170934327541112 | 0.008280707490287352

L\_1 Model |0.10307037765003084 | 0.0955193656946656 | 0.11230224439893213

L\_inf Model |3.0921113295009253 | 2.865580970839968 | 3.369067331967964

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02019884426718423 | 0.02025993821346957 | 0.02235793158861904

L\_1 Model |0.16129303001282844 | 0.1610523580629436 | 0.16886848442409966

L\_inf Model |161.29303001282844 | 161.0523580629436 | 168.86848442409965

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020260982026113556 | 0.028330768188348288 | 0.029569912276552663

L\_1 Model |0.1492884416893378 | 0.1253787611597231 | 0.2082780713008079

L\_inf Model |4.478653250680134 | 3.761362834791693 | 6.248342139024237

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01956095135176804 | 0.019612467225806222 | 0.022908676182677615

L\_1 Model |0.1577077184100725 | 0.15742862351273054 | 0.17006363698262444

L\_inf Model |157.7077184100725 | 157.42862351273055 | 170.06363698262444

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01340753580099246 | 0.01598899975716276 | 0.02082383287608217

L\_1 Model |0.13778452945078418 | 0.12550148110449652 | 0.18151797303406053

L\_inf Model |4.133535883523526 | 3.7650444331348956 | 5.445539191021816

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.0189337528611844 | 0.018976438518614292 | 0.02206798711003176

L\_1 Model |0.15514417631047098 | 0.1549863978124863 | 0.1686206397750051

L\_inf Model |155.14417631047098 | 154.9863978124863 | 168.62063977500512

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.012575106315398061 | 0.013484720542825505 | 0.01448383674737769

L\_1 Model |0.1342100108958772 | 0.12609891221409195 | 0.15235516443225358

L\_inf Model |4.026300326876316 | 3.782967366422758 | 4.570654932967607

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019471325962201386 | 0.019548782311113642 | 0.022447593521861715

L\_1 Model |0.15478938764824562 | 0.15442358001644338 | 0.16693458104740957

L\_inf Model |154.78938764824562 | 154.42358001644337 | 166.93458104740958

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.016453368497315594 | 0.017744454452130184 | 0.020173813620290398

L\_1 Model |0.1488542319744318 | 0.14309015821585203 | 0.1745457024424775

L\_inf Model |4.465626959232954 | 4.292704746475561 | 5.2363710732743245

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01832766839189119 | 0.018381684700945474 | 0.019398262024238284

L\_1 Model |0.15256588653419376 | 0.15231940155604323 | 0.1580627267186584

L\_inf Model |152.56588653419377 | 152.31940155604323 | 158.0627267186584

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018326357430113686 | 0.019887402011460154 | 0.02445502267020485

L\_1 Model |0.1494618503182919 | 0.14279813978997694 | 0.18879570154670924

L\_inf Model |4.483855509548757 | 4.283944193699308 | 5.663871046401277

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02058546048009225 | 0.02070300733107181 | 0.02985364372736991

L\_1 Model |0.16274790053546667 | 0.1622702667998228 | 0.1950208385206415

L\_inf Model |162.74790053546667 | 162.2702667998228 | 195.0208385206415

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02232958705158429 | 0.02608763906815111 | 0.03355029236896926

L\_1 Model |0.1729497377177192 | 0.165111628179511 | 0.22798980123630078

L\_inf Model |5.188492131531576 | 4.95334884538533 | 6.839694037089023

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02036048342409574 | 0.020425728176678093 | 0.021708676170754306

L\_1 Model |0.16241458778577775 | 0.16203798789120744 | 0.1661627897229586

L\_inf Model |162.41458778577774 | 162.03798789120742 | 166.1627897229586

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.015464924356253058 | 0.015672833934431286 | 0.020962212563825057

L\_1 Model |0.13633425213403333 | 0.1335258292455894 | 0.18176351851834477

L\_inf Model |4.090027564021 | 4.005774877367682 | 5.452905555550343

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018822352955328213 | 0.018858430736725842 | 0.021332442077822245

L\_1 Model |0.15511659861247307 | 0.15484805979980745 | 0.16496505126033256

L\_inf Model |155.11659861247307 | 154.84805979980746 | 164.96505126033256

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.017142281333688322 | 0.018629269842613143 | 0.02256163620113681

L\_1 Model |0.15205409714245827 | 0.14671784190795473 | 0.1905110211451346

L\_inf Model |4.561622914273748 | 4.401535257238642 | 5.715330634354038

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020877138635673562 | 0.02096253500350962 | 0.024002585690544278

L\_1 Model |0.16431975747381775 | 0.16395742400909197 | 0.1772519853945032

L\_inf Model |164.31975747381776 | 163.95742400909197 | 177.2519853945032

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.021221285142853288 | 0.02256200126571976 | 0.030872563895701036

L\_1 Model |0.15963315453181617 | 0.14898372448629285 | 0.20880224709718911

L\_inf Model |4.788994635954485 | 4.469511734588785 | 6.264067412915673

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019413705888568618 | 0.019452660625268233 | 0.021110059531041846

L\_1 Model |0.15580619547705224 | 0.15561469520600282 | 0.163214018161097

L\_inf Model |155.80619547705223 | 155.61469520600284 | 163.21401816109702

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.009611052805368522 | 0.011957372041937801 | 0.010156503223820753

L\_1 Model |0.11677589092745601 | 0.1074277789294698 | 0.12016379485248795

L\_inf Model |3.50327672782368 | 3.222833367884094 | 3.6049138455746386

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01874280927081187 | 0.018861036773991847 | 0.02283522174277905

L\_1 Model |0.15509670726406072 | 0.1547606285728827 | 0.1736731193769552

L\_inf Model |155.0967072640607 | 154.7606285728827 | 173.6731193769552

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01941765571628717 | 0.026106354664568118 | 0.021413583836105733

L\_1 Model |0.16925702085922964 | 0.1534093925421335 | 0.18125751143425836

L\_inf Model |5.07771062577689 | 4.602281776264006 | 5.437725343027751

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.021877849025272986 | 0.021910623231331527 | 0.025744459140189527

L\_1 Model |0.16630039395472765 | 0.16611657920853506 | 0.1838071129723158

L\_inf Model |166.30039395472764 | 166.11657920853506 | 183.8071129723158

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01820648675433529 | 0.019475301640558648 | 0.0199614121602102

L\_1 Model |0.13900983648183546 | 0.13267048369846793 | 0.14937896612859358

L\_inf Model |4.170295094455064 | 3.9801145109540377 | 4.4813689838578075

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01998291713682894 | 0.0200305126997382 | 0.025115659143989843

L\_1 Model |0.15946625027343106 | 0.15929428406880947 | 0.17929760357975547

L\_inf Model |159.46625027343106 | 159.29428406880947 | 179.29760357975547

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.018450985689664766 | 0.02087267649093504 | 0.020250731975862378

L\_1 Model |0.1571344779481045 | 0.15245249193865174 | 0.16365067773261666

L\_inf Model |4.714034338443136 | 4.573574758159552 | 4.9095203319784995

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020546212458750567 | 0.020611762462718862 | 0.022710315086175166

L\_1 Model |0.16347898054210772 | 0.1632613760823652 | 0.1723273707717182

L\_inf Model |163.4789805421077 | 163.26137608236522 | 172.3273707717182

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.012538260998224924 | 0.013895753748328264 | 0.015483867542945753

L\_1 Model |0.11798899370380368 | 0.10929116951202014 | 0.14213990332185902

L\_inf Model |3.5396698111141105 | 3.278735085360604 | 4.264197099655771

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020644562033257416 | 0.020744167613941007 | 0.024420418653756364

L\_1 Model |0.16224740281800512 | 0.16191510198551226 | 0.1773493558883998

L\_inf Model |162.24740281800513 | 161.91510198551225 | 177.3493558883998

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.014883050848255376 | 0.017535087698325773 | 0.016990602094114553

L\_1 Model |0.14722864470045113 | 0.14203253923304604 | 0.15837830514427959

L\_inf Model |4.416859341013534 | 4.260976176991381 | 4.751349154328388

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020099003711389618 | 0.020196375013712818 | 0.022034123636716002

L\_1 Model |0.16272194869271706 | 0.1624966577191577 | 0.16967216137353625

L\_inf Model |162.72194869271706 | 162.4966577191577 | 169.67216137353626

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.0136741696460227 | 0.01471764930070211 | 0.01712761555472535

L\_1 Model |0.13781504559071364 | 0.13326260289213862 | 0.15996108778287324

L\_inf Model |4.134451367721409 | 3.9978780867641586 | 4.798832633486197

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019633634041891754 | 0.019699209867904837 | 0.022072202249733375

L\_1 Model |0.1584848456998949 | 0.15828531303866655 | 0.17049934711065798

L\_inf Model |158.48484569989492 | 158.28531303866654 | 170.49934711065796

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.012617784579312749 | 0.014634409479186823 | 0.015955675183031877

L\_1 Model |0.13545522653070446 | 0.12416364927752394 | 0.157196208014388

L\_inf Model |4.063656795921134 | 3.724909478325718 | 4.71588624043164

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.021161847293423974 | 0.021214443359054737 | 0.03144216725400376

L\_1 Model |0.16318352084404883 | 0.16292907513451618 | 0.2009338669405303

L\_inf Model |163.18352084404884 | 162.92907513451618 | 200.93386694053032

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01121163616597836 | 0.011956096584437192 | 0.014450725333954794

L\_1 Model |0.12182601411781875 | 0.11368820529382803 | 0.1457801954867022

L\_inf Model |3.6547804235345627 | 3.410646158814841 | 4.373405864601066

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.022285461649996674 | 0.022318869321018616 | 0.027674151288457717

L\_1 Model |0.16646826389307173 | 0.16629088731401998 | 0.19087537906643134

L\_inf Model |166.46826389307174 | 166.29088731401998 | 190.87537906643135

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.024255218524903328 | 0.028876679324237924 | 0.03269146238872828

L\_1 Model |0.18132797820556104 | 0.16116862965291773 | 0.22144725297570889

L\_inf Model |5.439839346166831 | 4.835058889587532 | 6.643417589271267

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01926795694083163 | 0.01934980466560897 | 0.030007287983810974

L\_1 Model |0.15617838831543523 | 0.15583534462550094 | 0.19751518601072132

L\_inf Model |156.17838831543523 | 155.83534462550094 | 197.51518601072132

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01642407941390444 | 0.01925072510977266 | 0.018515770228463558

L\_1 Model |0.15483367644250573 | 0.1482198068239102 | 0.1771632520842692

L\_inf Model |4.645010293275172 | 4.446594204717306 | 5.314897562528076

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019539903232658286 | 0.01955678099833178 | 0.02223026953356725

L\_1 Model |0.15461631475725443 | 0.1545372285219442 | 0.16558206842677872

L\_inf Model |154.61631475725443 | 154.5372285219442 | 165.58206842677873

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.016025211771295716 | 0.018334721101934136 | 0.021889805412379417

L\_1 Model |0.1503219529777137 | 0.13853659566944412 | 0.19133045407385774

L\_inf Model |4.509658589331411 | 4.1560978700833235 | 5.739913622215732

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020929762385419957 | 0.020975577966051342 | 0.02332256032391784

L\_1 Model |0.16463303084717876 | 0.16438821190898673 | 0.17482670808916292

L\_inf Model |164.63303084717876 | 164.38821190898673 | 174.82670808916293

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.015394597456514674 | 0.0178460426789773 | 0.017624100138851077

L\_1 Model |0.1450437320294986 | 0.13437730447460156 | 0.16559210234006785

L\_inf Model |4.351311960884958 | 4.031319134238046 | 4.967763070202036

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02011536694157592 | 0.02013040397831599 | 0.021389403941831073

L\_1 Model |0.1595135409558534 | 0.15942949345067664 | 0.1644217043398826

L\_inf Model |159.51354095585341 | 159.42949345067663 | 164.4217043398826

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.014212430783113985 | 0.018607694331048488 | 0.02102614307949245

L\_1 Model |0.13845517509548974 | 0.12661791169980274 | 0.18676048818191818

L\_inf Model |4.153655252864692 | 3.7985373509940823 | 5.602814645457546

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.019487957757413896 | 0.019587542139139952 | 0.023490403613730115

L\_1 Model |0.15421246003999886 | 0.15383892735793964 | 0.17223159029975987

L\_inf Model |154.21246003999886 | 153.83892735793964 | 172.2315902997599

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020257043417988882 | 0.022506615752218427 | 0.0221413343321421

L\_1 Model |0.16807247500780223 | 0.15757741660607477 | 0.17897020973821143

L\_inf Model |5.0421742502340665 | 4.727322498182243 | 5.369106292146343

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02000316044272387 | 0.020017756301345658 | 0.028595818027440222

L\_1 Model |0.15751410287189047 | 0.15739369537646766 | 0.1907444525806079

L\_inf Model |157.51410287189046 | 157.39369537646766 | 190.7444525806079

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.015319153784914408 | 0.016511243401664433 | 0.02170489548230377

L\_1 Model |0.12748815477375872 | 0.11982831693335408 | 0.16703901854832484

L\_inf Model |3.8246446432127614 | 3.5948495080006224 | 5.011170556449745

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.020681091247583164 | 0.020788298340718575 | 0.02363020606328643

L\_1 Model |0.16382165989881442 | 0.16355250844388874 | 0.1758101816371747

L\_inf Model |163.82165989881443 | 163.55250844388874 | 175.8101816371747

Table: Different training losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.01172373047121488 | 0.01296918189487667 | 0.02022248294213947

L\_1 Model |0.12133830889560228 | 0.11791427160145775 | 0.17715897624559848

L\_inf Model |3.640149266868068 | 3.537428148043732 | 5.314769287367954

Table: Different test losses for different models

Model | L\_2 loss | L\_2 loss | L\_inf loss

L\_2 Model |0.02041044688464786 | 0.020456250711245783 | 0.026553992242393767

L\_1 Model |0.16101776823280192 | 0.16086314626903786 | 0.18451582220849053

L\_inf Model |161.01776823280193 | 160.86314626903786 | 184.51582220849053

1e)

What I seen in the tables is that I realized that the difference in the training and test losses tables are that the test table have more larger values than in the training table. This is because in the test table we are generating a lot more test data points which causes our model to be more accurate. Also, in the L\_inf model, the values are higher than in the other models. It is because L\_inf norm is going off the maximum value, making it more normal for it to be high. It is influence by the magnitude of the highest value, so outliers very easily cause it to diverge.

2d)

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.5887073208488474 | 1 |9.685454059110057e+110 | 0.1 |-8.544383053831851e+104

50 |6.681187863472287e+78 | 2 |8.264222796482778e+103 | 1.0 |2.710698337871812e+209

100 |1.3264423843113885e+108 | 4 |1.399214959758374e+108 | 10.0 |nan

200 |9.257088690786201e+145 | 8 |-1.115438668225958e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |3.5398926683963936e+210 | 1 |3.874201127686805e+217 | 0.1 |-2.3995542115574723e+214

50 |-6.092265290698519e+211 | 2 |2.4318346396293315e+214 | 1.0 |nan

100 |6.531618010146513e+210 | 4 |1.052244909446209e+211 | 10.0 |nan

200 |7.849208802619096e+218 | 8 |-2.7914629345974663e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-3318337365.0413804 | 1 |3.166171798797541e+108 | 0.1 |4.508780977663195e+109

50 |2.7797360116551276e+77 | 2 |3.607196350170701e+113 | 1.0 |1.3575235219470153e+204

100 |-7.336865284021071e+107 | 4 |-7.02001164203185e+104 | 10.0 |7.71688856270352e+305

200 |-5.48717545945836e+141 | 8 |1.8614665094700543e+106 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |3.980400111873265e+210 | 1 |-1.2355647798273682e+221 | 0.1 |2.5293258269495735e+214

50 |2.366454113395718e+211 | 2 |5.505183994251537e+213 | 1.0 |nan

100 |9.808201183776644e+213 | 4 |1.2861791272388231e+211 | 10.0 |nan

200 |4.936166010965149e+218 | 8 |-2.4592667474051196e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2893601962.2312965 | 1 |1.1794132549973481e+111 | 0.1 |2.2901452273076475e+113

50 |-3.705764488902992e+85 | 2 |-7.795352480019006e+110 | 1.0 |1.1571776433939313e+205

100 |-5.3891732936010125e+107 | 4 |-1.0515009578453442e+108 | 10.0 |nan

200 |-8.980355782685844e+136 | 8 |1.7423283717480669e+109 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-1.0171551261277492e+208 | 1 |2.9305768985817667e+217 | 0.1 |1.1961328084877877e+213

50 |6.375877122642293e+211 | 2 |3.68075236621739e+213 | 1.0 |nan

100 |5.797072998795849e+212 | 4 |-1.2187543236836965e+212 | 10.0 |nan

200 |3.707405660964195e+218 | 8 |-1.6506216795273347e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |107392.49930268108 | 1 |2.023251110142703e+114 | 0.1 |-3.0315126380895086e+111

50 |8.038428206286163e+76 | 2 |1.9257604389927175e+110 | 1.0 |-1.929502896009264e+213

100 |3.6429957695797507e+108 | 4 |1.3828508737098547e+107 | 10.0 |nan

200 |-1.2886023254284662e+135 | 8 |1.3486832206866585e+106 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-2.0864123594224526e+209 | 1 |5.923967675270361e+218 | 0.1 |-2.647238793252183e+214

50 |9.214106087317534e+211 | 2 |2.8130425159078393e+215 | 1.0 |nan

100 |-3.0975112776816678e+212 | 4 |1.6044731437039901e+211 | 10.0 |nan

200 |3.8370717189394284e+216 | 8 |3.198155354712311e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |0.8806395024314195 | 1 |5.62337913758968e+114 | 0.1 |3.451683832318016e+103

50 |1.1263503627056595e+82 | 2 |-2.572769313820822e+110 | 1.0 |7.052244394804065e+214

100 |1.206663493917033e+116 | 4 |-3.9728119054783846e+100 | 10.0 |2.820695859868985e+307

200 |-1.9418084651194348e+142 | 8 |3.535002289781789e+106 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-7.849990509766216e+208 | 1 |-2.6863715107312336e+219 | 0.1 |5.285532934298051e+214

50 |-1.5137229109368103e+211 | 2 |-3.768247995947978e+215 | 1.0 |nan

100 |1.4880774603231338e+212 | 4 |2.650023365220947e+210 | 10.0 |nan

200 |2.2801852272284108e+216 | 8 |3.4594232544168816e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |7169186344.458417 | 1 |8.648674473840346e+112 | 0.1 |4.121593375930555e+107

50 |1.8406326897815568e+86 | 2 |-5.136207851112755e+109 | 1.0 |2.2931719964554517e+210

100 |3.0911635444732773e+114 | 4 |7.732866276359234e+100 | 10.0 |nan

200 |7.360918496594647e+141 | 8 |1.8549240146478945e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-7.265877925408103e+207 | 1 |1.5448797152692444e+220 | 0.1 |2.958408790538185e+215

50 |8.854418333435177e+211 | 2 |1.6531413354282244e+213 | 1.0 |nan

100 |1.1871220490486944e+215 | 4 |2.790184303947682e+211 | 10.0 |nan

200 |1.7962305817881252e+220 | 8 |7.346571867308058e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |16831699355387.305 | 1 |2.1904248402060403e+115 | 0.1 |1.1644055635502312e+108

50 |3.6603538849979154e+76 | 2 |-1.6624087513223502e+110 | 1.0 |9.70667384054138e+202

100 |1.3330069335761428e+107 | 4 |3.115565636912145e+107 | 10.0 |nan

200 |7.69222027054193e+141 | 8 |-5.2427165324156814e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.051046737928901e+211 | 1 |4.9065539887242374e+219 | 0.1 |-1.653656851678039e+213

50 |-5.654124985922029e+210 | 2 |7.350139682038665e+211 | 1.0 |nan

100 |-1.4863750228315118e+214 | 4 |5.1191269277506893e+210 | 10.0 |nan

200 |-2.389906198734426e+218 | 8 |-4.4050371417359185e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |604311584.5062443 | 1 |4.765646026329451e+108 | 0.1 |1.6342563739725154e+113

50 |1.4002301159201677e+79 | 2 |1.74064274492961e+107 | 1.0 |7.290943358391204e+206

100 |5.263067743185593e+102 | 4 |6.070961962595482e+109 | 10.0 |nan

200 |2.7991295952958515e+139 | 8 |-8.091620930121673e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |9.11299713820462e+210 | 1 |3.128123255128227e+216 | 0.1 |-6.016825260606337e+213

50 |-1.5177524259339362e+212 | 2 |1.0302360944691367e+216 | 1.0 |nan

100 |-7.262099196146719e+213 | 4 |9.751703706378137e+210 | 10.0 |nan

200 |-9.134196693946885e+218 | 8 |-4.851825641379698e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.9213669396576416 | 1 |1.0889921853777e+109 | 0.1 |4.193558815286653e+111

50 |-3.797729510521591e+78 | 2 |4.152797683150825e+109 | 1.0 |1.222285595272223e+209

100 |1.9350171191799553e+103 | 4 |2.0626817998779343e+104 | 10.0 |8.775151258826349e+304

200 |-1.297090743320634e+147 | 8 |-6.915344458132493e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.9345517841500225e+208 | 1 |4.024223352144386e+218 | 0.1 |8.675006463806541e+213

50 |3.1838202208019717e+211 | 2 |2.9888995951769885e+214 | 1.0 |nan

100 |8.402143598777942e+212 | 4 |6.853861329494663e+209 | 10.0 |nan

200 |7.989022062043416e+215 | 8 |6.107049126319675e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-1.2158129084230158e+37 | 1 |-1.4588951973536691e+121 | 0.1 |2.713803304589833e+109

50 |-9.503662539638893e+75 | 2 |9.279016782561824e+113 | 1.0 |1.5040388982981859e+208

100 |1.586212234542708e+106 | 4 |3.7236351656455037e+105 | 10.0 |nan

200 |1.361740270081602e+144 | 8 |1.0967412564621194e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.4570704920739101e+209 | 1 |2.307302537484164e+217 | 0.1 |-1.2564702067557648e+215

50 |1.284207519250175e+211 | 2 |5.3535119797824e+213 | 1.0 |nan

100 |6.735949987892597e+213 | 4 |-3.378422757447662e+210 | 10.0 |nan

200 |3.231816463011576e+218 | 8 |3.4776581412881606e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |369.30154935034744 | 1 |2.2929618409431954e+114 | 0.1 |-6.8678779097729604e+112

50 |-1.4600620926110207e+90 | 2 |3.2441951084455915e+112 | 1.0 |6.0218615505159605e+208

100 |4.056319748890233e+113 | 4 |-4.91576298143231e+110 | 10.0 |7.128723296597454e+300

200 |7.343582558530846e+136 | 8 |3.0773573560178103e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |3.0553463482842495e+209 | 1 |5.727152937708778e+217 | 0.1 |7.242680421504799e+211

50 |1.791273588533185e+209 | 2 |6.698400772240744e+216 | 1.0 |nan

100 |2.637888973470054e+214 | 4 |5.362010610364563e+212 | 10.0 |nan

200 |1.0062654705439546e+217 | 8 |4.51422202693039e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-6.613615170608433e+16 | 1 |3.636108310071247e+107 | 0.1 |1.3228798845283705e+108

50 |7.825675471999792e+86 | 2 |-7.907676858694154e+111 | 1.0 |1.5717353742435615e+210

100 |-1.0231438174019459e+109 | 4 |1.8902635794979782e+107 | 10.0 |nan

200 |8.673341591101821e+138 | 8 |3.908486745465365e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.7218360130071745e+212 | 1 |1.7486954899078707e+217 | 0.1 |9.130514752784957e+215

50 |1.109247365412263e+208 | 2 |5.675071366886216e+214 | 1.0 |nan

100 |2.7405801781388564e+212 | 4 |-2.857043007404294e+210 | 10.0 |nan

200 |2.2322209828721466e+219 | 8 |5.834405576790464e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |93765.50874636958 | 1 |5.591623561240852e+115 | 0.1 |1.8665036772152743e+108

50 |5.474092890215978e+81 | 2 |4.1953167007143043e+104 | 1.0 |-8.54754834614123e+214

100 |1.6940933148244257e+113 | 4 |2.7943629163097755e+112 | 10.0 |1.0269811510907708e+307

200 |4.8134180108812044e+138 | 8 |5.5632919507235984e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |4.162054515785872e+207 | 1 |-1.0107209325887732e+220 | 0.1 |-1.551240971677313e+213

50 |-2.440739320815385e+212 | 2 |-4.3275515220302434e+212 | 1.0 |nan

100 |1.0214416791741383e+214 | 4 |1.1218162332069701e+209 | 10.0 |nan

200 |2.8478759007582135e+218 | 8 |1.0217709028623184e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |281053.6892717875 | 1 |5.645626436971277e+111 | 0.1 |5.918078726822386e+112

50 |-4.4842570911100263e+73 | 2 |5.320968802608037e+108 | 1.0 |4.789168087923847e+205

100 |9.845635966644243e+104 | 4 |2.931309934417528e+110 | 10.0 |nan

200 |6.777737240847135e+139 | 8 |7.006271666891828e+100 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-6.550946717218225e+210 | 1 |6.42093691031691e+217 | 0.1 |-2.0928712242286923e+214

50 |-1.0127742946069809e+211 | 2 |2.54630935660651e+213 | 1.0 |nan

100 |1.4886830392030376e+212 | 4 |-3.524732988256185e+209 | 10.0 |nan

200 |3.269008876202718e+218 | 8 |3.416337185103225e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-1.7411448468616157e+26 | 1 |7.788495751752272e+110 | 0.1 |1.3477699699231438e+118

50 |-6.602253990349314e+74 | 2 |-1.536240861339966e+108 | 1.0 |2.332964714724819e+202

100 |-1.0359795721416109e+111 | 4 |1.0374121538805898e+104 | 10.0 |nan

200 |1.464593455754941e+138 | 8 |1.1150805981974087e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.022360393823991e+209 | 1 |9.78054272589155e+216 | 0.1 |-2.764253585313786e+214

50 |2.063383113331989e+212 | 2 |7.468935963226194e+214 | 1.0 |nan

100 |3.5350446318700576e+214 | 4 |5.033577207129591e+210 | 10.0 |nan

200 |-1.7165487601080114e+218 | 8 |2.0070746927034743e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |4.2232543210113676e+21 | 1 |1.315384268233333e+108 | 0.1 |6.544747501168845e+110

50 |2.6580852886768416e+79 | 2 |-4.674470909301346e+110 | 1.0 |-2.693747753752943e+201

100 |-3.1838828096894275e+109 | 4 |-6.873449921082053e+104 | 10.0 |9.285412314020703e+306

200 |1.32032147034131e+139 | 8 |9.844447608996356e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-1.35836562407335e+210 | 1 |1.1177644950917274e+219 | 0.1 |-8.436808598920422e+213

50 |1.5945489511751028e+213 | 2 |4.365942154149638e+214 | 1.0 |nan

100 |-1.1184438914754177e+216 | 4 |3.558188179927858e+211 | 10.0 |nan

200 |-3.8687146062733124e+218 | 8 |8.92796504347209e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |6.004562332051188e+19 | 1 |1.5492059244547805e+117 | 0.1 |1.806185734240266e+109

50 |-1.0661866672818016e+82 | 2 |1.696698562084778e+108 | 1.0 |1.1410016201781382e+214

100 |5.005379384186941e+110 | 4 |2.284474418146504e+104 | 10.0 |5.946013130444508e+300

200 |9.480757368791391e+137 | 8 |-2.426680625277647e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-6.330147690445354e+207 | 1 |6.460173730482653e+216 | 0.1 |6.565702858109736e+212

50 |1.1959313927514087e+210 | 2 |-2.6805803206344567e+214 | 1.0 |nan

100 |-9.877670791508315e+213 | 4 |6.233657905821862e+209 | 10.0 |nan

200 |1.8390672640608714e+219 | 8 |8.407436719003657e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |4.463864438462133e+17 | 1 |-2.2994293350976792e+115 | 0.1 |2.0602014693139175e+110

50 |2.4518736071830122e+73 | 2 |3.611264330407994e+104 | 1.0 |-7.183071168454619e+207

100 |5.1043433579896994e+109 | 4 |4.846879378229176e+109 | 10.0 |nan

200 |3.4774105543266276e+139 | 8 |5.7709329019467184e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.184400423183505e+210 | 1 |-1.400069004604957e+217 | 0.1 |2.2731287880053356e+212

50 |2.978914066261042e+211 | 2 |1.279711897554009e+213 | 1.0 |nan

100 |2.8838329380073256e+213 | 4 |5.690180558628837e+210 | 10.0 |nan

200 |9.987767146048197e+216 | 8 |1.3741338388989592e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |5.416661291547309e+28 | 1 |-4.4562476857954064e+108 | 0.1 |-1.3841717747749747e+108

50 |3.0397970749549805e+72 | 2 |1.1147903772147834e+112 | 1.0 |-9.426090419059176e+207

100 |2.7178090530671017e+104 | 4 |-2.5521788178587846e+102 | 10.0 |nan

200 |7.567472162832822e+135 | 8 |2.24479689231959e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.7696698816820767e+210 | 1 |2.1409649595058696e+221 | 0.1 |1.6940814763926445e+214

50 |1.8449120432180036e+211 | 2 |7.8992811982243e+212 | 1.0 |nan

100 |-5.137600112268776e+214 | 4 |-2.7808718341841046e+210 | 10.0 |nan

200 |5.19271968526884e+219 | 8 |-1.6817045450857308e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.3081517575880001e+21 | 1 |1.275832215457325e+117 | 0.1 |4.486809063271805e+112

50 |6.904284166852322e+74 | 2 |3.9912626719178105e+107 | 1.0 |-1.4110303430508433e+208

100 |-8.495042357197093e+104 | 4 |6.748545228130605e+101 | 10.0 |3.621219300988426e+302

200 |-1.4553813962741192e+145 | 8 |4.223424900655802e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |6.299122428670017e+210 | 1 |4.444501808544312e+218 | 0.1 |3.05906638616829e+213

50 |-8.40725265096572e+212 | 2 |9.875077764774303e+213 | 1.0 |nan

100 |-1.960400108375862e+213 | 4 |-9.045174176831048e+210 | 10.0 |nan

200 |2.787960328457965e+218 | 8 |-2.93964519595011e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |5.971280466861903e+16 | 1 |-9.543520662744071e+114 | 0.1 |-1.1904511180946586e+112

50 |-9.065774877191779e+80 | 2 |4.3954484618606835e+108 | 1.0 |3.122511258791159e+208

100 |2.6111564744706916e+110 | 4 |3.7050434737996586e+105 | 10.0 |-inf

200 |8.419788350625676e+140 | 8 |1.664052932136176e+107 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |6.5545010985415465e+209 | 1 |3.887616770139285e+219 | 0.1 |7.108531614400062e+212

50 |3.84468874402702e+209 | 2 |2.0601701668113042e+215 | 1.0 |nan

100 |3.5125150936876847e+214 | 4 |7.457275969545165e+210 | 10.0 |nan

200 |-1.444508139570283e+218 | 8 |1.6144107817204119e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |96080953602771.42 | 1 |-2.8856252522796293e+111 | 0.1 |7.976899098554747e+105

50 |2.901886840459323e+80 | 2 |-2.0954035273116273e+111 | 1.0 |2.6022679607400807e+211

100 |-3.6422325320145064e+109 | 4 |4.2034436203032323e+107 | 10.0 |nan

200 |-4.9125478066632345e+144 | 8 |3.89851294694231e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.1643058351439338e+208 | 1 |3.27699898110098e+220 | 0.1 |9.513158949294682e+213

50 |7.822256053476066e+210 | 2 |-6.139011785140284e+213 | 1.0 |nan

100 |7.356531409251062e+211 | 4 |-1.7311281442287552e+211 | 10.0 |nan

200 |1.8655079973969728e+215 | 8 |1.0196648554552669e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.0806436752361614 | 1 |-8.783773344799516e+122 | 0.1 |-4.727471646490108e+110

50 |2.55455457682823e+66 | 2 |1.9571558329299465e+109 | 1.0 |1.019694057727286e+209

100 |2.7954559472035576e+109 | 4 |6.864728302997345e+111 | 10.0 |nan

200 |-1.3331856187153733e+138 | 8 |1.7382459192618727e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-5.952485158625767e+210 | 1 |3.0989980890835124e+216 | 0.1 |1.1396236682174333e+215

50 |6.551282890587061e+210 | 2 |4.9603728214339904e+213 | 1.0 |nan

100 |1.7812736837398793e+211 | 4 |5.437571887476225e+210 | 10.0 |nan

200 |1.742712999849798e+218 | 8 |-8.099783088221839e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.4581409727839834 | 1 |-9.870213524123831e+116 | 0.1 |6.090673579694257e+103

50 |3.5664453137457625e+82 | 2 |-3.8932616089332465e+110 | 1.0 |8.41569374725559e+212

100 |4.0412462465080466e+101 | 4 |-6.130105602839589e+111 | 10.0 |nan

200 |-3.0486514375963774e+139 | 8 |-6.91580643173801e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.5033872084565555e+207 | 1 |8.25132061536061e+219 | 0.1 |1.0086416038793612e+214

50 |1.1835781469004135e+213 | 2 |-1.0829870080328593e+215 | 1.0 |nan

100 |2.4825890211018316e+213 | 4 |-1.1504021284198792e+211 | 10.0 |nan

200 |-6.850262523910636e+217 | 8 |-1.255738453676547e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.0459879347748593e+22 | 1 |-1.5548132173685076e+114 | 0.1 |9.411760834298559e+111

50 |-1.1667229186183717e+75 | 2 |5.8626950103773e+104 | 1.0 |3.227889899749799e+210

100 |-4.2675632120937486e+111 | 4 |6.505450535225832e+103 | 10.0 |nan

200 |1.715028982321362e+141 | 8 |1.8900052182594408e+106 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.0302270320230998e+208 | 1 |-1.6648972911765418e+218 | 0.1 |1.3022526295677732e+215

50 |6.908723272189485e+212 | 2 |1.869094607488309e+213 | 1.0 |nan

100 |6.488620472178913e+215 | 4 |4.218356271187507e+208 | 10.0 |nan

200 |1.8236235160360646e+218 | 8 |4.9096400734993924e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |0.8259596472464922 | 1 |4.5534017994871324e+113 | 0.1 |1.3983922416533835e+109

50 |-8.713205564404602e+79 | 2 |6.185154945556519e+112 | 1.0 |-1.0480319111748471e+212

100 |-1.2317504409596497e+114 | 4 |5.717568339582999e+106 | 10.0 |nan

200 |1.9453838789681044e+140 | 8 |-1.4211414275318612e+109 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |3.0131470611271984e+210 | 1 |2.069218695638281e+216 | 0.1 |-4.0455971660564436e+213

50 |2.3351190451570568e+212 | 2 |1.9148447677438194e+215 | 1.0 |nan

100 |7.781092152893194e+215 | 4 |1.779746920696795e+210 | 10.0 |nan

200 |-7.256395786793258e+216 | 8 |3.0330009375994767e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.214714684759575e+16 | 1 |-5.8879080865327996e+109 | 0.1 |1.1237494793981428e+113

50 |9.516434292321881e+75 | 2 |2.0619736235817875e+107 | 1.0 |2.1050079409442934e+212

100 |3.2558651269199755e+111 | 4 |6.088694433609547e+107 | 10.0 |nan

200 |-2.3238514268134593e+134 | 8 |-4.995020601090372e+102 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.0358586633536702e+209 | 1 |1.4167257896385341e+217 | 0.1 |-1.0011127949007629e+215

50 |-9.513553600448536e+211 | 2 |1.9155759774702817e+214 | 1.0 |nan

100 |5.559195479774964e+212 | 4 |2.4907181092959023e+209 | 10.0 |nan

200 |4.250045475640002e+218 | 8 |2.9997499179664157e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |0.5298605665065259 | 1 |3.8867896088192975e+106 | 0.1 |-1.3648512417160558e+108

50 |-4.480361267793741e+82 | 2 |1.8913344803195262e+110 | 1.0 |2.5206626072607245e+204

100 |1.943403859583451e+109 | 4 |1.9966069607499632e+101 | 10.0 |-inf

200 |9.812805474188259e+130 | 8 |5.197402083305349e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.100028656327172e+210 | 1 |1.4534942347165304e+219 | 0.1 |1.6027074417925677e+214

50 |6.556374842958114e+211 | 2 |-2.311957127250301e+213 | 1.0 |nan

100 |-5.55502602983419e+214 | 4 |-9.858502031273959e+209 | 10.0 |nan

200 |1.4128438571884584e+217 | 8 |-2.9005651626283476e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |6.662633180153943 | 1 |1.5643986099997502e+111 | 0.1 |6.457176337833762e+113

50 |3.7966841985424504e+87 | 2 |3.1053309474634935e+117 | 1.0 |2.9913855649715425e+210

100 |2.6003066205814703e+114 | 4 |-1.6626171325241104e+102 | 10.0 |nan

200 |9.41027285331256e+136 | 8 |4.070427230184081e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |7.259757060088334e+210 | 1 |1.4802657539753507e+218 | 0.1 |3.308143747502485e+213

50 |1.0147457753309175e+211 | 2 |2.0066948410929748e+216 | 1.0 |nan

100 |6.523961095487844e+212 | 4 |2.2957881249161795e+209 | 10.0 |nan

200 |2.5123868638113626e+219 | 8 |3.0478910452528077e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |14777339353034.037 | 1 |7.984475886084914e+118 | 0.1 |-1.8338244511068686e+111

50 |-1.661088823243705e+82 | 2 |5.945680638679516e+108 | 1.0 |1.0989444483257659e+212

100 |2.0921796112122514e+109 | 4 |1.79510381198169e+110 | 10.0 |2.761577293065664e+304

200 |-9.014625911131131e+143 | 8 |-6.08676889924308e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.7940645649589543e+210 | 1 |-5.763730418765131e+218 | 0.1 |8.400751404760248e+212

50 |2.92053683157505e+210 | 2 |-5.046987558929888e+214 | 1.0 |nan

100 |2.460141049044527e+211 | 4 |-7.867348303133868e+209 | 10.0 |nan

200 |-1.829395452311602e+218 | 8 |-3.230658074339598e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |4.956376263776349e+18 | 1 |1.3856175511780997e+118 | 0.1 |1.7666137086109734e+114

50 |8.604947546469333e+81 | 2 |8.341936260544356e+114 | 1.0 |8.556716997032355e+200

100 |3.234801967759543e+109 | 4 |-1.3485171869742768e+112 | 10.0 |-inf

200 |2.0939370194680003e+137 | 8 |1.4336318594252256e+109 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |8.925692878064917e+208 | 1 |-1.5512516441360568e+219 | 0.1 |3.686705957043992e+212

50 |1.2768500492375363e+211 | 2 |1.893513664323486e+213 | 1.0 |nan

100 |-2.304426762908139e+216 | 4 |7.378270356673428e+208 | 10.0 |nan

200 |8.167298458565111e+215 | 8 |-3.227834958101882e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |5.650858542839681e+23 | 1 |3.762205995187919e+120 | 0.1 |1.1971213446620274e+111

50 |-2.1562800183344038e+80 | 2 |-7.929072180030249e+110 | 1.0 |-4.906439171809778e+211

100 |1.5265774901467426e+109 | 4 |1.2676938606075887e+103 | 10.0 |nan

200 |-5.296383296434461e+139 | 8 |1.4908210646109467e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |7.60272182295116e+210 | 1 |1.9810494353465441e+217 | 0.1 |4.8056895195447226e+212

50 |-1.3822118698567983e+213 | 2 |2.0199626222091002e+214 | 1.0 |nan

100 |6.499077882141983e+213 | 4 |-1.5342239845161313e+209 | 10.0 |nan

200 |9.465902912931629e+216 | 8 |1.2631969162804098e+209 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |94540455138.14452 | 1 |3.990809601983324e+110 | 0.1 |2.3626418233662927e+111

50 |2.385658605593896e+64 | 2 |5.3669843316774415e+107 | 1.0 |-1.0569009165393628e+213

100 |9.289198296131137e+112 | 4 |5.5231773750855316e+107 | 10.0 |5.155779208331805e+305

200 |-1.6985438906292428e+142 | 8 |-2.287072734314766e+102 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |3.013310526161607e+210 | 1 |2.1882116046069156e+218 | 0.1 |1.402639473306714e+214

50 |-1.9753735983216157e+210 | 2 |5.933572358107158e+213 | 1.0 |nan

100 |5.886974536849354e+213 | 4 |2.9580302718667746e+209 | 10.0 |nan

200 |3.2108748485629163e+217 | 8 |-6.64911769775409e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-1.1379491888280188e+25 | 1 |2.9575248074270634e+111 | 0.1 |-1.263520342437962e+103

50 |-1.5360689910627052e+85 | 2 |6.969904243676614e+107 | 1.0 |1.411142907358768e+212

100 |6.164272278343707e+99 | 4 |2.202784322765544e+106 | 10.0 |-inf

200 |9.083093618328774e+144 | 8 |1.4337113698867502e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.5520125366455845e+207 | 1 |2.7879749322479744e+218 | 0.1 |-5.67096791370079e+216

50 |6.1462383132689275e+211 | 2 |-1.3341993839247395e+215 | 1.0 |nan

100 |9.536395437322947e+212 | 4 |3.015841742507014e+208 | 10.0 |nan

200 |6.69664165355444e+216 | 8 |1.9389420812491805e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.245003165499533e+25 | 1 |1.85495101352403e+116 | 0.1 |2.6054799809805314e+111

50 |9.727060304473276e+83 | 2 |1.43965562806216e+109 | 1.0 |6.265920217966832e+221

100 |3.5945815875546006e+107 | 4 |6.422081316727344e+106 | 10.0 |2.5882677228478615e+307

200 |2.7773689153180224e+140 | 8 |2.1677771452637387e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-5.6422820555021865e+212 | 1 |2.994491022849056e+219 | 0.1 |-5.36717763856509e+214

50 |-9.329977478923965e+212 | 2 |6.20461510583578e+213 | 1.0 |nan

100 |-7.330506309272866e+214 | 4 |6.866338962836471e+210 | 10.0 |nan

200 |-3.6593460360178753e+217 | 8 |4.283197043464501e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |0.12636518665254334 | 1 |8.852747961560105e+114 | 0.1 |4.221943366066231e+112

50 |6.229184251057646e+79 | 2 |7.541935258546174e+111 | 1.0 |6.567363458679411e+207

100 |-1.258426504235913e+111 | 4 |1.849979864933637e+103 | 10.0 |nan

200 |4.562412524471502e+141 | 8 |2.8228765775660424e+102 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.1655427228237805e+211 | 1 |2.8272527684251735e+220 | 0.1 |2.5202120484726854e+215

50 |3.017745944090743e+211 | 2 |-3.2538531239803553e+212 | 1.0 |nan

100 |1.0841518502644878e+215 | 4 |1.4027753180293044e+209 | 10.0 |nan

200 |8.24086862876778e+217 | 8 |2.879479332593709e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |6.559892665127857e+16 | 1 |-9.178663352516927e+123 | 0.1 |9.705325848109949e+113

50 |9.55090976858361e+77 | 2 |4.83027109463331e+106 | 1.0 |1.6986425968003927e+208

100 |7.88948883450899e+115 | 4 |6.475851306669798e+108 | 10.0 |nan

200 |1.227992103998095e+139 | 8 |1.8927202861086152e+107 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-3.5762966436168394e+209 | 1 |2.0216002721131936e+217 | 0.1 |5.329515757552972e+216

50 |4.194302535606083e+212 | 2 |1.0204329044534335e+215 | 1.0 |nan

100 |5.640798122739543e+214 | 4 |7.721657177208923e+209 | 10.0 |nan

200 |-6.375662538109223e+218 | 8 |2.4185981334012324e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |949089807705.9165 | 1 |4.4717493414878845e+110 | 0.1 |5.682724967297716e+111

50 |-5.1546513355397026e+85 | 2 |1.2447883037936401e+104 | 1.0 |-1.0915855715788622e+208

100 |6.285172872652486e+112 | 4 |4.280948765013741e+106 | 10.0 |-inf

200 |-1.512136401049104e+143 | 8 |1.6302396712210776e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-6.702967975526717e+207 | 1 |1.2162328586057143e+219 | 0.1 |1.6753738530557166e+213

50 |4.985041132020246e+212 | 2 |-6.746568243276345e+213 | 1.0 |nan

100 |-1.139763919652464e+214 | 4 |-1.19003640625925e+211 | 10.0 |nan

200 |2.616481073811317e+216 | 8 |-8.622683030967093e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-3.252162904997476e+20 | 1 |5.244879057491058e+108 | 0.1 |1.289726121068709e+106

50 |6.973548007551732e+83 | 2 |-2.132092806304701e+109 | 1.0 |-4.7744922010413295e+203

100 |-1.5358338528299207e+105 | 4 |-8.267831187310923e+103 | 10.0 |nan

200 |3.1450811264117505e+141 | 8 |6.825512567546038e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.851370024130372e+210 | 1 |1.559504503480341e+217 | 0.1 |1.365005035255471e+215

50 |8.41196531727669e+213 | 2 |7.246092183672974e+213 | 1.0 |nan

100 |9.520454129532132e+211 | 4 |5.29028274693494e+210 | 10.0 |nan

200 |-6.447420984141193e+216 | 8 |-3.1847257048697655e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |3613798848523130.5 | 1 |-3.725306032334616e+113 | 0.1 |4.1893313079570827e+105

50 |2.0929902247522256e+78 | 2 |2.444741682563173e+108 | 1.0 |1.8206546937585136e+211

100 |-6.658568730867329e+108 | 4 |5.208833392264194e+105 | 10.0 |6.066168313889218e+303

200 |-5.909759797908521e+142 | 8 |1.2747639867066503e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |3.764233623645812e+210 | 1 |1.0957276999432635e+218 | 0.1 |1.10903443879857e+216

50 |7.875529202272559e+210 | 2 |-2.4089116664488088e+216 | 1.0 |nan

100 |1.6254633581988735e+214 | 4 |7.856039456633101e+210 | 10.0 |nan

200 |2.585257245483959e+217 | 8 |-2.2371312875637034e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |9878248881349338.0 | 1 |1.5945506581029987e+113 | 0.1 |3.0033347938243606e+108

50 |1.27264411488576e+93 | 2 |-1.8250431244395668e+111 | 1.0 |-1.7011451404390948e+210

100 |-1.3311649228189559e+107 | 4 |2.211987942958182e+106 | 10.0 |nan

200 |4.672456337200252e+144 | 8 |2.455953773118838e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.5471009492432492e+211 | 1 |1.1500426816909039e+219 | 0.1 |3.2405979036598316e+211

50 |1.381175582022442e+213 | 2 |1.2841487579820567e+215 | 1.0 |nan

100 |-8.578692995582524e+215 | 4 |-3.871060255574359e+211 | 10.0 |nan

200 |-9.797986266970411e+216 | 8 |-3.9959663118425665e+204 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.19343022641847e+19 | 1 |1.091353133656954e+126 | 0.1 |1.4759699245250634e+110

50 |-6.314729024204743e+72 | 2 |8.328421037493386e+110 | 1.0 |1.0384640247175158e+208

100 |-6.465453790994439e+105 | 4 |4.415983767175015e+102 | 10.0 |nan

200 |5.801370721063024e+136 | 8 |1.650992666325736e+100 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.75065967994115e+209 | 1 |1.4143726097636114e+217 | 0.1 |-6.031434826018988e+214

50 |7.076956321386588e+208 | 2 |2.5672402671583986e+214 | 1.0 |nan

100 |4.7388674479681797e+213 | 4 |1.3236777701111363e+210 | 10.0 |nan

200 |8.498595049160442e+219 | 8 |-6.809253824793349e+205 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.2740879972881434e+16 | 1 |3.79498699004196e+113 | 0.1 |-8.363454364710364e+109

50 |2.0204048048518844e+87 | 2 |4.679368435767431e+110 | 1.0 |1.2901204343688433e+213

100 |4.540046535927828e+106 | 4 |1.4165284250197474e+104 | 10.0 |nan

200 |4.0701888641683445e+142 | 8 |1.490101497254434e+106 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-2.025450907024407e+209 | 1 |3.3492126034648985e+217 | 0.1 |-4.264668947550336e+214

50 |-1.3890290939622216e+212 | 2 |3.002087973296729e+214 | 1.0 |nan

100 |1.0476462503617112e+213 | 4 |-7.969868502136216e+208 | 10.0 |nan

200 |6.531416335514712e+216 | 8 |7.800924867523963e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |15681282330.15809 | 1 |1.7624855740669847e+118 | 0.1 |3.7486953119202526e+108

50 |1.6622772591310773e+76 | 2 |-3.66536188480009e+109 | 1.0 |-1.1921501176492642e+212

100 |6.93881367307489e+108 | 4 |-2.072941277805008e+112 | 10.0 |nan

200 |6.948003835243768e+140 | 8 |-1.2732017321351674e+110 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-3.4823656321710892e+208 | 1 |3.7398315007248076e+217 | 0.1 |-3.2694984491828736e+212

50 |9.027728422449662e+211 | 2 |5.394399844292877e+216 | 1.0 |nan

100 |-1.8150250716004065e+214 | 4 |1.1171227826967069e+209 | 10.0 |nan

200 |7.767053740129707e+217 | 8 |-2.401094340585271e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.490624455186795 | 1 |4.987040572550879e+115 | 0.1 |5.395955869269326e+106

50 |-6.467084922687309e+81 | 2 |6.2132524499291654e+106 | 1.0 |3.176969747860321e+210

100 |1.530526255298016e+105 | 4 |4.56735874164266e+103 | 10.0 |nan

200 |2.5026746812793062e+141 | 8 |8.084409151468766e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.9808968934252653e+210 | 1 |1.0995403305563135e+217 | 0.1 |3.389809543073405e+214

50 |5.702231083091101e+212 | 2 |3.665626474745028e+212 | 1.0 |nan

100 |1.4711411332143684e+214 | 4 |4.357697946564616e+211 | 10.0 |nan

200 |6.632791949511248e+217 | 8 |-1.5724581375722287e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |9759023364.763916 | 1 |2.317504152689068e+112 | 0.1 |-1.769132828706467e+108

50 |9.787168878354811e+84 | 2 |1.0221556562911468e+110 | 1.0 |-9.320645032306726e+211

100 |3.7486979117735694e+110 | 4 |-4.8695235339132525e+107 | 10.0 |4.0957246280640686e+304

200 |-1.3225640443793607e+138 | 8 |-7.44068664757278e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.3231535307987037e+209 | 1 |-5.416330606825574e+219 | 0.1 |1.9411684051300887e+213

50 |-6.404257423044086e+211 | 2 |1.9816779197433267e+214 | 1.0 |nan

100 |7.347840923732669e+212 | 4 |-2.0588844264000111e+211 | 10.0 |nan

200 |1.2104880855907883e+217 | 8 |-1.4442556152528681e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.6817674285383334e+17 | 1 |1.0748778822433184e+114 | 0.1 |1.8835364936264577e+105

50 |6.233477703416443e+80 | 2 |1.0961348584752557e+112 | 1.0 |2.3659638319645814e+212

100 |4.290526225116407e+111 | 4 |-1.1255277271465175e+104 | 10.0 |-1.7087331319389085e+307

200 |4.708990515103246e+141 | 8 |-1.2641426735079495e+105 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |6.5002900875246825e+209 | 1 |1.7033769988659139e+214 | 0.1 |-1.9857178906039834e+214

50 |3.0604808293397125e+209 | 2 |2.2252000003341774e+215 | 1.0 |nan

100 |-2.825265932187031e+213 | 4 |-4.4996131703389285e+211 | 10.0 |nan

200 |3.3584017863761893e+217 | 8 |1.1982952107670198e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.504752613792009e+22 | 1 |3.189173166179953e+111 | 0.1 |5.8543222388129914e+107

50 |4.1187350553884775e+79 | 2 |7.162860857142426e+112 | 1.0 |3.2361544304628853e+201

100 |7.953089174394209e+107 | 4 |-6.0406361433933225e+106 | 10.0 |nan

200 |-1.83559299046973e+137 | 8 |6.823364020722354e+106 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |4.550547501843048e+209 | 1 |2.0594190131037292e+219 | 0.1 |-1.3065267312809363e+214

50 |6.411605588424347e+214 | 2 |3.461108526964528e+215 | 1.0 |nan

100 |9.335654277124306e+212 | 4 |2.0955556475721482e+211 | 10.0 |nan

200 |1.118193030873965e+218 | 8 |-1.4079218185275713e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |9396396694023.254 | 1 |1.0410319197809215e+110 | 0.1 |1.2799658034559096e+106

50 |2.8757545887374716e+75 | 2 |1.990203490983569e+109 | 1.0 |-1.5422401501428837e+207

100 |9.259556232609803e+112 | 4 |7.633917273526524e+106 | 10.0 |nan

200 |3.963566917559621e+137 | 8 |-8.17945497588391e+107 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.5318217264049585e+209 | 1 |-1.1821137418369019e+219 | 0.1 |5.433021503308024e+213

50 |1.8705787406100959e+211 | 2 |9.140728811870805e+213 | 1.0 |nan

100 |-3.151708555559506e+213 | 4 |5.443331440432598e+209 | 10.0 |nan

200 |2.2880255505959934e+216 | 8 |-3.978040889168139e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.7204123959395037e+19 | 1 |2.9518681384661614e+112 | 0.1 |2.5032682435026836e+106

50 |5.3551149102712404e+79 | 2 |-6.813436314625104e+112 | 1.0 |4.994861919470534e+205

100 |3.8220299628129665e+111 | 4 |8.243861542169223e+107 | 10.0 |nan

200 |-8.982810512783562e+140 | 8 |3.6752342197211925e+105 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |4.4528520408058605e+209 | 1 |-5.499956456690793e+218 | 0.1 |-1.4480618116029009e+214

50 |2.881631545527945e+211 | 2 |3.257716174599243e+212 | 1.0 |nan

100 |2.7216792269310283e+213 | 4 |1.6555223925521058e+209 | 10.0 |nan

200 |2.0949287254626503e+218 | 8 |-4.975906647237704e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-1065095581593.3256 | 1 |1.0942117007165013e+106 | 0.1 |-1.343414748727026e+112

50 |6.621698678030423e+71 | 2 |4.848872911006043e+108 | 1.0 |2.265086756536901e+204

100 |-5.834912024598974e+108 | 4 |3.785742276911938e+103 | 10.0 |nan

200 |6.301626470816987e+137 | 8 |-7.739752779723256e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.0950409844151055e+212 | 1 |6.057832438012426e+217 | 0.1 |3.558891321926778e+213

50 |2.1685656869585886e+212 | 2 |8.199312140650124e+214 | 1.0 |nan

100 |-2.8178383614335007e+215 | 4 |1.9997824644210318e+209 | 10.0 |nan

200 |1.1790086439349593e+218 | 8 |2.33792350095656e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |7827889050.09373 | 1 |6.9259965067089045e+109 | 0.1 |9.718787966869479e+108

50 |-2.2754569734514574e+86 | 2 |-2.4682897026989876e+112 | 1.0 |2.2408230963225785e+214

100 |-1.1040860463218733e+111 | 4 |-1.11702091612878e+111 | 10.0 |nan

200 |-3.5145879243129904e+144 | 8 |-1.5889385025420004e+102 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-5.207457907311324e+210 | 1 |6.899996611041438e+216 | 0.1 |-4.733627007544547e+213

50 |-1.746623827649505e+212 | 2 |-8.610139016634478e+213 | 1.0 |nan

100 |6.311271976500989e+212 | 4 |-1.0914268618448818e+211 | 10.0 |nan

200 |-5.3818048831166985e+218 | 8 |2.207237222868191e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |4.256810920242916e+19 | 1 |-6.121234066330069e+112 | 0.1 |-6.563201853837586e+120

50 |2.573338571740396e+85 | 2 |-1.1453241054597907e+104 | 1.0 |8.18539078831745e+201

100 |6.4980326138534e+106 | 4 |8.631348826986876e+106 | 10.0 |nan

200 |5.654640788453179e+139 | 8 |1.192336726198981e+102 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.882391509483832e+210 | 1 |1.8177382556347655e+218 | 0.1 |4.430451665644911e+213

50 |2.5071414295037645e+212 | 2 |4.7165306436001734e+213 | 1.0 |nan

100 |4.0342960465651544e+212 | 4 |-2.0289613099797282e+210 | 10.0 |nan

200 |1.0620108376846074e+217 | 8 |6.754973996882745e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2121581650.2714329 | 1 |-1.1541864054863006e+116 | 0.1 |1.2724184249531955e+112

50 |1.0085787509909746e+82 | 2 |2.2370164089652033e+106 | 1.0 |-2.785820373709816e+216

100 |-5.042811125067139e+111 | 4 |-2.654893909384744e+103 | 10.0 |nan

200 |7.081455747300285e+137 | 8 |-5.423912874285194e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.587907020090219e+210 | 1 |-8.704909167498747e+217 | 0.1 |-3.195197237669536e+213

50 |2.5159908555719062e+212 | 2 |4.242978541448562e+213 | 1.0 |nan

100 |2.0004330339812048e+213 | 4 |1.0664432174120649e+209 | 10.0 |nan

200 |-6.237278198840779e+215 | 8 |2.3421290772832503e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |10.028277503777344 | 1 |3.140572485535413e+110 | 0.1 |1.4257138499383032e+114

50 |7.334945734070441e+85 | 2 |5.396621779645526e+107 | 1.0 |2.65685119349575e+210

100 |1.7678612809638606e+106 | 4 |3.70611624262352e+104 | 10.0 |nan

200 |8.460463733926745e+138 | 8 |-1.7122320801180627e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |7.095993965187165e+207 | 1 |-8.50826964496959e+218 | 0.1 |5.168366521278811e+211

50 |2.3346730964699206e+210 | 2 |-3.0621820636539497e+211 | 1.0 |nan

100 |-3.203663807036103e+214 | 4 |1.9104795819438442e+209 | 10.0 |nan

200 |2.8682084243611317e+218 | 8 |3.2504951719469544e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.1926902637387076e+16 | 1 |-3.086493901366968e+120 | 0.1 |-1.6050287498576632e+112

50 |-6.944840574159301e+89 | 2 |2.4493234835768317e+112 | 1.0 |7.484037371634772e+211

100 |3.3383778522421765e+106 | 4 |-7.649538526394943e+107 | 10.0 |nan

200 |4.6308991290088555e+141 | 8 |1.2664413216716686e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-8.82552896970993e+211 | 1 |7.628838514230881e+217 | 0.1 |-4.488896150938129e+214

50 |-4.318518870643816e+210 | 2 |3.3963829122562495e+212 | 1.0 |nan

100 |4.534857332655187e+213 | 4 |1.551913202889483e+211 | 10.0 |nan

200 |4.1823764795461295e+217 | 8 |-1.1887203391597866e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |7.07152138444167e+26 | 1 |3.007596190848825e+104 | 0.1 |-8.554808500637436e+108

50 |7.905556673520751e+84 | 2 |4.241562731870862e+112 | 1.0 |1.0226449267477716e+210

100 |1.3175544418910511e+104 | 4 |3.718507125256038e+100 | 10.0 |nan

200 |7.613743281393662e+142 | 8 |6.787253242185234e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-1.1821035512273377e+209 | 1 |-4.164551820523059e+221 | 0.1 |5.350302877645293e+212

50 |-6.298701733785658e+210 | 2 |2.841233729544876e+214 | 1.0 |nan

100 |5.815249507000918e+214 | 4 |2.09412143813611e+210 | 10.0 |nan

200 |5.3428865979215474e+219 | 8 |-1.7681568648013733e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-10056682924891.986 | 1 |1.0128619105314044e+107 | 0.1 |-1.369868885772458e+108

50 |1.4793536020191758e+81 | 2 |3.936168634279598e+106 | 1.0 |-1.5052431038445123e+214

100 |-1.0177662126453206e+117 | 4 |3.244590363362442e+107 | 10.0 |nan

200 |-2.1589636791315838e+139 | 8 |2.1013007505753288e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.466736139724313e+211 | 1 |-1.3759854912815847e+220 | 0.1 |1.9716066235850896e+215

50 |6.735181797322247e+211 | 2 |-1.2361764766267937e+215 | 1.0 |nan

100 |2.0019869890043705e+214 | 4 |-1.6687035506492015e+210 | 10.0 |nan

200 |-4.127713978649064e+217 | 8 |-1.1662426833059805e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.2093537446184723e+21 | 1 |-1.0982764122812758e+115 | 0.1 |7.081953231807516e+108

50 |1.9897915720227497e+81 | 2 |9.9746121364285e+112 | 1.0 |5.633664757465649e+210

100 |9.022319707304648e+113 | 4 |2.5168138816903214e+102 | 10.0 |-inf

200 |-1.7807313798898218e+142 | 8 |6.121645744173871e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.2842563441814205e+209 | 1 |-1.9033368780737115e+217 | 0.1 |7.142146680799863e+214

50 |-6.889395263873023e+210 | 2 |-6.859930816088507e+212 | 1.0 |nan

100 |8.222816061840974e+211 | 4 |1.5225841383381095e+209 | 10.0 |nan

200 |2.6830356980026526e+217 | 8 |8.98792289761991e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-4.277970431671806e+17 | 1 |-1.2649069818368726e+115 | 0.1 |1.7458208526195822e+105

50 |3.824551997660066e+82 | 2 |-6.0936824678305055e+112 | 1.0 |-1.097437925337643e+211

100 |1.6332362667454773e+112 | 4 |5.206182624759085e+108 | 10.0 |3.750333094399853e+300

200 |8.074337915559008e+139 | 8 |3.7786285502491755e+102 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-2.559452158163762e+210 | 1 |1.2849886550985227e+216 | 0.1 |3.002908794245318e+213

50 |7.266163002667783e+211 | 2 |4.564727421775664e+214 | 1.0 |nan

100 |-5.235155854854218e+214 | 4 |-4.2157510367367314e+210 | 10.0 |nan

200 |2.4488378278844283e+215 | 8 |-1.5867537492015035e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |219121051434258.34 | 1 |7.108447436855589e+115 | 0.1 |-1.6697496643079038e+113

50 |-3.5711381112550656e+85 | 2 |1.5961401261487713e+110 | 1.0 |5.049091324859215e+210

100 |2.1089025229947636e+103 | 4 |6.998098869468207e+112 | 10.0 |nan

200 |-2.099008091200316e+140 | 8 |-7.000032415671044e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.2322110648430344e+210 | 1 |7.054534702956105e+215 | 0.1 |-6.106780547554879e+215

50 |-2.1039871805960814e+212 | 2 |2.944624501041266e+214 | 1.0 |nan

100 |-8.89785965905056e+212 | 4 |3.6047130924025643e+210 | 10.0 |nan

200 |1.485079847523801e+218 | 8 |1.2817364525792155e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |4.60512348172449e+20 | 1 |-9.349924227581036e+117 | 0.1 |4.846683782018512e+108

50 |1.081546965366502e+75 | 2 |2.3793308566038944e+106 | 1.0 |1.7977542009521972e+202

100 |-4.630374314002817e+109 | 4 |-5.471578833142131e+108 | 10.0 |5.834487948375929e+306

200 |5.2072222646659926e+141 | 8 |-2.371232929766845e+100 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |6.336851397168994e+210 | 1 |5.718102940516254e+217 | 0.1 |3.671594683926799e+211

50 |1.3767394037380424e+211 | 2 |-4.374097402113475e+214 | 1.0 |nan

100 |-3.179965476547389e+216 | 4 |-3.107205674874421e+210 | 10.0 |nan

200 |-3.2185254181566307e+218 | 8 |-2.76234662265948e+209 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.428053965417113e+25 | 1 |-1.209890400683223e+111 | 0.1 |8.32174825330209e+115

50 |1.0509999476197717e+82 | 2 |-3.123253311776324e+107 | 1.0 |-8.188346727796368e+215

100 |-6.812109492019455e+113 | 4 |7.139497483043145e+103 | 10.0 |2.7418992454343806e+302

200 |4.5336209204848474e+141 | 8 |-2.1492558358941062e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-1.3175456043308853e+213 | 1 |-2.6065719195154722e+219 | 0.1 |-1.728264645891017e+211

50 |2.0107279432807041e+211 | 2 |2.2272290566377113e+212 | 1.0 |nan

100 |2.0773419163086564e+213 | 4 |-1.753027603508954e+209 | 10.0 |nan

200 |1.2119832466220943e+218 | 8 |-1.7008051539778589e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |12389649617.83265 | 1 |1.4025227695380196e+113 | 0.1 |1.7696123196754414e+115

50 |4.1498573993593456e+86 | 2 |7.475281944920952e+108 | 1.0 |-8.573231619471374e+215

100 |-1.7554803185443624e+108 | 4 |8.350532718354252e+100 | 10.0 |nan

200 |6.293886936212287e+138 | 8 |4.0087627114597867e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |6.231248548456369e+210 | 1 |-2.632127792514549e+218 | 0.1 |9.99909701338298e+215

50 |1.7619171066965422e+212 | 2 |7.847051077806931e+213 | 1.0 |nan

100 |1.487777160112981e+214 | 4 |-8.072247933807115e+209 | 10.0 |nan

200 |-1.4672369592599676e+219 | 8 |2.4547313985152095e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.05947172376922e+19 | 1 |1.032885857729942e+115 | 0.1 |-4.1986119883369146e+102

50 |-1.5966385271887558e+82 | 2 |1.7643048733448788e+112 | 1.0 |2.4185988827858228e+213

100 |-3.750498422625589e+111 | 4 |2.3457395126918134e+104 | 10.0 |nan

200 |9.904406196102862e+138 | 8 |-4.71929971979729e+105 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.9242712169667037e+210 | 1 |-4.2941664502506527e+220 | 0.1 |8.407362127311171e+212

50 |-5.957401041240883e+210 | 2 |-3.725778597468803e+216 | 1.0 |nan

100 |1.952421514549935e+214 | 4 |-3.329494463683729e+209 | 10.0 |nan

200 |9.490714159478682e+217 | 8 |-1.8380389885595267e+205 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.7728822330702636e+18 | 1 |9.715840551501112e+115 | 0.1 |1.5980415558279862e+115

50 |3.2123755810714718e+84 | 2 |-1.1153500597057759e+108 | 1.0 |8.327458651916776e+210

100 |3.9252622998387306e+110 | 4 |5.584276485739044e+102 | 10.0 |-inf

200 |-2.0616352008182664e+138 | 8 |-3.019197410169204e+102 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.6656464668521965e+211 | 1 |-8.968025775766336e+217 | 0.1 |-7.689938992029065e+213

50 |8.548575290507933e+209 | 2 |6.054744010102108e+211 | 1.0 |nan

100 |4.9268666974333405e+212 | 4 |3.948658293019738e+209 | 10.0 |nan

200 |8.457675980098504e+215 | 8 |4.370907910530119e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |124.6068749358352 | 1 |2.1461557087641585e+118 | 0.1 |8.26704254353695e+109

50 |1.7486949682926112e+85 | 2 |-9.416347688213537e+109 | 1.0 |1.0981414995558826e+204

100 |4.2160465333667e+112 | 4 |3.5169896692774426e+109 | 10.0 |nan

200 |-8.295654190294832e+139 | 8 |-2.0742095794521872e+99 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |3.8595947700922696e+209 | 1 |4.0984677816874886e+221 | 0.1 |-9.774745532242525e+215

50 |-1.6589351313813022e+211 | 2 |-2.795532992727941e+213 | 1.0 |nan

100 |1.1130285232184933e+213 | 4 |-1.0873364061917808e+209 | 10.0 |nan

200 |-3.211977656807966e+217 | 8 |1.6785504705763092e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |586.0014258582762 | 1 |-2.275598633390972e+113 | 0.1 |2.7902525998978364e+107

50 |3.9963144567517174e+92 | 2 |-7.188124381747037e+111 | 1.0 |-7.522206857924125e+209

100 |7.494042174905312e+108 | 4 |-3.863669682784224e+109 | 10.0 |nan

200 |1.4593235803886291e+138 | 8 |-4.0004429225096704e+102 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.9405540440733067e+207 | 1 |1.2442571789988925e+218 | 0.1 |2.4264310229684528e+213

50 |1.1799248011847854e+212 | 2 |3.28197048698552e+214 | 1.0 |nan

100 |-4.382004810266824e+214 | 4 |5.679031096861521e+211 | 10.0 |nan

200 |8.025685810831244e+216 | 8 |4.545122221804089e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1586312300.8341682 | 1 |2.48524154948518e+114 | 0.1 |-8.613287731540227e+118

50 |-3.528462156663725e+78 | 2 |1.2700080311616126e+106 | 1.0 |-2.372626959404846e+211

100 |1.918931739387602e+106 | 4 |4.837107877148766e+102 | 10.0 |nan

200 |7.171272302438429e+136 | 8 |4.8774820038862314e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.0037655470356235e+211 | 1 |9.265012245041102e+216 | 0.1 |2.568584262099785e+213

50 |5.780276353383284e+211 | 2 |-6.014424556016512e+213 | 1.0 |nan

100 |2.675782384307962e+213 | 4 |4.623375895693051e+211 | 10.0 |nan

200 |2.3801239736057623e+217 | 8 |1.8243548133165008e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-4793979486.172105 | 1 |8.316682554577971e+105 | 0.1 |-1.7340009822518483e+104

50 |2.7689020592554265e+87 | 2 |-2.0511130582940814e+113 | 1.0 |2.0976678569285316e+207

100 |1.6676593256187987e+115 | 4 |4.633462181563116e+105 | 10.0 |nan

200 |5.276484221470446e+139 | 8 |2.4297282165024557e+105 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.0141356777924804e+208 | 1 |4.083025005703933e+218 | 0.1 |-6.697944554586975e+214

50 |-4.183565232281659e+210 | 2 |1.171122082931671e+214 | 1.0 |nan

100 |8.026833611479678e+212 | 4 |2.928483839824136e+209 | 10.0 |nan

200 |8.81775679880251e+217 | 8 |-5.449544473416421e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.1274260378534926e+33 | 1 |4.268182736772302e+115 | 0.1 |1.1234761122270622e+110

50 |2.932262852514874e+73 | 2 |9.813747114328925e+112 | 1.0 |1.4965564849988106e+214

100 |-7.844988838878303e+103 | 4 |-6.247152286555251e+110 | 10.0 |nan

200 |6.35152029550069e+140 | 8 |-1.3522408970870843e+102 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-1.230856464234532e+210 | 1 |6.381550906825645e+216 | 0.1 |2.2615564394475886e+213

50 |1.2812129304732847e+210 | 2 |6.221479821638222e+211 | 1.0 |nan

100 |1.4262026509901212e+215 | 4 |-6.2414197216768884e+209 | 10.0 |nan

200 |5.669105892599309e+217 | 8 |7.632064314622643e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |9.463172948959124e+23 | 1 |4.545580875819461e+106 | 0.1 |9.257072174576317e+99

50 |2.763597412547109e+84 | 2 |4.824020828059737e+113 | 1.0 |6.976346928108106e+201

100 |2.458482467043733e+113 | 4 |4.3905100992513477e+108 | 10.0 |-2.9226632773053035e+300

200 |4.789984752701201e+140 | 8 |1.9293768070774775e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.4856409122087888e+210 | 1 |3.038863129977156e+216 | 0.1 |-8.054205714456736e+214

50 |-1.3301475080717356e+211 | 2 |-1.9630716055201428e+214 | 1.0 |nan

100 |2.477363850573087e+214 | 4 |-1.1126160402255168e+212 | 10.0 |nan

200 |1.770181550703144e+220 | 8 |5.021427897507678e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.5024802782000747e+23 | 1 |2.375165063901204e+115 | 0.1 |2.5496372619890895e+102

50 |1.6467706341591498e+76 | 2 |-6.423413840593868e+107 | 1.0 |5.436083507690072e+200

100 |2.0811170159144658e+109 | 4 |1.5566965757356198e+107 | 10.0 |nan

200 |-1.2761135426690757e+141 | 8 |2.3205364822257557e+102 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-4.4870359346120733e+210 | 1 |2.1759966972141229e+217 | 0.1 |5.0437618866491745e+213

50 |3.7395091212551503e+211 | 2 |-1.2203785366082375e+213 | 1.0 |nan

100 |-7.530891666344876e+214 | 4 |1.0712069419134105e+212 | 10.0 |nan

200 |1.1512220007191568e+217 | 8 |1.0616987092027706e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |180642133.60725474 | 1 |-4.2776235568409354e+120 | 0.1 |7.309369393055415e+109

50 |1.8663553612269406e+79 | 2 |1.9875572522386216e+113 | 1.0 |-7.2595159075958415e+211

100 |6.534645882876218e+106 | 4 |-3.1287467673240004e+103 | 10.0 |nan

200 |2.8206267817815265e+137 | 8 |-2.1975013977608078e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-2.810126781564308e+210 | 1 |-1.8338606480109386e+218 | 0.1 |-5.797468387307523e+213

50 |2.0124360891240184e+213 | 2 |5.001045589126764e+215 | 1.0 |nan

100 |-7.736169817597755e+212 | 4 |1.7920539390079883e+209 | 10.0 |nan

200 |9.425193629791585e+216 | 8 |4.0571723920307434e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-4.297453500405771e+29 | 1 |1.2779856434424265e+114 | 0.1 |3.363489014477811e+114

50 |5.061914233503368e+73 | 2 |-3.6100856962245572e+112 | 1.0 |1.77885898646992e+214

100 |1.7194671106111564e+114 | 4 |2.7966726163805136e+108 | 10.0 |nan

200 |-1.5277714699669322e+140 | 8 |-2.41944922114597e+108 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.2417820504420954e+210 | 1 |-2.3634029207321836e+220 | 0.1 |9.468875952291304e+213

50 |9.223161196613771e+211 | 2 |2.568831871387737e+215 | 1.0 |nan

100 |5.435230415159559e+213 | 4 |4.860984283347268e+211 | 10.0 |nan

200 |5.452838649413212e+217 | 8 |6.589560352848724e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |5.773820805604853e+23 | 1 |1.8142959773840674e+118 | 0.1 |2.0981696536407788e+102

50 |1.5168972049887213e+84 | 2 |1.715324889101169e+113 | 1.0 |-2.485563815923274e+211

100 |1.5771005645757826e+104 | 4 |7.540422511446602e+109 | 10.0 |5.799871841313753e+307

200 |3.0987144732049094e+145 | 8 |2.891342152248225e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |4.475495849583765e+209 | 1 |5.022491001853893e+216 | 0.1 |1.354848961647384e+213

50 |-3.181577098240367e+211 | 2 |2.168836909650311e+213 | 1.0 |nan

100 |4.497628258161258e+214 | 4 |2.806125003213243e+210 | 10.0 |nan

200 |3.920384173466373e+216 | 8 |1.1160408457585735e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.960212568244029e+38 | 1 |4.052583161313552e+112 | 0.1 |1.7375241324327423e+115

50 |5.297694199102047e+86 | 2 |9.416960958602079e+109 | 1.0 |1.4771073335162112e+214

100 |-1.7917353635961024e+114 | 4 |-1.0250994441682056e+101 | 10.0 |1.0563685800165717e+305

200 |2.001057447711207e+137 | 8 |5.262550918387536e+107 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-3.9942699359746136e+209 | 1 |1.4463142205200143e+217 | 0.1 |-9.291799311422099e+214

50 |-1.9873213353826429e+214 | 2 |6.400581237370447e+213 | 1.0 |nan

100 |-3.1933169883082972e+212 | 4 |4.564996025064154e+212 | 10.0 |nan

200 |-1.0703040577447539e+218 | 8 |1.1664999886531881e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-1512511410533.029 | 1 |4.79537621056658e+112 | 0.1 |6.899254420299416e+108

50 |2.0000157941196543e+79 | 2 |6.037456694341677e+115 | 1.0 |1.7205155776975242e+213

100 |7.109898504393373e+105 | 4 |-1.2673770150823505e+109 | 10.0 |nan

200 |2.395761124962038e+140 | 8 |8.06841149006814e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-7.039959206880416e+209 | 1 |-4.662824665723458e+218 | 0.1 |1.3687533394217064e+214

50 |-1.5239093434912065e+211 | 2 |1.9746016908132543e+215 | 1.0 |nan

100 |6.57289027778544e+214 | 4 |9.546795525015412e+209 | 10.0 |nan

200 |3.52068871240161e+217 | 8 |2.7104980107370363e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-4.757333660277977e+19 | 1 |-2.493302719685074e+119 | 0.1 |5.297566298206638e+118

50 |3.441938104548407e+73 | 2 |6.971677304369064e+108 | 1.0 |6.529829388378185e+209

100 |2.3618327705620883e+108 | 4 |1.5389446920362367e+111 | 10.0 |-inf

200 |-1.5655915078449812e+140 | 8 |-4.575118736050669e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.7937239811708937e+211 | 1 |-2.5547344981198294e+218 | 0.1 |-6.493971182166055e+214

50 |4.333990269255578e+212 | 2 |1.1323905096225124e+213 | 1.0 |nan

100 |2.890184101498187e+215 | 4 |1.6870586344780237e+209 | 10.0 |nan

200 |-1.4239707844932668e+217 | 8 |-3.613538886357251e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2312686825602.6743 | 1 |2.3615046593456623e+116 | 0.1 |-2.6167826761824895e+103

50 |-9.665694780116781e+80 | 2 |3.147524702629351e+105 | 1.0 |1.265330827004492e+211

100 |-9.451676505277336e+112 | 4 |-3.774923747467878e+104 | 10.0 |nan

200 |-1.0223750724276194e+139 | 8 |3.1617015821317307e+105 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |7.228808292030386e+210 | 1 |3.654628648892706e+218 | 0.1 |1.4453859509763066e+211

50 |6.164551523922901e+211 | 2 |-6.060894727779879e+215 | 1.0 |nan

100 |7.909958870814523e+214 | 4 |3.244331707193799e+211 | 10.0 |nan

200 |9.560147529481475e+217 | 8 |-2.6269562777609943e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |16.534368868192804 | 1 |-1.4028369261034433e+116 | 0.1 |3.0832049148270723e+107

50 |9.544258390804004e+81 | 2 |1.1027677029729915e+107 | 1.0 |1.5827510704368958e+207

100 |3.5079513433247546e+111 | 4 |2.821300147159528e+103 | 10.0 |nan

200 |6.357842045279456e+136 | 8 |1.0323893014230589e+107 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.23864905923315e+208 | 1 |1.6492880999057154e+217 | 0.1 |1.7221338817334472e+213

50 |-9.631078185999422e+211 | 2 |5.2711273225207244e+212 | 1.0 |nan

100 |5.693006497562238e+215 | 4 |4.323632763919489e+209 | 10.0 |nan

200 |2.7735465228963215e+216 | 8 |4.1193525894004835e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |31629653.54722107 | 1 |4.9424364600594764e+119 | 0.1 |-9.384067921392853e+115

50 |3.866392810036961e+73 | 2 |-4.185522755792585e+117 | 1.0 |6.732083600110247e+205

100 |-5.881020461152334e+109 | 4 |1.1884651114275092e+110 | 10.0 |2.468580853090591e+307

200 |-3.7994055507195297e+136 | 8 |-2.0453554329020384e+102 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.117576602466677e+209 | 1 |3.3534994258753144e+214 | 0.1 |-2.2367060599806936e+215

50 |1.6778877199291408e+212 | 2 |1.143044456416592e+214 | 1.0 |nan

100 |2.5138832825745677e+215 | 4 |1.2431418242379432e+210 | 10.0 |nan

200 |5.664905805458244e+217 | 8 |1.5603370402776056e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |3.2852567218469663e+19 | 1 |-6.812532087184896e+117 | 0.1 |1.791128100069256e+112

50 |-1.3599801224507158e+78 | 2 |4.758655038938794e+109 | 1.0 |3.3548720979853076e+200

100 |-3.159104976751027e+116 | 4 |1.0258996756054082e+108 | 10.0 |5.813304379322257e+304

200 |3.205064113158169e+141 | 8 |-2.5836547481311277e+105 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |6.861775420881827e+208 | 1 |2.8415752034524604e+218 | 0.1 |4.8868458162785736e+215

50 |2.4313958147001828e+213 | 2 |-7.949033916957899e+212 | 1.0 |nan

100 |-2.748847375367414e+214 | 4 |-1.6477711913566989e+212 | 10.0 |nan

200 |-9.62897424341301e+217 | 8 |-2.023940752827107e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |222210838589.31702 | 1 |1.048498682202552e+115 | 0.1 |1.3688696322466742e+109

50 |4.331434069256869e+68 | 2 |-2.0447571514836e+107 | 1.0 |2.4957238892139768e+206

100 |5.274163067686632e+100 | 4 |-6.538832571261431e+102 | 10.0 |nan

200 |1.031249162317036e+134 | 8 |-9.862676818049747e+102 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.0991785244859063e+211 | 1 |2.018629486284991e+218 | 0.1 |1.1666546792222745e+213

50 |-1.1556324470655328e+211 | 2 |3.6610635816492964e+213 | 1.0 |nan

100 |1.869850829829545e+212 | 4 |-2.7752181031330745e+210 | 10.0 |nan

200 |9.41827661060726e+218 | 8 |4.9963895597982694e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |34486.99553457892 | 1 |7.971186299368589e+108 | 0.1 |1.962894937549345e+113

50 |5.3682770757674206e+79 | 2 |1.3274255745294257e+111 | 1.0 |1.4262489057588542e+208

100 |8.087157865994455e+112 | 4 |3.288654377702259e+111 | 10.0 |nan

200 |1.0189532117848103e+142 | 8 |6.628621549504055e+105 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |7.24046685105125e+207 | 1 |1.3261402344818238e+216 | 0.1 |5.795881679561243e+215

50 |7.180892594247852e+209 | 2 |6.798717609946388e+212 | 1.0 |nan

100 |5.208894669842217e+214 | 4 |1.536252413110984e+211 | 10.0 |nan

200 |2.8640681784164634e+218 | 8 |1.6515416495739516e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |0.2819572084565329 | 1 |1.5209556191952232e+109 | 0.1 |-3.539851436559606e+108

50 |1.2080233555974786e+76 | 2 |-2.0370980783508147e+107 | 1.0 |5.0275062602879866e+206

100 |3.0759995681514826e+104 | 4 |9.39016043319223e+103 | 10.0 |-inf

200 |-6.37160051823048e+141 | 8 |2.8280928663633546e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-1.789194054737817e+210 | 1 |5.653988483925868e+219 | 0.1 |-8.424980555228248e+214

50 |1.0420787746382297e+210 | 2 |2.9258858299320865e+213 | 1.0 |nan

100 |3.254590873259859e+212 | 4 |3.758755946023058e+209 | 10.0 |nan

200 |-9.410806117007835e+216 | 8 |4.922142003006977e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-111030989195.0353 | 1 |1.6605394849938873e+108 | 0.1 |2.647043679342675e+112

50 |2.9630071456308273e+69 | 2 |7.771493058438369e+105 | 1.0 |8.629389966108618e+213

100 |3.0751777888999723e+113 | 4 |-2.6413837866338894e+103 | 10.0 |nan

200 |4.147782675843174e+143 | 8 |1.8626416759119118e+100 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.2082690441551187e+209 | 1 |2.074987615257864e+218 | 0.1 |9.320842998156603e+211

50 |-3.095807529458024e+210 | 2 |1.0061978904055831e+213 | 1.0 |nan

100 |5.1896201894310066e+212 | 4 |4.6076555325565584e+210 | 10.0 |nan

200 |6.4100381449629616e+218 | 8 |-9.8459806547834e+209 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |0.39946545374224174 | 1 |-1.3750336711125222e+108 | 0.1 |1.195464822638583e+106

50 |1.0298996934812005e+86 | 2 |1.0527322028949962e+110 | 1.0 |2.8432728670152656e+204

100 |5.8123611858801455e+113 | 4 |9.971545739206162e+100 | 10.0 |2.243725905236675e+301

200 |3.39866402036759e+143 | 8 |-2.050144287692182e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.5494631968652987e+210 | 1 |1.3515686151304811e+219 | 0.1 |-3.023722795230347e+213

50 |2.05627558454757e+212 | 2 |2.1670934641196004e+212 | 1.0 |nan

100 |-7.427742822466881e+213 | 4 |1.0110191656431314e+210 | 10.0 |nan

200 |1.0454333739952234e+217 | 8 |1.8018362537823652e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.0173766401275547e+17 | 1 |7.377284841675986e+119 | 0.1 |6.12359667238757e+108

50 |-8.953539959263947e+85 | 2 |1.0609083110224302e+104 | 1.0 |-1.0978898000646534e+216

100 |1.6412858732066194e+108 | 4 |-4.431563737879841e+105 | 10.0 |nan

200 |1.0558853839212934e+139 | 8 |6.669058297244642e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-2.07176904540019e+210 | 1 |4.847590308635903e+217 | 0.1 |3.501520344524358e+214

50 |2.932680612646422e+213 | 2 |1.5005307381745495e+216 | 1.0 |nan

100 |2.4051722639317367e+213 | 4 |7.869817860774859e+210 | 10.0 |nan

200 |-1.2822745966436432e+217 | 8 |1.859194279106361e+206 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |7607657681.959407 | 1 |2.3523304148498917e+117 | 0.1 |1.3424227092022626e+113

50 |1.2510294253402544e+87 | 2 |1.8626011116782328e+112 | 1.0 |-3.4252774504026406e+214

100 |4.333494006966124e+110 | 4 |5.508926259686591e+104 | 10.0 |1.0718833349534291e+306

200 |-1.7456151715758696e+139 | 8 |4.852648268757867e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.4847756700443987e+208 | 1 |3.0491365353730256e+219 | 0.1 |1.5878848054884207e+215

50 |-2.351295907593162e+213 | 2 |5.396936227780599e+215 | 1.0 |nan

100 |1.944168495438501e+214 | 4 |2.8398685635458535e+210 | 10.0 |nan

200 |2.536377354056239e+219 | 8 |-1.2459479069179761e+209 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.34279531467458e+23 | 1 |2.676343878117755e+114 | 0.1 |2.2507451555428074e+109

50 |2.6925428089427353e+77 | 2 |4.635851101467152e+112 | 1.0 |-3.1150323024469405e+211

100 |1.5780946079654223e+110 | 4 |1.466406613410248e+100 | 10.0 |7.303308756353538e+305

200 |1.7255405787286054e+136 | 8 |3.126865905445952e+104 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-8.224081940224896e+209 | 1 |8.203245763937687e+217 | 0.1 |5.6825041335030605e+212

50 |3.937508283616992e+212 | 2 |-8.395452854425826e+214 | 1.0 |nan

100 |5.57091308492374e+215 | 4 |4.861488940297035e+208 | 10.0 |nan

200 |1.8332366619617454e+218 | 8 |-2.2694496879890996e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |33.105582706219536 | 1 |1.5630491588013527e+115 | 0.1 |1.4727007957547217e+112

50 |-6.3414889009447405e+78 | 2 |6.199105835058454e+113 | 1.0 |-9.307304647200071e+208

100 |1.477835025061924e+111 | 4 |-1.5295463149263359e+106 | 10.0 |-1.74294270221403e+305

200 |-1.3115603084634156e+142 | 8 |-4.77009194416044e+106 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.873972668316769e+209 | 1 |1.373546009524462e+218 | 0.1 |1.0578578638484961e+215

50 |3.595468240434825e+211 | 2 |7.539515897157456e+214 | 1.0 |nan

100 |-1.0162745197972116e+215 | 4 |1.7153280802927605e+208 | 10.0 |nan

200 |4.6618795191914155e+216 | 8 |-7.898400918184576e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |89994629695.42435 | 1 |-3.208314467281618e+114 | 0.1 |-1.2518958987789966e+105

50 |1.444738299903465e+82 | 2 |3.619371359215702e+114 | 1.0 |2.175857248653924e+201

100 |1.5897896290177855e+102 | 4 |-5.501833361503429e+105 | 10.0 |nan

200 |2.047570016260969e+139 | 8 |5.403400960668278e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |8.848725321818371e+208 | 1 |3.3062178159867884e+218 | 0.1 |3.494100574486856e+215

50 |-7.115310991424298e+210 | 2 |-1.584765859958402e+215 | 1.0 |nan

100 |8.431010787221646e+213 | 4 |-1.0329736134021042e+210 | 10.0 |nan

200 |4.395497729266238e+215 | 8 |-2.4751174069678843e+208 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-17152707.216563214 | 1 |-1.6465087566325106e+118 | 0.1 |3.219298049499785e+112

50 |-4.265533157264016e+83 | 2 |1.917523908998881e+107 | 1.0 |7.5067469798538e+210

100 |-4.462915271504992e+110 | 4 |-3.0904895085645947e+111 | 10.0 |-inf

200 |2.8629310726257233e+137 | 8 |4.7756328206790796e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |4.068912961491508e+212 | 1 |1.522667941803755e+217 | 0.1 |1.6504300392110376e+215

50 |6.412664311668106e+211 | 2 |1.4034162141939654e+213 | 1.0 |nan

100 |1.0585121870743924e+213 | 4 |9.213106964507121e+210 | 10.0 |nan

200 |1.8045832491496834e+217 | 8 |-1.0980292604829625e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.1770998343497352e+17 | 1 |5.047352802684239e+114 | 0.1 |2.2696399534563454e+113

50 |6.268825050991161e+75 | 2 |1.2783235846402743e+107 | 1.0 |-3.001389098428917e+206

100 |-1.3074613842700667e+113 | 4 |1.1971839419182339e+110 | 10.0 |nan

200 |1.1661443975666693e+142 | 8 |3.4134063410387392e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.62478106309587e+210 | 1 |-6.534151327883731e+218 | 0.1 |2.8324925217693804e+214

50 |8.2143060827167e+211 | 2 |4.21448216383771e+215 | 1.0 |nan

100 |-9.58058361817666e+214 | 4 |4.837453638750577e+211 | 10.0 |nan

200 |3.085493673770858e+217 | 8 |-4.1975075109402567e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |4.435342061440804e+24 | 1 |-3.8811753084542345e+113 | 0.1 |1.0901504632933053e+109

50 |6.282695240051364e+80 | 2 |1.1114834239119378e+115 | 1.0 |9.633530399950132e+206

100 |4.858703459778015e+105 | 4 |4.069225306586355e+113 | 10.0 |nan

200 |-2.995684584941822e+137 | 8 |5.834678622014509e+102 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-3.862262252937514e+210 | 1 |7.004046467872871e+217 | 0.1 |-4.8938811413925154e+213

50 |2.346008584552684e+213 | 2 |3.579987900835508e+212 | 1.0 |nan

100 |-8.756078561599857e+212 | 4 |6.043446677041595e+210 | 10.0 |nan

200 |3.142337975229202e+217 | 8 |-1.6133585948663743e+205 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.4619902258800357e+30 | 1 |9.88961281654279e+112 | 0.1 |4.986817586407131e+107

50 |1.1339567821752384e+87 | 2 |-4.9300775127757725e+109 | 1.0 |8.335699028630668e+207

100 |2.860835552437352e+109 | 4 |6.096607826792125e+102 | 10.0 |nan

200 |4.446987731317298e+143 | 8 |2.5358681834882754e+103 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.4284456088697287e+207 | 1 |-2.7094011360412828e+218 | 0.1 |4.973895400368286e+213

50 |1.488694399642186e+212 | 2 |2.0705802400736193e+214 | 1.0 |nan

100 |1.9588899193613342e+213 | 4 |1.8434709595352786e+210 | 10.0 |nan

200 |-5.13094594325114e+220 | 8 |2.4939525532758424e+209 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |2.1609609580098617e+31 | 1 |1.0858514710929044e+105 | 0.1 |1.5620475703756194e+106

50 |-7.859725662753783e+86 | 2 |-1.4128151751261687e+114 | 1.0 |-1.5439256686766505e+213

100 |-1.5995383708859058e+109 | 4 |3.611190978164333e+102 | 10.0 |-inf

200 |5.495569877393121e+138 | 8 |1.441437064302413e+105 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-9.740011772478286e+212 | 1 |9.381402300776995e+217 | 0.1 |2.4287051174030895e+215

50 |1.1246420752563472e+212 | 2 |4.885153678522617e+215 | 1.0 |nan

100 |8.1487792926290535e+214 | 4 |5.55319822077979e+210 | 10.0 |nan

200 |5.470372376975281e+217 | 8 |3.083254670672823e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |3900377344491932.0 | 1 |-5.887775137223824e+112 | 0.1 |3.006108223019421e+105

50 |-1.542943610570632e+77 | 2 |9.377051425520647e+112 | 1.0 |-1.1200379664711533e+209

100 |4.5250197509263984e+111 | 4 |1.8184681715117387e+103 | 10.0 |nan

200 |1.0016886119644676e+140 | 8 |2.8611343669399314e+107 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |3.226178796000382e+210 | 1 |2.1716421468530835e+216 | 0.1 |3.920901825295131e+213

50 |4.801961978507288e+212 | 2 |-4.539338147523738e+214 | 1.0 |nan

100 |-3.740930037902427e+215 | 4 |9.200630584344656e+208 | 10.0 |nan

200 |-2.59798310222858e+218 | 8 |4.096517059904767e+207 | 100.0 |nan

Table: Training accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |-1.8767451639943852e+18 | 1 |-3.7360565423263436e+111 | 0.1 |-4.9213704024168455e+107

50 |7.184319948227954e+78 | 2 |-7.468730060892563e+101 | 1.0 |-1.5819334641347882e+204

100 |1.3258701633847002e+106 | 4 |2.0220198154728718e+109 | 10.0 |-inf

200 |-1.922588738481227e+142 | 8 |2.3285122341524413e+101 | 100.0 |nan

Table: Test accuracies with different hyper-parameters

n Train Accuracy | d Train Accuracy | eta Train Accuracy

10 |1.6772170644314406e+208 | 1 |-3.2129071383124196e+218 | 0.1 |-5.470861587194502e+213

50 |9.442963359332107e+212 | 2 |7.346182358752401e+213 | 1.0 |nan

100 |1.153943092266797e+213 | 4 |7.771180621489425e+210 | 10.0 |nan

200 |-1.6000391726757255e+218 | 8 |4.3407632922284933e+207 | 100.0 |nan

2e)

The data that I seen in the tables is that I realized that the eta hyper parameter has a huge impact on the data points. It is because I seen they cause the training parameters to have high values than the cause of other hyper parameters. If the step size if very high, then the average of the w vector has a high value. Also, as the dimensions (d) increases, then the test data decreases. It is because when the vector w has more parameters, then it gets closer to the curve. The other thing we notice on the tables is that the more of the test case points can causes the averages to have a bigger value. If there are more amount of data points then there will be a bigger average.

Names: Michael Balcerzak, Kuba Potera

Student Numbers: 101071699, 10115432