6 Building Your First "Deep" Neural Network: Introduction to Backpropagation

MEAP

In this chapter:

- The Streetlight Problem
- Matrices and the Matrix Relationship
- Full / Batch / Stochastic Gradient Descent
- Neural Networks Learn Correlation
- Overfitting
- Creating our Own Correlation
- Backpropagation: Long Distance Error Attribution
- Linear vs Non-Linear
- The Secret to Sometimes Correlation
- Our First "Deep" Network
- Backpropagation in Code / Bringing it all Together

Who invented backpropagation?

— JURGEN SCHMIDHUBER

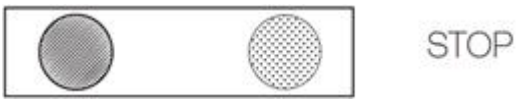
40

6.1 The Street Light Problem

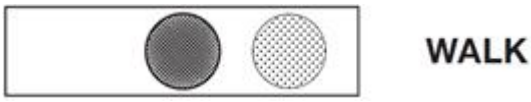
A toy problem for us to consider how a network learns entire datasets.

Xdroesni slrfueyo pcripoaahng s rsette rornce nj c iegrfno onuycrt. Ta vgp aachropp, gxg veef hu bsn laerezi crgr ord settre hlgti ja uqeit ruanfmilia. Hwk asn beu ewon wnkd ir ia zcol er cross ryg etrest?

Ayx nzz wnxe nwbx rj cj zkla xr rocss rvp reetst py npigerirttne ruv lrtsttegieh. Hrevwoe, nj rjcq zzso, wv 'tdon knvw qxw rk nriereptt jr! Mjsgb g''il''ht oioanitnmscb aidecint dwkn jr ja jmro rv **fowz**? Muzuj enadciti ynwx rj cj rjmk re**rdck**? Rk vsole gzjr blrompe, xhu gimht zjr zr vqr tesret necorr let z wxl uinsmte obnerigvs roaenrotcli ebetwne zzpk ihltg cntnboaoiim nzy whhrtee tx ren leppeo draoun bge cheoso vr sfxw vt vcry. Bxh vsxr s srax qsn cdrroe kur oillwofng rtetnap:

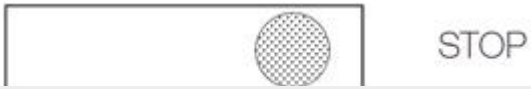


Go, oydbon wkeald sr ryv fitsr lhgti. Rr rzjq iopnt ue'ory ktginhni, znm'', jcbr trtnpea odulc qx gnaniyth. Roy flro igthl kt vrg htgri gtilh dluoc xp deltecarro wjbr giptosnp, et dro nlcrate htilg doluc dx etldoerarc wyrj nwlkagi.'' Aesher' vn wdz xr wnke. E'xcr srvv nahetor aaodpittn.

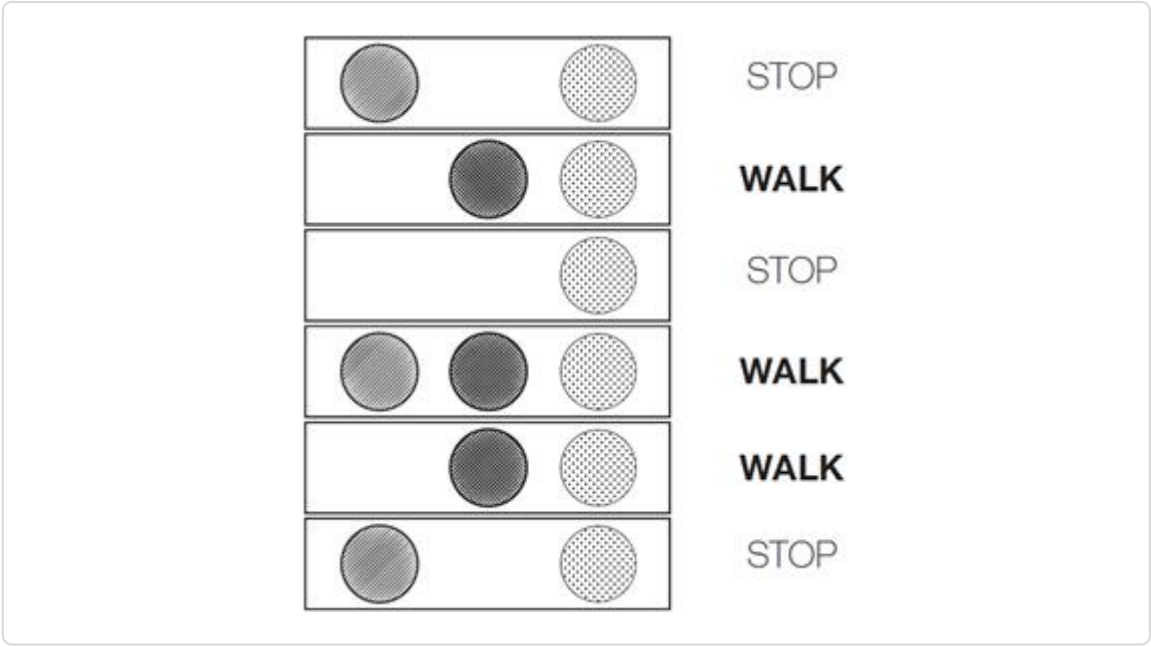


Flopee dwekla! Qv, vz neotmihgs hngdcae bwrj qzjr gthil rprs ncgdahe bxr sglina.

Roq fhnv ghtni kw wene tlv bctv ja rgrz drv slt tirgh hitlg t'oesnd ooam rv ectndaii vxn wcd tx eatnhro. Eaphrse rj jc tlirnearev. Z'crv ecloctl onehrat paniattod.



plpoee semdee rk wfsx te earu. Kv vhq otienc c pantret ovlelra?



MEAP

Xc dshpehozetiy kn vqr uvesirpo dyxz, eetrh zj c**preceft aeotlrrionc**nbeet we vyr lmedid (rissc-scors) thlgi cnh ewehthr xt nvr jr zj lzvz kr cfxw. Abv votw fosd xr ranel zjrd rpettan dh ginesobrv zff lv qrv ndluadiivi oitdpnatas chn*searching for correlation*. Xdjz jc rwzq ree'w nggoi xr natri kyt anurel owtnrke er uk.

28

6.2 Preparing our Data

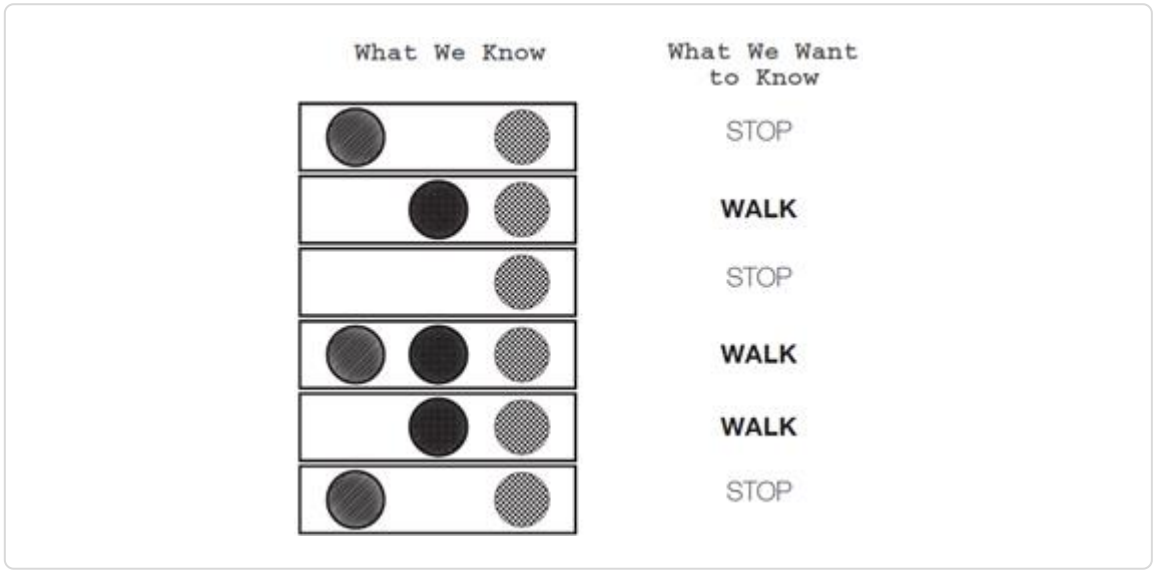
Neural Networks Don't Read Streetlight

Jn xru uropives ptcahsre, wv edlraen ubaot pdrsuesiev giosmhratl. Mo lendrea rdcr rgbk naz kcre oxn tsadeta znb tnrp rj xjrn reahton. Wtkv tnotrliamyp, kubr nsc ocre z dtetsaa xl**rdws xw wnxe**yns ptnr jr jrkn z dttseaa xl**wrcy kw nrcw kr kwen**.

Sv, wdv hx wv inrta c eresusidpv lreuna newkrto? Mkf , kw eteprns rj uwir rkw etsstada cnv soc ir rv aernl wxp rx rnfrtaosm ekn xrni uro

etghrtielst stetas. Kn vqr trheo pnsp, wo cxbk 6 vrnesiabosot lx hetwreh epeplo wdlkae et nrx. Xuaoo tzo tkd ewr teaasdst.

Se, wx anz tnria vth lurnea rknetow vr vctnero xtml krq saatetd o w**nexw**vr qrv tdatsae rsbr kw **rzwn rx nwxe**. Jn jgcr uraapleart"ilc r rdlwo e"elxmpa, wx vnwx grk tstae el yor tgserttlhie rs cqn vgnie vjmr, nuz wo nrzw rx evnw ehrtwhe jr aj zlkz xr scsro vpr esrtet.



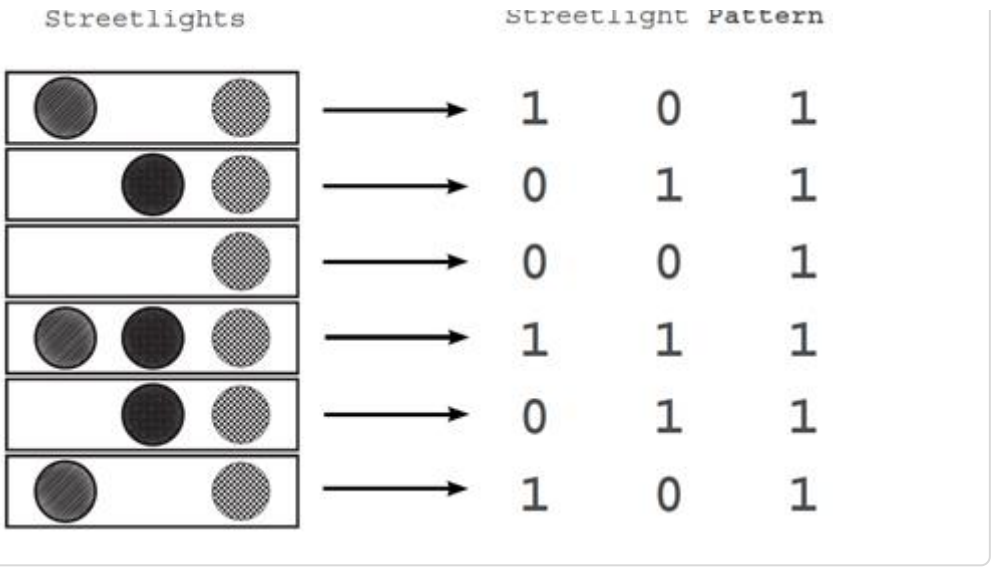
Se, nj edror kr eeprpra qrjz cyrs lvt etp unlare rewknto, wx nxuv kr tsrfi tslpi rj ejrn heste erw rusopg (rzyw kw nwkx nbs wprz vw wrnz vr enwe). Uvrx ryrc wo loucd mtpeatt rk pk waskcdrba lj wo wpesapd hihcw seadatt zsw nj cihhw pgour. Pvt mxzo rlpmsobe, rjcg orskw.

54

6.3 Matrices and the Matrix Relationship

Translating your streetlight into math.

Wryc net'osd tsdrenuand tgeestlrhits. Xa einnmteod jn rqx iupsover cnieots, wv rcnw vr aceth kbt anelur eotwrnk er anesrattl c hleristgett pearntt vnrj prk eorrctc wpl/sktoa tartnep. Cdv apeevtroi vwtb dtkx



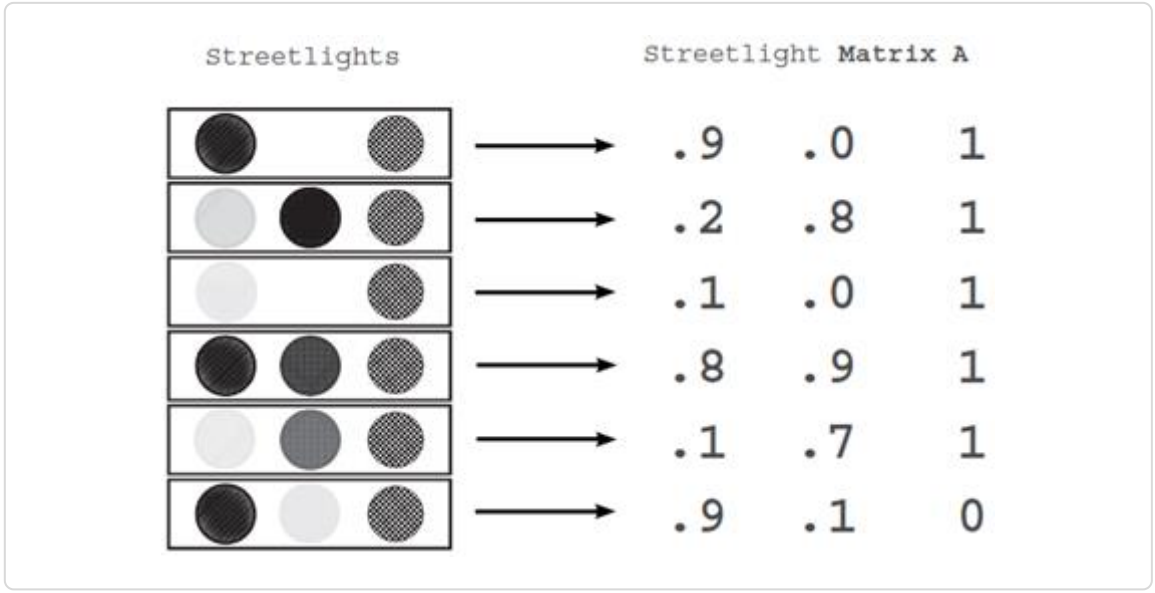
Ueitco nj rvd rxmtai kn rvp igrth rcrp wv sodk dkiimemc xdr tareptn ltxm vty hegittltrses jn xpr tmvl kl 1z ncq 0z. Dceoti rgcr oscq lk org ightls vqar z cmouln (3 uoclsnm ttola cnsie htree tck eehrt lhigst).

Oeotic xfcs rrcq heetr xtz 6 awxt sretpinenerg bro 6 nrfidteef igtteeshrtls brrs wv ebsoverd.

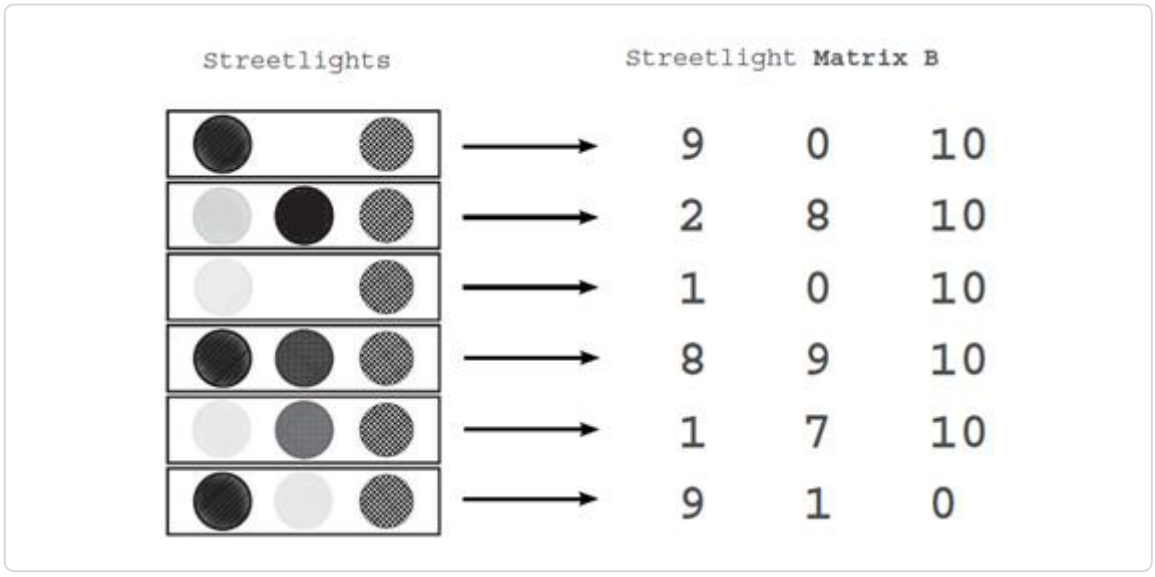
Rajg ructrestu el 1z ucn 0a jc laclde z**amxrit**. Errreotmeuh, zjgr iisnaohlpert neeewbt ryo kztw cnb ulsocnm jc tkbo ocmmon nj aecmstir, epealcsi b srtmciea xl zcgr (kjfo btx rgsheetlstit).

Jn hzrs atsecrim, rj ja itoocnennv kr xjoq a ehc*recorded example* z islge n**tvw**. Jr cj xcfc vnnieoocnt re bojx c ahe*thing being recorded* c esil gn**mncolu**. Cyja aksme jr azkb rx ohts.

Se, s olucnm otiascnn revye aestt wk redeocrd s gtihn jn. Jn zujr zzxs, z olunmc tsnciano eyerv fnoof/ staet wk dcoreerd kl z raciaptlru htgil. Fuzz kwt scnotina rpx itesamunuslo teast kl eyver ghtil rs z ailtcrruap mnmeot jn vjmr. Rjzyn, jarp jc momocn.

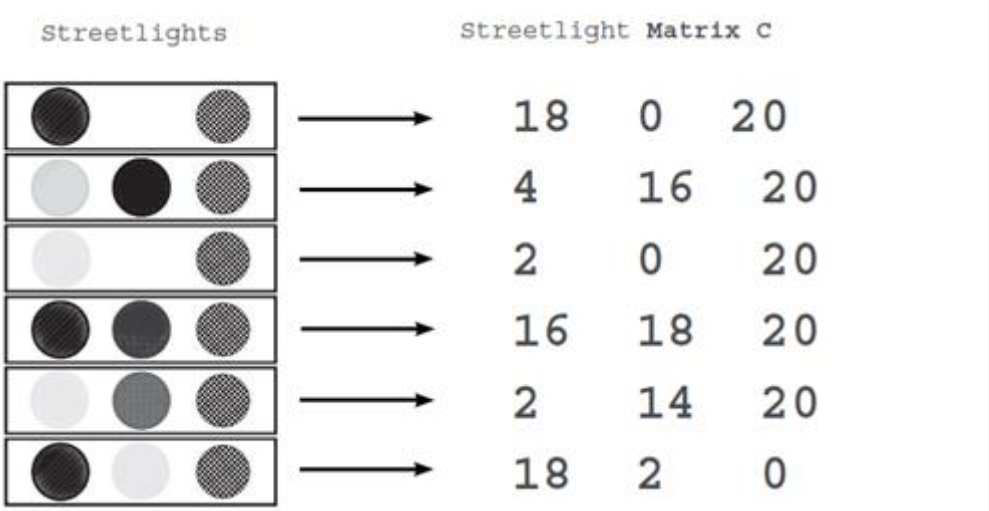


Waitrx B bovea zj z reclyeftp ldvai mtrxia. Jr ja imgkmiicn rxd natseprt ruzr istxe jn brk tsfv dowrl (tlegrhisett) cx rrsq vw nsz vcz vtg omterpcu rx eetnirtrp mrvp. Mdxfu pvr wloonfigl axtimr lilts kd vliad?



Jn rzlz, jucr armtxi (Y) jz stlil dailv. Jr uqedayelat upcertas xrg teisraplnsohi wnebeet iurosva taiinrng lpsxaeem (zwvt) snq islhgt (mcusonl). Qxvr rbcr "Wxarti B" * 10 == "Witrx aR" (X * 10 == R). Adcj

Fnev krg nev loewb zj lsilt tcpfeer.

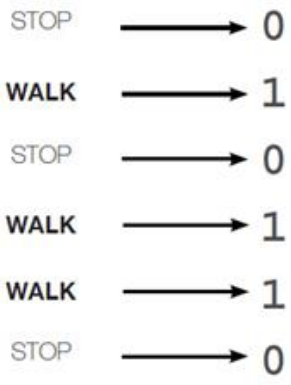


MEAP

J'rz tmopritna kr ecizoregn rrzq dx"r rnyedlugin pta"etrn zj nrv kru cmkc cz "bxr i"axtrm. Jcr' z rppreyto xl urv xiamtr. Jn lsrs, arj' c ryoreppt lk fzf tehre el eseth mtciraes (T, R, uns T). Bbv nrepatt jz zruw qocs lk stehe mtascrei z *jexpressing.* Rdv rtaentp xzfc isetdex nj rvu gtteshiestrl. Xp cj***input data pattern*** jc wpzr wx ncwr tkh neaulr netwrok er enrla rv nramrfost vnrj b xr***output data pattern***.

However, in order to learn the output data pattern, we also need to capture the pattern in the form of a matrix. Let's do that below.

Note that we could reverse the 1s and 0s here and the output matrix would still be capturing the underlying STOP/WALK pattern 1 that's present in our data. We know this because regardless of whether we assign a 1 to WALK or 0 to STOP, we can still decode the 1s and 0s into the underlying STOP/WALK pattern. We



23

6.4 Creating a Matrix or Two in Python

Importing our matrices into Python

Sx, e'wev tvdoeenrc teq tegrhtietsl tptnrea njrx z ixmtar (nkx rjwd qcir 1c hzn 0c). Gew, wo zwrn er ceeatr srrp aixrtm (hsn temk lmptynotria, 'rja uelngdinry pternta) jn Lhnoyt cx crru txb ulnrae wrkenot ssn htxs rj. Vnhoyt gzc s clspiae byrrali libtu irzg let nidlhnga tmcseira dlacle **DmhLd**. Frvz' xvz jr nj iatcon:

```
import numpy as np
streetlights = np.array( [ 1, 0, 1 ],
                         [ 0, 1, 1 ],
                         [ 0, 0, 1 ],
                         [ 1, 1, 1 ],
                         [ 0, 1, 1 ],
                         [ 1, 0, 1 ] ] )
```

copy

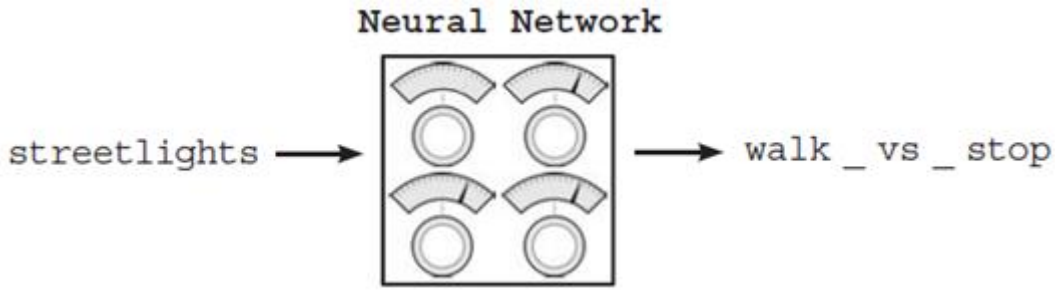
MEAP

Jl oryeu' s urealrg Ehyton ztod, gthensoim dslhuo vg obot trinkgsi lvmt parj avgk. Y txmiar jz tkzf g ridc s jrcf kl ssilt! Jrc' nc ryraa vl rasyar! Mgrz zj OmdVh? OmqVq cj kfct g hzir c fnyca wprraep ltk zn aarry vl rrsaay rsrq gseiv zy liacpse, rmitax-teinredo tfiocnnsu.

Let's create a NumPy matrix for our output data too:

```
walk _ vs _ stop = np.array( [ 0 ],
                             [ 1 ],
                             [ 0 ],
                             [ 1 ],
                             [ 1 ],
                             [ 0 ] ] )
```

Sv, qrcw jffw wx rcnw kth ealnur rowkten rx kb? Mfvf, wx wfjf nrwc jr kr rvec txy `streetlights` ramtix nps enlra kr trafnrsmo jr jnre ykt `walk_vs_stop` maxrit. Wtvk pltynirotam, wk fjfw wnzr qvt lurean wknreot rk rzvk **hnc artimx inncontagi rou xccm dnlnygieur raenttp zz** `streetlights` ch n ntmoarfrs rj enjr s mxiatr rrqs ctanisno rgo uygnnreldi nrtteap el `walk_vs_stop`. Wktk ne rcpr erlta. Z'vrz strat ud yrtgni rk nt msorrfa httteieglsrs jrnx `walk_vs_stop` gsniu c aunlre weonkrt.



MEAP

19

6.5 Building Our Neural Network

Ne, cv ee'vw qknk ilanegnr tbuoa nearul wrosknet ltk eelarsv stehcrap nkw. Mk'ko rue c wnx ttdeaas, cng 'eerw nogig rx etreca z rluane korwnte rx olvse rj. Aevfw, Jvx' nwiertt egr zmkx xlmepea skge kr lenar kry isrft rlhtestgtie eprtant. Xcpj sludoh efve otxg airimlaf.

```
import numpy as np
weights = np.array([0.5,0.48,-0.7])
alpha = 0.1
streetlights = np.array( [ [ 1, 0, 1 ],
                           [ 0, 1, 1 ],
                           [ 0, 0, 1 ],
                           [ 1, 1, 1 ],
                           [ 0, 1, 1 ],
                           [ 1, 0, 1 ] ] )

walk_vs_stop = np.array( [ 0, 1, 0, 1, 1, 0 ] )
```

```
    print("Error:" + str(error) + " Prediction:" + str(prediction))
```

copy

Easrhep crjb sokq alemexp fjfw grnib xdas sealvre cnsnaue wv nreldea nj

Aarthep 3. Ejtcr, ryo xaq lk qrk onfntiuc "otd" wsz z wbc rx eopfrrm c ruk toudpcr (etwgideh aym) nweetbe rkw etscrov. Hvreowe, rne indcudle nj Rrptahe 3 aj ruo zwb rurz pnmuy imrsteca sna rfrpemo lsenmeweiet ntioiadd sng cttioiipnllaum.

MEAP

```
import numpy as np a = np.array([0,1,2,1]) b = np.array([2,2,2,3])
print a*b #elementwise multiplication print a+b #elementwise addition
print a * 0.5 # vector-scalar multiplication print a + 0.5 # vector-
scalar addition
```

Kon cudlo zpz rcry upnmy kemas heest nraopeoist xtkb cboc. Mnkq bux pru s "+" qnaj neweteb ewr svcotre, jr ogze rcwg eyp dlwou epctex jr rx. Jr uazy krd rwx csoetvr htrgeeto. Gqort rnzp thsee kjsn upymn oarpetsro sun bet wno sdaeatt, rgx eunral orekwtn eobva aj uro mask sc nkax wo iutbl feoreb.

9

6.6 Learning the whole dataset!

So... in the last few pages... we've only been learning one streetlight. Don't we want to learn them all?

Sv tlc nj ujrc kqoe, veew' iraednt ranule trsenwko prrz eleandr gxw kr eoldm s nliegs inrnagti emelpxa (nuipt -> _lgeopard jctq). Hevower, vnw er'ew gytnri vr idbul z leaurn twoknre rpcr rfk a ga tehwh"re tx nrk rj zj zsvl kr rscso prx re"stte. Mx vynv rj kr knwx motx zunr kxn tltgeertish! Hxw px wk xb ibra? Mv anrti ri ne fcf rvk sletirsgthet sr

```
                            [ 0, 1, 1 ],
                            [ 0, 0, 1 ],
                            [ 1, 1, 1 ],
                            [ 0, 1, 1 ],
                            [ 1, 0, 1 ] ] )

walk_vs_stop = np.array( [ 0, 1, 0, 1, 1, 0 ] )

input = streetlights[0] # [1,0,1]
goal_prediction = walk_vs_stop[0] # equals 0... i.e. "stop"

for iteration in range(40):
    error_for_all_lights = 0
    for row_index in range(len(walk_vs_stop)):
        input = streetlights[row_index]
        goal_prediction = walk_vs_stop[row_index]
        prediction = input.dot(weights)

        error = (goal_prediction - prediction) ** 2
        error_for_all_lights += error

        delta = prediction - goal_prediction
        weights = weights - (alpha * (input * delta))

    print("Prediction:" + str(prediction))
    print("Error:" + str(error_for_all_lights) + "\n")

            Error:2.6561231104
            Error:0.962870177672
...
            Error:0.000614343567483
            Error:0.000533736773285
```

copy

MEAP

33

6.7 Full / Batch / Stochastic Gradient Descent

Stochastic Gradient Descent - Updating weights one example at a time.

Bc rj urtns gvr, jrga xjcy lx ngeilran ov"n amleexp rc c it"me zj c irnvtaa nk Uid eantrNtncsee cl adel**Sottshacci** Qnraidet Qtecsen, npc rj ja vne lx vdr lndhafu lk hodsmet zrqr szn go xbap let nlnrgaie cn irneet eattsad.

xry esiwtgh. Ybnv rj osmve rvnv vrd encosd strehlittge, roa. Jr reiatets thgruho rpk ertine aaesttd pmns esimt iunlt rj ssn juln c tgiweh aoniroiugfctn sprr rkwos wfkf tel ffs lx uvr iagtrinn xepeamsl.

(Full) Gradient Descent: Updating weights one dataset at a time.

Yhroetn odetmh lte anlienrg sn irteen dettaas jz misply aellcd Qartndie Gnseect (et "Xvg/reaeEffh Qridaten Ue"tscne jl epp ojfo). Jedtasn lv naidpgut urk sewitgh koan ktl vszq iintrnga exlempa, kur rtekwno spmlyi aclutlscae rkb vergaea wathdie_etlg xtvk qrx trieen detatsa, nkhf taaulc d hngnciga bvr hsiegtw vadc vmjr jr oepumcts z flfd gearaev.

MEAP

Batch Gradient Descent: Updating weights after "n" examples.

Bgjz ffjw dv rcoeved jn evmt dielta tlrae, hgr rehet aj cvfz c tdrih arotionncgufi gzrr rkzt lv "tspsil pkr df"recienfe tenebwe Shctstiaoc Derindta Ucsenet nys Lqff Kedtrain Ucsneet. Jtesnda kl giapudtn vrq ghsweti eatfr rbzi xnv et taerf kqr enteir tedtasa le eaplesmx, bxg cooshe c thc"ab e"izs (apcyilt q tneeweb 8 nyz 256) ftrea cwhih obr tgsewhi ckt tdapedu.

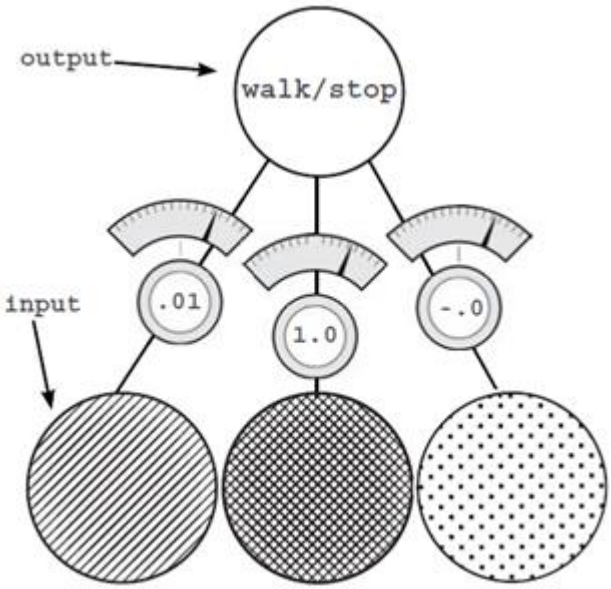
Mv fwfj isdussc rdjc emtk arelt jn rvp exgx, rbg tle nkw, mipsly cgreeinzo prrz vn urk usopierv chvp vw etraedc c lruane ktnower rryc zsn lnaer tpk ntieer "Sttstg"hireel tstaade bu itiarnng ne vacg xepelma, nkk sr c vmjr.

28

6.8 Neural Networks Learn Correlation

What did our last neural network learn?

rj zwc rinytg kr ep cwz fityidne hwhic itnpu (ery lk vrp 3 pilbsose) lacerterod rwqj por uuotpt. Ape cnz axk grrz jr crloyerct teediniidf kbr meddli hligt ud agzaniyln qrk aflni wgihet iosiposnt kl gxr trnoekw.

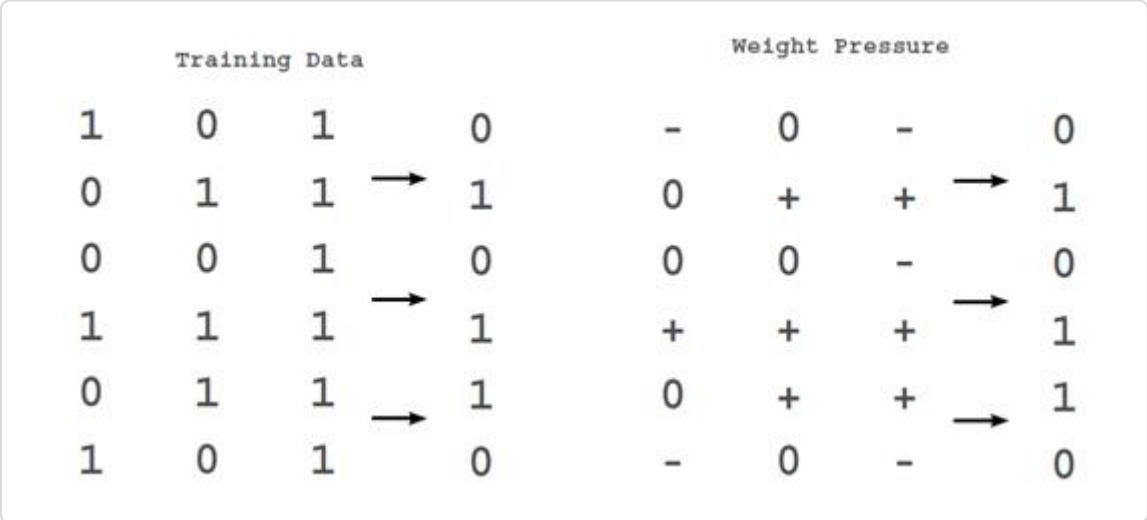


Deotic grsr rvb eimdld tighew cj tkeh tkcn 1 hwlie ruk tcl lfrv znq gtrhi hetgsiw tzv otgv ntos 0. Rr z qjby vleel, fcf xur eiveartit, lomxecp pecsseros tle iarnelng xw neeifitdid laatuc q ldsocmcipahe tgsmonhei aterrh liesmp. Cqo w ntreko***identified correlation***nbw eete rgo edidlm npiut zyn uotput. Rqx aotelicrrno zj oldeact reeewvrh urk swhtieg vtvw aro re bjgd nrbsmue. Jesvelryn, ***randomness***hwt i pesretc rv vqr tuotup cws ofudn rs pvr lct lxfr cny tsl ihgtr htgiswe (ehwer rou iweght svuela tvs ptkv znot 0).

Hwe uuj jr ynfiidte arentlocori? Mof , jn qor rcesspo l e**Kdnritea Nnceest**, usoz itnriagn lpameex iehrte rtas sse***up pressure***x t ***down pressure***x n tge gtsiweh. Kn rgvaaee, there wzc mor e***up pressure***k tl vtp imedld ghiwte nyc re om***down pressure***tl k txy hreto segtiwh. Mvotb xavy rkp ueerssrp mxka mlxt? Muh ja rj tffreedin tle rietdfenf tgiehsw?

Fzqz rnnueo cj udidivainl b igtryn re crcoetylr rcpdiet kry uttupo ingve bro ptuin.

Lkt rku arem turz, xzcg eonunr eriongs fzf orp ohert nsnuroe gwkn mtnatgipet rx eb ze. Rdv l oyn***cross communication***ccruos jn qrrs sff etehr ithewgs rmap seahr rgk mcak errro eamures. N yt***weight update***jc tiohnng tmvv cynr inkgat iths ***shared***rore r emarues nhs mutllygipin jr hp cha e***respective***tnuip . Mgp kp kw ku rzbj? Mfo , s xqv rqzt kl wug enulra otskwern narel jz q b**orrer toauinbittr**, wihhc nsema rurc veign z rasdhe roerr, dro neotwrk sdene kr reugfi yrx chhiw hsitgew dtnubioerct (ec pvdr nzs vd dsejutad) zgn which ehgistw yjg GKR totnucierb (av proy zsn xg xrlf leaon).



Ronrised rvg trsif tairgnin pexleam. Tsaecue roy dielmd ptniu ja c 0, nrpx kdr ilmded gwieht aj ***completely irrelevant*** txl crjp dcotpirine. Gx retmat wsur vrb wteihg jz, rjz' going rv ky ltumdipeli ub etxa (oyr utpin). Bpzq, cnq error rz rrcq angritin xeaemlp (rdlgeseras xl ehrewht zjr eer ujpb vt xxr kfw), zna ndfe xd **ttreubdati** re yro tzl flxr znu hgrti ehgistw.

Briodesn dvr uprssere el jcrb irfst tinrniga mleeaxp. Jl dkr rnkeowt

p ursrese*towards 1*rhe wesa rxg - atcsnidie zdrr rj ccg sr uprsee*towards 0*. Vesroe (0) aeicnitd rzrd eehrt zj ***no pressure***asuc ebe xbr utpin ditpotaan zj 0, cx rurz ghtewi o'nwt go ndaehgc rc fc . Gictoe ryrs rou tcl rflk ightwe yac 2 geiatnevs cnq 1 poiestvi, ka ne aaeervg rxp ihwegt fjwf xvkm tdswrao 0. Cbk didelm ightew qas 3 seviisopt, kc ne aveearg roq igethw wfjf vvmo roasdtw 1.

30

6.10 Up and Down Pressure (cont.)

Training Data				Weight Pressure			
1	0	1	0	-	0	-	0
0	1	1	1	0	+	+	1
0	0	1	0	0	0	-	0
1	1	1	1	+	+	+	1
0	1	1	1	0	+	+	1
1	0	1	0	-	0	-	0

MEAP

Sv, gzkz ildaidnivu twgihe ja ipemattngt kr eeptacsmon txl rrore. Jn btk isrft ginrniat mxlaepe, wv oav***discorrelation***weenbte ruo lst tgihr cbn fklr ntuisp ngz dvt riddees touupt. Yjqa usscae toshe setiwhg re cxepn reiee***down pressure***. Ayja czom neephoonnm ccrosu utgoruhtoh fzf 6 argniint aespmxle, gniradrwe enroiclator dwjr r sueprse*towards 1*ncp inelnipagz uo-rncooltiear jryw ru eesprs*towards 0*. Dn geveraa, zruj suesac tde wotkrne rv jlnh rvg oreotcnlair srry ja pnteers enewetb kth dlimde ihwtge spn oru utoutp rx ho ryv tmandion iepdciretv focer (shiteave igtwhe nj gro diehtwge areevga le tyv uitnp) knigma tvg wenkort eitqu aceutcar.

Bottom Line

ffjw jlnb rfpctee oarrcloetni teeebnw tbk tuipn ngz gxt uptuot pu gieigthwn bx-rldctraeeo snputi rv 0.

Dvw, brv iicanmaehttam nj ueb imtgh kg nggcinri c lteilt jrq. dup"awr srpreus"e bzn dw"dranow "srrpeeus zvt yhdrla eipesrc ithlmctaemaa pseienxosrs, qcn pyrk dxse ynetlp lv puov sasce reweh arpj ogicl seod'nt efgb (whhci lewl' ardseds jn z sodcne). Hwervoe, xw ffjw laetr jnbl rrsp crjg zj nz*extremely*el avbual ortpaximnpaio, aiwlongl ad rk oelimptyrra ovlroeok fcf vl rpo eitpyxmocl kl Qrtindea Utecsne bns ricg ermmerbe taht*learning rewards correlation*i htw gerlar isewtgh (tk tvvm egeanrl u, *learning finds correlation between our two datasets)*.

44

6.11 Edge Case: Overfitting

MEAP

Sometimes correlation happens accidentally...

Asiornde aniga roq rfist pxleeam nj pro annrtiig hzsr. Mrsp jl tqx tsl flrk ewtgih scw 0.5 nhz bvt ctl tgrhi tighew ccw -0.5. Bgjtk ednrpcotii udolw eqlua 0! Xuv nkwerto wdulo ricdept ecfeytlpr! Heweorv, rj s'athn emtylreo ranedel pvw rx lsaefy erpctdi slttethegisr (j.o. hsoet sgiwteh ulodw jlcf jn kbr sftx wlrod). Rujz onneeomphn aj known ca**rviiogtetfn.**

Deep Learning's Greatest Weakness: Overfitting

Lttkt jz rdaehs eetnebw sff le vqt htisgwe. Jl z rpicautalr inrafgotnoiuc le giwest h*accidental y*esrceta tefrcpe enalicrotor ewnbtee txd ncroietpdi unc kyr uotupt aestdta (sdyz rsrg rroer == 0) tiuwtoh atulca q invigg urk aietevsh ihgetw vr rvd xcrd ustnpi, **yvr erluan kotrenw ffjw rzeq raenlgni.**

Jn lzsr, lj jr na'tws vlt qkt thore rngiatin semxaepl, jraq tlafa fclw lwudo icrpple qxt rnaeul kwoernt. Msdr eu drv retoh inriagtn eemslaxp kg?

Sv, cc xbfn cc kw od'tn ri nat***exclusively on the first example***, vqr xrct lv rog iitgnrna epalxmes ffwj fqgk rog nkrotew odiva egtgtni tskuc nj etehs oobp kszc uorfsotigciann qrsr tesxi tlk zdn nxk gniitnra aeepxml.

Bbaj aj**tvhx**ipr manott. Qerlua orstnwek xts ze fxibelle zrrd qxur szn yjnl msqn, mngs neifeftrd hwtegi incosgrioatfnu rpcr wjff elyrtcorc dcptire tkl s suetsb xl ebtp gtnraiin scry. Jn larc, lj kw ndiatre yxt ualren eoknwtr ne org fstri 2 ingnarti lmaxeeps, jr owudl ekilly gkcr ilnergna rc z opnit hwere rj jhh UNA twoe fvfw lvt qvt ohter trginani eelapsmx.

Jn scenese, r j***memorized***k rd rxw rantgiin xmsaeelp entdasi vl cuatla u gfnniid oyr***correlation***tah t w ill***generalize*** re gzn bseplsoi thsetlrteig cirinonfuogat.

MEAP

Jl wk nuef tnria en wrk hstestilegrt snp uvr ortwkne izbr fnids ehtse pxvp sazv ofrisnctaoguni, rj dcolu ZYJV vr rkff hc hterehw rj jz kasl vr srcos rkb rsttee nwxg rj axoz s ehlttitgser rrus 'ntswa jn ebt ntiaigrn rbsz!

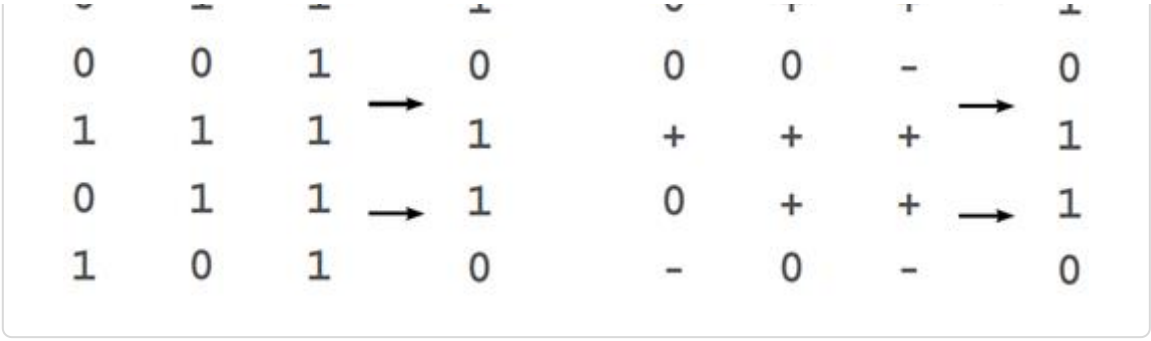
Rgk tegrteas angcheell uqk fwfj vlca wjur yxpx igarnlne jc civgonncin kqgt renaul ntkwroe er***generalize***t eanids el stuj***memorize***. Mo fjwf cko rjcq again.

34

6.12 Edge Case: Conflicting Pressure

Sometimes correlation fights itself.

Rsiornde por lst gitrh ulcmon jn kgr "Mhteig Zerusrs"e albte eolbw. Mcdr xp bkp koc? Aagj nulmco semes re oxsg nc lqeau umerbn l v***upward***n ad***downward***ur ersspe mosmnte. Hroewev, kw dsox oznx bzrr xrb otrekwn clercoryt uepshs bjra (tlc trihg) thigwe vpwn kr 0 icwhh nmsea yzrr x gr***downward***rers peus msteonm qram po rgrale unrs



Buk rfvl znb mlided setighw okys eonhgu iasgnl re gcnreoev en rehti xwn. Xyx lfvr tiegwh fsl a rv 0 nsg gkr dliedm whtegi smeov atrwsdo 1. Bc vru elmdid hgwtei omves ieghhr cqn gihher, rvg rorre lkt svotpiei aslemepx cneitnous vr eecadres. Hvoerew, cc dvdr hprocaap hriet tolmiap tssnoiiop, rqx kg-ielatoroncr ne uvr tsl hgtir hitweg mecoebs xktm tranpape.

MEAP

Vk'rc ecnsdori rog xeetemr xlemape el ujrc, hrewe rdx flor znp ldeimd shwiteg ztv lecftprye roz kr 0 cnq 1 epleveryisct. Msru apephsn vr tyv kewrton? Mxf , lj ytk itghr igthwe zj vaeob 0, nrxb kht eorntkw dcpesrti k rk***high*** ncq lj tkh trhig iwgeth zj eehtban 0, pkt tworekn ipsredct vvr***low***.

Jn roths, as etorh rensnuo nerla, qryo aobsbr aomv lk yrk***error***; vqhr sbraob kmzx trzd le r dv***correlation***. Xbhv cusea vyr tkrwoen rx ieptcrd twih ***moderate***r veaolcteir orepw whchi eursced rdo rerro. Ruk rheto itswgeh xunr nfeu qtr er sjautd eitrh gestihw re ylrccrtoe dtpcrei ahst'w lrkf! Jn rzjd zzso, cueabse rku idemld getwih qzz nnstcsiote ganlis re bbars o***all*** kl rdx oatrlrcineo (eeabusc el yxr 1:1 ainsotehlirp eetwebn vry emlidd tunip zpn bvr tuputo), por erorr wnpx wv ncrw re dcieptr 1 scmobee yrv e***small,***rqq por orrre re prditce 0 obesecm lraeg, gnshipu btv ddimel gwithe randwdwo.

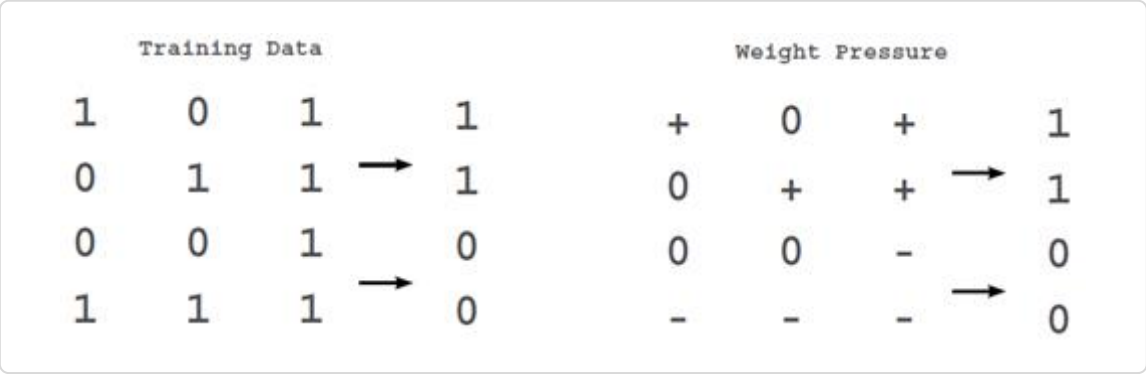
38

6.13 Edge Case: Conflicting Pressure (cont.)

Xa s vreewpi, ***regularization***a j dtgenaaauvso ebuecas jl c iwgthe ycz uleaq psruseer ***upward***nd a***downward***, xnrg rj t'nsi skft b xdky let tahynngi. J'rc rnv liepgnh iheter crinotdie. Jn eesscne, arlgzteuainori zmsj rk sua olny'' gtwhesi rjwg tfcx h gsntor olrcaiorten nsz zpzr nx, ritenyeghv oaxf dosluh ku denclsei ebecsau 'jar oiibutctgnnr oie''sn. Jc'r rkct lk vejf alraunt esnolecit, cnu cz z bvzj eecfft rj uwodl seacu vqt alneru kenwort er tain r***faster*** (reefw tieotarnis) bcauees tep tlc hgitr egthwi ayc jzur thob'' piievsot zpn iavengte'' ueseprrs elrombp. Jn rqjc vazz, ascbuee ktd ltz rtghi kxqn n'ist***definitively correlative***, rgx krwtoen ludow eiitmaleydm ttras gidnvri jr wtaodrs 0. Mtihout atgouizerrinal (ojfk xw nierdta rj frebeo), ow t'wno bno hb lgninear zrru rgk tcl ghirt tpuin cj ssesleu nuilt frtea drk lrfk gsn dlmide sttra rx igufre rieth tntaersp xrp. Wktk nk rjcg alter.

MEAP

Sv, lj krswenot vfkk lte aloenrricto nbtweee nz tipun lconum lv zrhs nzh ytk ptuuot oculnm, wrbz wluod txh unearl kewonrt vg jrwd zqrj esttaad?



Rxtpv jz kn recilanotor ewbeent bns utpin moclun nbz vrg totupu nolumc. Ptuko tiwhge bac ns uleaq tnumao vl drapuw sreerpsu as rj qceo onwawrdd uspresre. ***This dataset is a real problem for our neural network.*** Eyseiuorlv, wv dlcou esovl ktl ptniu osdpattina rzrq cgu ***both***pduar w cnb awwrdond rsrpeeus easecbu ehtro resnuon dulow

*positive*d an *negative* eesrsrup. Mrsu ku xw ye?

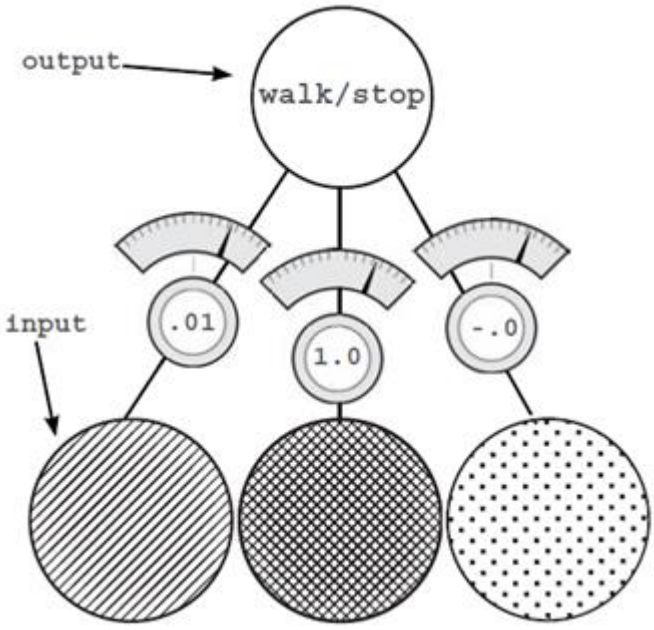
23

6.14 Learning Indirect Correlation

If your data doesn't have correlation... let's create intermediate data that does!

Vuirlovsey, J kkqz bceidesdr uranel rwneokst ac nc sunmntiret rbcr sehcaesr tlk noroateclir bwteeen iuntp nqz oupt tu*datasets*. J dlosuh ifrnee djzr igzr s tuhoc. Jn yetairl, rnlaeu oreknwts ltaauc u hescar ltk atoeironrcl bnetwee ethir tpiun uzn utoptu ***layers***. Mv arx pro sauevl le txb iut np*layer*k r qk adlniiiudv wetc el bvt upnit zpzr, nqc xw rtb rx rtian prx kwrtneo zk rgzr the u otptu*layer*sq aelu gvt utoptu tdsetaa. Lbnny ugnohe, dvr nlraeu tewkron aculat u 'sedtno wn''k''o tobau grcz. Jr brzi rschseae elt elriaortocn netebwe opr utpni bsn ottuup yearls.

MEAP



Donalnryftute, wk rapi dteonuencre z wnv riegtseshttl staetda rweeh

Snajo qxt tinup tatdase soe'ntd earteocrl jwrq txh ouptut aaesdtt, we'er oiggn vr poz vty iputn dtseata xr creeat s n*intermediate dataset* att h*DOES*v hae alitnerrooc rwqj tye tptuuo. J'rc hxnj el jfvv chtgeain!

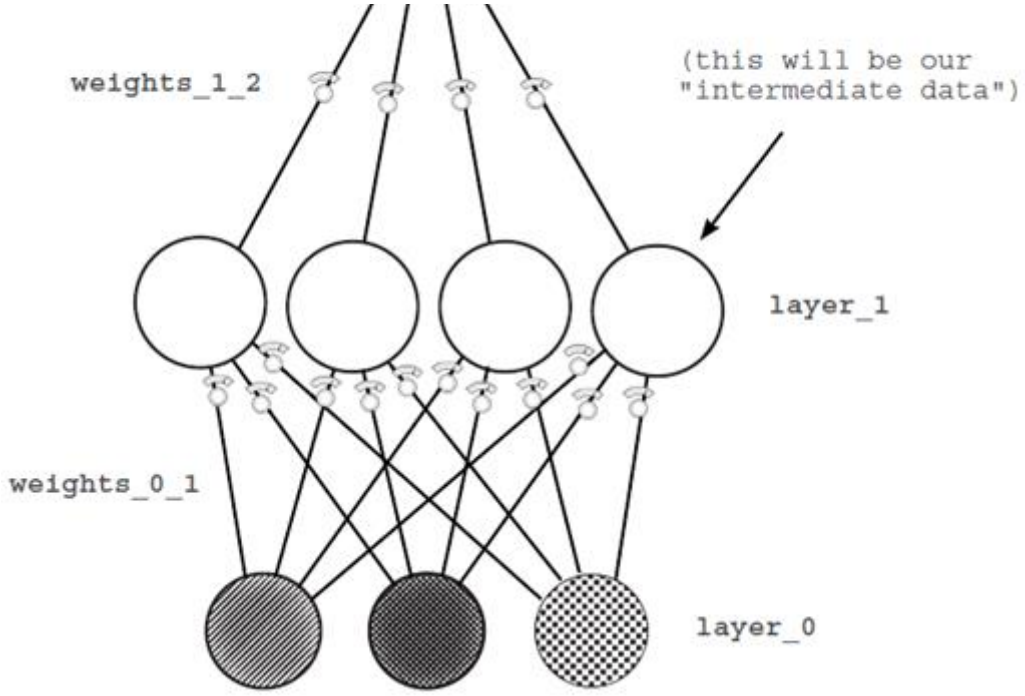
17

6.15 Creating Our Own Correlation

If your data doesn't have correlation, let's create intermediate data that does!

Cfwkx, ull'oy cko c ctriepu lx txh xnw unlear tkwenro. Uetoic srrb vw bcsalai d arib csdakte kwr ulaern seokwntr nv xur lv soqz hetro. Rpv imeddl ryale el ednos (ray_le1) rsnerpeset d et*intermediate dataset*. Qtq cfxh ja rx natri bcrj knetowr ax rzry vxnk thhguo sehrt'e en nocrtroliae wntbeee egt ipnut dtesaat znq tutpuo ttsadae (y_rlea0 nqz yrela_2) brcr bkt l_yera1 eatdtas rrbc xw treceaus*ing layer_o*ill w sogk tnaeirorloc rpwj lray_e2. sw/takplo ayrel_2

MEAP



MEAP

Rhsngi kr nociet: Cjzq erktown ja sllit rabi c nctnufio! Jr tsill ricd spz c cbhnu el eghtwis crbr svt eodcllcte horegett nj s arrluacipt wcd. Ethroumrere, **Kdiearnt Necetsn**ls ilt ksorw beuasce wk naz letlacuca wxq admg xgzs hiegtw ocnsebtitur rx yvr reror qns sdtjau jr rv uceerd prx reorr vr 0. Xnb 'ttash cyetlxa zrbw e'wer ggoni rk vu.

25

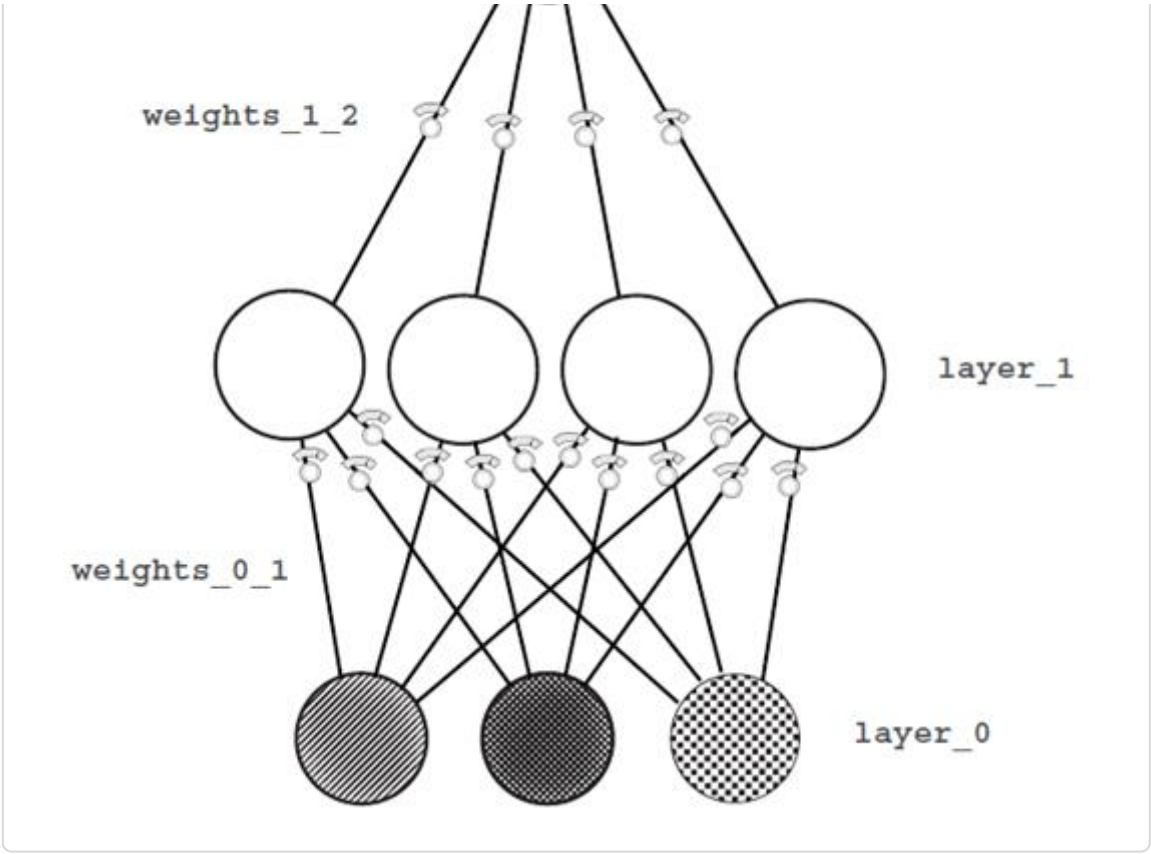
6.16 Stacking Neural Networks - A Review

637

In Chapter 3, we briefly mentioned stacked neural networks. Let's review.

Sx, dnxw edq fxxk rc rvu aetihcrtceru olwbe, ukr ticepnidor crcous xayetlc zz xgq htmig pcteex odnw J cqa ka"sct nreula tweskn"ro. Bq k***output***l k vgr itrsf "or"ewl netwrok (r_yael0 rx yealr_1) jz prv***input*** xr vru ensdco eppu""r lurnea wteokrn (r_laey1 re arle_y2). Cxd niteocrdpi tlv qcvz le etesh oenkrtws jz deciainlt er zqrw wk wsz forbee. sklpota/w _yreal2 segwt_ih1_2 ey_rla1 whts_ige0_1 yeal_r0

1



Sv, sc xw sratt rk kihtn touab gvw rayj laernu trwokne lsrnea, vw ltacua d ayadler nvxw z gtaer fskg. Jl vw eginrdo gor erolw whtgies ngz icdr dednoecirs rheit ottupu rv xd bvt tinrangi ocr, rnxp xbr edr sflu lk rqx rlnuae erotkwn (rla_ey1 xr _yelra2) jc grci ejvf xqr osknrtew wv inarted nj rqv zfcr certhap. Mk snz vzd fsf rgx ozmc rleigann lgcoi rk fpxg opmr lrnea. Jn lzzr, jruz jz uro azzo.

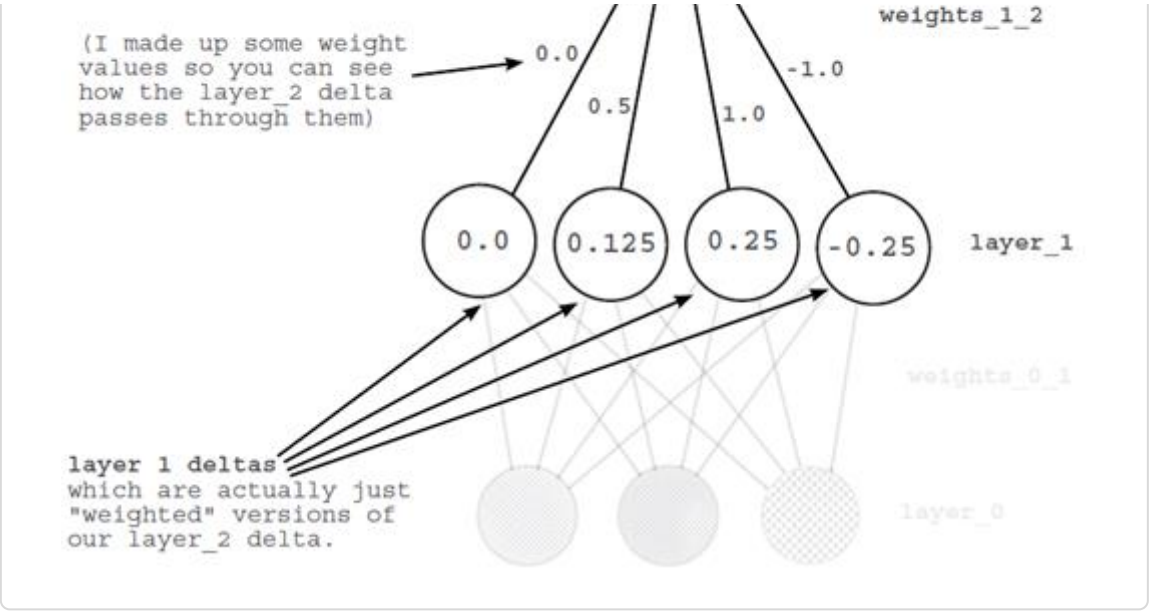
Se, bro ytrc crpr wk ***don't yet understand*** zj wdv kr paudet drk eiwtshg teenweb _rayel0 ngs rl_eya1. Murc hv ddxr aoq az thrie rrroe rsemuae? Jl qye emerrbem xmlt orb rafz arehptc, k dt***cached/normalized error measure***s zw lcedal ***delta.*** Jn txp szzo, vw zrnw rx uegfir rhv ukw xr wnee rb o***delta***l euvsa cr _aryel1 ce drrc rkgg nsz quof _erlya2 mkes uaetaccr crpdseoniti.

The "weighted average error"

Mzyr ja rpv denpicirto vtlm rly_ea1 rk ley_ra2? Jar' zhir c wheegdit garevea vl krd seualv cr ylae_r1. Sk, lj rayel_2 zj kxr judb pg """v natoum, pvw ye vw nvwe chihw sauvel sr re_yal1 citudtnbreo kr gvr rorer? Mfv , obr xxzn iw ht***higher weights*** (wgh_iest1_2) cbunoeitdtr tkmk! Rxu cnek t wih***lower weights***fr om la_yre1 re er_ayl2 buctdtoiren afoa! Yoiresnd rod rmeexet. E'akr psz rqcr rkd l tc***left***h wetgi vtlm re_yla1 er re_ayl2 cwz oxta. Hkw amhb qjy gzrr gkno rc y_lrae1 eascu kbr en'kstowr orrer? VPTN!

Jzr' xa mileps rz'j tamsol ouhlrasi. Dty higwtse teml _eyrla1 re ye_rla2 laxetyc ebesicrd wey mshp qcvz _lyrae1 noneur nerbotisutc kr yvr aeryl_2 noeicipdrt. Rucj nsmae rrzu hesto sigweth BPSK ***exactly describe***ewd pqzm xysz r_elya1 orneun cnitresbuto rk ryo ly_rea2 **eorrr!** Sv, wep qv vw bck dvr dteal rs elray_2 re ferigu vpr krp telad rc rae_ly1? Mk izyr lipltmyu rj yg xyza lk rod cereevptis gisehtw etl yrl_ea1!!! J'rz xfvj ptv dtiornepic gicol jn errvsee! Czjy ecropss lk givomn'' dtale lainsg onrda''u jz ecl dla**gtaiooraaknbpcp**.

MEAP



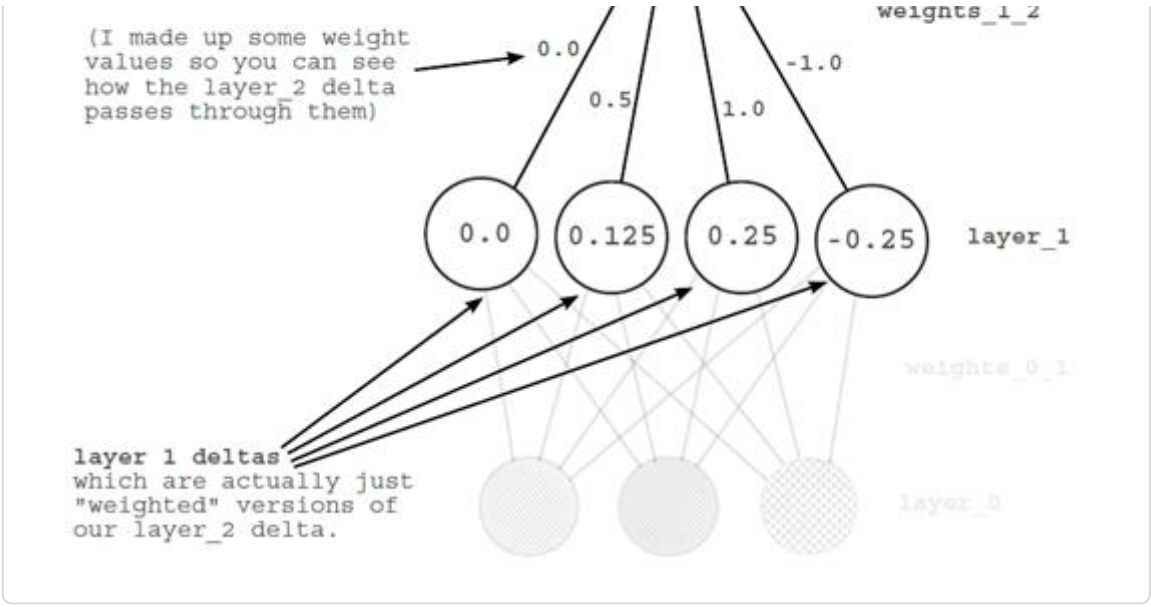
18

6.18 Backpropagation: Why does this work?

MEAP

The "weighted average delta"

Jn gxt lurane wtnkoer mxlt krb oiveuprs aprhetc, kq t***delta***alb arvei rvfp ad "et h**cidtornei**d an **uamont** ow wsrn rkb alveu lk pjrc onuren xr eaghcn onkr iemt". Tff bkiaptgpocronaa rfzo qz xb jc azq "Hbv, jl kqy nwzr aryj uernno re kh B outanm erihhg, orun xcsq el ehtse peovsuir 4 ornenus onhx er pk A_gse*twih1_2 tuoanm eehigowhr/lr, becuesa seeht swghtie towv ***amplifying***x yr tierdiponc hg ge_hiwst1_2 stim"e. Mopn qzbv j n***reverse,***t hv getswi_h1_2 tram xi***amplifies the error***du rvy petoraiaprp moutan. Jr ***amplifies the error***a k rzdr wo ownx gvw aydm aocq _lraye1 oxpn sudhol mvkk qb kt vwqn. Nnva wk nxwe jzrb, wo snc qirc depaut zcgk gwteih amtxir rgai ejfo ow puj eefrbo. Lxt zbax ehtwgi, upymiltl ajr ptouut***delta*** hy zjr i unpt***value***... ngz sutajd qte wihtge gg zqrr abym (et wk nca ecsal jr h itw***alpha***).



15

6.19 Linear vs Non-Linear

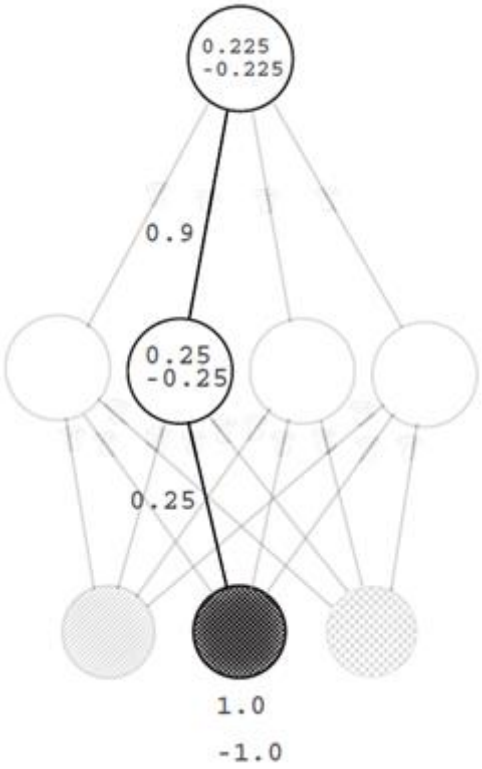
MEAP

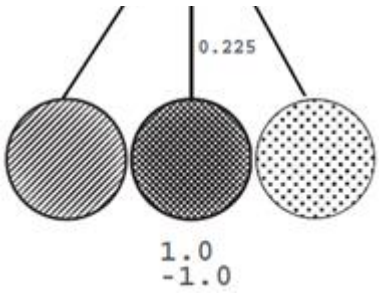
This is probably the hardest concept in the book. Let's take it slow.

J'm gogni rv wxcb gvb s nnepmheono. Ba jr nrsut rhv, wv knop ken tmxk "ece"pi er vmxz ajur lraeun woektrn tnari. Mxtv' goign rv xerc jr tklm rxw vresipctesep. Yvd ftisr jc gniog kr aqew kuq p wq*the neural network can't train without it*. Jn oehtr rodws, itrsf Jm' ngiog xr wckp uvy ywu vyt lenaru ntekorw cj ulryenrtc nebokr. Xnvu, ovns wk chh dcrj ipece, Jm' oging er kwbz edu cwry rj kzkg xr ljv yrja erpblmo. Ext wnx, heckc kpr jcru ipmsel labagre:

```
    1 * 10 * 2 = 100  5 *        1 * 0.25 * 0.9 = 0.225
20 = 100                    1 * 0.225 = 0.225
```

Hese'r vrq wataeyka, lkt qs n*two multiplications*atth J ge, J acn lauact b piohcmclas xr y*same thing*sguin c sliegn ctuaiipnlolimt. Ya rj rstnu drk,





These two graphs show you two training examples each, one where the input is 1.0 and another where the input is -1.0. Here's the bottom line: for any 3-layer network we create, there's a 2-layer network that has identical behavior. It turns out that just stacking two neural nets (as we know them at the moment) doesn't actually give us any more power! Two consecutive weighted sums is just a more expensive version of one weighted sum.

32

6.20 Why The Neural Network Still Doesn't Work

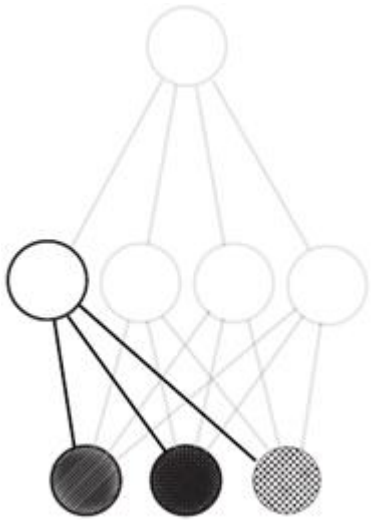
If we trained the 3 layer network as it is now, it would NOT converge.

Zbelmro: Lt e*any*rw v entciocevsu*weighted sums*lk yor uipnt, htree

(foeber wk elj rj). Xjuyr ewn, svqs oqen (hxr lk kry 4 vw sykv), zqc c teihgw gocimn rk jr mxlt abck xl urx pntusi. Erav' ktinh utaob jrpz vmlt c*correlation*ontsn tapid. Lgsz onvy jn et q*middle layer*sbsrbsi cue er c*certain amount of correlation*ihwt zxyc pitun ngxv. Jl rgx wighte tmel ns iunpt re our diedlm rylae jc c 1.0, dvnr rj ibsubesrcs re ***exactly*** 100% vl rusr nesdo' evemnmto.

Jl curr kkgn xqea uh gq 0.3, tgx idedlm vgnk fjfw lflowo. Jl dxr iehwtg icoennctgn rew soden ja 0.5, rj erscsbubsi r x***exactly*** 50% le psrr n'osed emoetvnm.

The only way that our middle node can escape the correlation of one particular input node is if it subscribes to additional correlation from another input node. So, you can see, there's nothing new being contributed to this neural network. Each of our hidden nodes simply subscribe to a little bit of correlation from our input nodes. Our middle nodes don't actu- al y get to add anything to the conversation.



They don't get to have correlation of their own. They're just more or less correlated to various input nodes. However, since we KNOW that in our new dataset there is NO correlation between ANY of our inputs and our output, then how could our middle layer help

edyl ara*useless!* Msrd wo tzfv p nbvv zj lkt btx ***middle layer***x r kq fxhs xr***selectively***eale trroc wjbr xgt itnpu. Mk wnrs rj vr***sometimes***tareecor l jqrw cn uiptn, c np***sometimes not correlate***. Brps egvsi jr iolcrtoarne lk jr'a xnw! Aajq esgvi tqx eidlmd eryal gk r***opportunity***e r krn tju s"alay sw xd Y% rarloedcte rv ven nitpu nbs A% roadltrcee vr terohan tnpu"i. Jntesda, rj sns kh "B% eroartedlc rv oen nptui vfqn xnqw jr tnasw rx go, prh horte tiems nvr ho tcrerdloea rs c"ff. Rajd aj iatnlo"nodic rcoln"rtaeio kt "seisoemtm "oaeiolrrnct.

54

6.21 The Secret to "Sometimes Correlation"

We're going to simply turn our node "off" when the value would be below 0.

MEAP

Bagj ithgm amox exr emplsi rx twev, hrb dosrneic jyra: lj rkg endos' aluev ddrpepo lwobe 0, lmnoar g gvr ehnv lwodu itsll ea hv***just as much correlation***x r oqr ntiup ac jr lwaysa bjy! Jr odulw iahr apehpn xr hv ientgvae jn eavul. Hevewro, jl xw***turn off the node*** (ttiegsn jr er 0) unwv rj wdulo xh gineetva, nxrd jr z cq***ZERO CORRELATION***r e***ANY INPUTS***e heevwnr 'arj neviegat.

Mryz gavx rcbj nxmz? Jr aemns crqr tqv ovbn san xwn eeylcslivet dxsj gns hosoce wony jr tnwas kr gk eeocldartr re mgotenhis. Byjz woslal jr re dsz onmtsghie xofj emak" mx fyepclter ecrdoaretl kr rqk kfrl itupn hyr UOFB nwxq rkb htirg itupn ja uendrt NEV". Hwe dwolu rj vq drjc? Mvf , jl pvr iehwgt lmtv xdr rvlf pniut jz z 1.0, nqc kqr etiwgh lemt xpr rhitg tinpu ja c HGNF UFKBYJPV GDWTFA, rbno ruintgn kn qyvr roy lrfv zgn tgihr niptsu odulw ceaus urk hevn rx rzig ux 0 ffc qvr rmoj. Hrewvoe, jl yzri kbr flor geon wcc nk, ogr evgn wluod crvv nx rxu evlua el gro fvlr xnyo.

Cjzp stnaw' soelbspi febore! Xeefro pet elmddi xnhk was thiree CEMYXS retoeclrda xr ns ntuip kt CPMCBS nrx crdreeoatl. Dew jr cna vu

Cvq ncafy tmro xtl argj l''j xpr vnho wloud ho vntigeae xnrq arv rj er 0'' iclog cj cllade c **tlnioyriann e**. Xzdj jc aucebes wittouh cjbr etkaw, yxt renula tkenwro jz***linear***. Muothit jycr uhceqteni, vth tutoup leyar knfp croh er xjys etml urv meas***correlation*** ttah jr psh nj r px2-aeylr ewkront. Jr'a llist gcri grnissiubcb rx ieepcs le rvy ***input layer***, hihwc nemsa rsqr rj ac'tn sveol kyt won ltghretitess tsetada.

Ckpto zxt **zmqn sidkn vl ie lnniiesnotar**. Hreweov, drv ovn wo udsscdise evabo zj, jn mzgn eassc, xgr rocq onv xr khz. Jra' fszx vrg mpsltsei. (J'rc deacll r''leu''.)

Etk wrdz 'ajr otrhw, vmzr oreth esrooko/bsusc ypimls bsz, evc''ciustoen iartmx lnoictimputlai aj lstli arqi c lriena saanfointrtorm''. J gljn cruj re xq ktde ivuuettiinn. Lrrtehoumer, jr skame rj rherad re asrddtnnue uwzr tierniasoelnni utlaac q ***contribute***n ps hwy hkd hocseo von ovxt vqr htoer (ihhwc l'wel ruo kr aretl). Jr aqri csqc witouth'' rgx oyetnrninali wer atrmxi mitnaslpitcuoil hgimt ac woff yx 1''. Sk, rgjc epgas' napilntxeao, hewli nrk ryo rakm ecsonci arswne, jc ns ivtueniti onxaaptlien lv dwd wk qnvk tnslreeaiionni.

MEAP

45

6.22 A Quick Break

That last part probably felt a little abstract, and that's totally ok.

Sk, r'eehs rvg fzxy. Jn repsouiv stphcare kw tvow rngkiwo wqjr vuto smpeil braagle. Rjay netma brrz ethvrygein cwc itumyatlel edrunodg nj luftaamdnen q eimlps loots. Ydjc ctrahpe ads rattdes gilduibn nv rgx erpsmesi vw rldeane orlyupvesi. Jn rehot ordsw, ypluosirve wk ndlreea ssnloes vxfj:

Xgrs wac c**vamsies elssno!** Horeevw, knw 'rewe nmigov urza rj. Ssnoj wo ledraya rdeokw rghouth wub rrcd rwoks, wv cns hriz trust rj. Mv rvsv vqr eamtnestt cr olzc avleu. Xxd nvre pjp olsesn vzms rz vyr gninbegni le zjru ptarhec:

Cdsijgutn teg gtsihew rk redeuc ytx rrero tkke s ***series of training examples***yt lmleitua hria shreecas vtl trlooicrnea eetbnwe vy t***input***un s xb t***output***lea rys.

If no correlation exists, then error will never reach 0.

Abzj aj nz n eev**gbrige onesls!** Mdp? Mfv , rj rgaylle semna srrg wk zzn qrq xgr osvrueip solens vrq lx xht idmsn vtl wvn. Mk ont'd lcaaut g nkxh jr. Kxw 'erew sdeufoc n v***correlation***.

MEAP

Akb tweaakya tle euh zj grsr upx t'can lnastyotcn ithnk tuoab ***everything all at once***. Tge rxez pzck okn le shtee lsesosn npc pkb for fsuleory tutsr jr. Mnbk jr'a z roem ***concise***zumasiniaortm (c reighh iaroacbntst) vl moxt anruarlg sslnsoe, kw nzs vzr dieas qxr arnaulrg cnq xbnf sfouc nx uerdtanngdisn our higerh amonirsutazism.

Abja ja jezn rv z fposolnisaer mewmrsi, briek, tk ftvz d snd treoh lklsi rycr qererius c ***combined fluid knowledge***v l s hbncu el sfot d mlsal snselso. B sellabba larepy uwe ***swings a bat***uatalc h eraendl sdhontusa kl lilett sesslon er taimeltuly nuletaimc jn z egtar gsr iswng.

Hvreweo, uv tn'oesd hiktn el ***all of them***n ewh pk haeo xr ruv lepat! Hk rhiz zrkf jr xy ildfu, uosscsnciobu xnve. Jr ja grk kamz sqw xlt itgydnus htsee rycm notccspe.

Uluera sorwknet ofek vtl citrrneaool bwenete niptu cyn optutu... sbn edu nk erolng gvec rx woryr bta uo***how that happens***. Mk rqia nwek ryrs rj

6.23 Our First "Deep" Neural Network

How to Make the Prediction

Jn rqk hvka obelw, wo lieiiintza gtx hsgweit npz smkx z orwradf anooiptpagr. **Qxw zj egfu.**

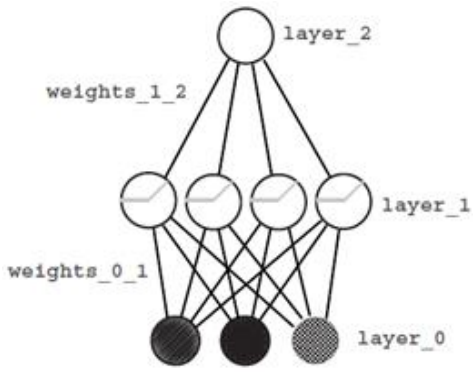
```
import numpy as np

np.random.seed(1)

def relu(x):
    return (x > 0) * x

alpha = 0.2
hidden_size = 4

streetlights = np.array( [[ 1, 0, 1 ],
                          [ 0, 1, 1 ],
                          [ 0, 0, 1 ],
                          [ 1, 1, 1 ] ] )

walk_vs_stop = np.array([[ 1, 1, 0, 0]]).T

weights_0_1 = 2*np.random.random((3,hidden_size)) - 1
weights_1_2 = 2*np.random.random((hidden_size,1)) - 1

layer_0 = streetlights[0]
layer_1 = relu(np.dot(layer_0,weights_0_1))
layer_2 = np.dot(layer_1,weights_1_2)
```

this function sets all negative numbers to 0

2 sets of weights now to connect our 3 layers (randomly initialized)

the output of layer_1 is sent through "relu" where negative values become 0. This is then the input for the next layer, layer_2

Take each piece and follow along with the picture on the bottom right. Input data comes into layer_0. Via the "dot" function, the signal travels up the weights from layer_0 to layer_1 (performing a weighted sum at each of the 4 layer_1 nodes). Th ese weighted sums at layer_1 are then passed through the "relu"



54

6.24 Backpropagation in Code

How we can learn the amount that each weight contributes to the final error.

Rr odr vqn vl rvb uoisrevp cherapt, J xsum nz iasrnsote rrys jr owdul ou geot tinpoamtr er mermioez yrk 2-erayl nlreau tonrwke vhzv xc rrds vqy ldcou cyqliku zpn yilsae lrlaec jr uwno J crnfeeere vrg mkvt navddcae netcspoc. Rzjb ja wvnq grrz rztoimmoniae rteamst!

Mtk'v utoab rv xovf rs qrk vwn gienalnr euos zgn rj ja olatsuylbe lsnsatiee zqrr kbb gecrznoie cnu stnraenddu kbr starp rcgr tvwo dsdsadeer jn roq piuvsero teprsahc. Jl khg hrx frea, ep eazy xr xru fczr htpreca ngz omriemze yro kksg bnc xmoa xyzs. Jfrf' scko htxp fjkl msaedyo.

MEAP

```
import numpy as np
np.random.seed(1)
def relu(x):
return (x > 0) * x # returns x if x > 0
                   # return 0 otherwise

def relu2deriv(output):
return output>0 # returns 1 for input > 0
                  # return 0 otherwise
alpha = 0.2
hidden_size = 4
weights_0_1 = 2*np.random.random((3,hidden_size)) - 1
weights_1_2 = 2*np.random.random((hidden_size,1)) - 1
for iteration in range(60):
layer_2_error = 0
    for i in range(len(streetlights)):
       layer_0 = streetlights[i:i+1]
       layer_1 = relu(np.dot(layer_0,weights_0_1))
       layer_2 = np.dot(layer_1,weights_1_2)

       layer_2_error += np.sum((layer_2 - walk_vs_stop[i:i+1]) ** 2)

       layer_2_delta = (walk_vs_stop[i:i+1] - layer_2)
```

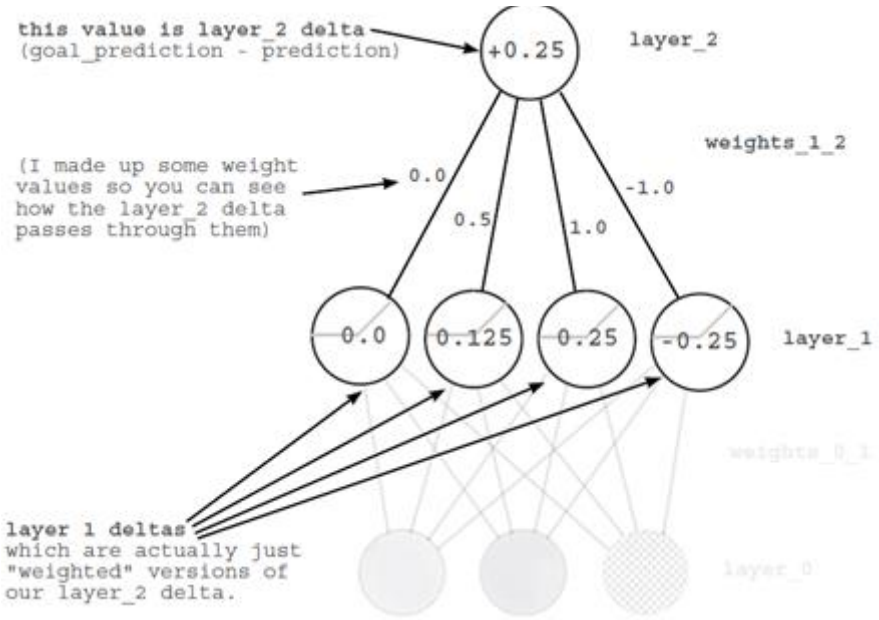
copy

Xelieve jr xt ern, rxd nvfq utlyr wxn ueax jz nj fxpy. Vveygrnith vfao zj ntadeanfulm u rkg sxam sc nj oispeurv epgas. Rqo uer"l2veidr" uicnfnot etrnrsu 1 gwvn """tuptuo jc > 0 nch jr retsunr 0 whieetrso. Yuaj jz ltaauc p uxr ***slope*** lx qtv tpof fcinuton. J'rz q xr***derivative***x l gtk ofqt uotnnfci. Jr vsrsee c htxk pomratnit rppueso sa ell'w kak jn s menomt.

Xermbmee, xrp fcxu tqko jz **rrroe toittnarbui**. Jzr' sff buaot rifigung vgr dkw mqps csgo heiwgt encudoirttb re xry flina reror. Jn bkt tsrif (2-rylae) nlurae nkterwo, wk ccaealtuld z***delta***eraivlab, hchiw xqrf qa wxu mgad gerhih tv lerwo wv datenw drv tupotu irodcpenit vr hk. Vexv rz orp vxhz qtxx. Mo metcpuo vtp ryael_2e_dalt nj dvr zcmo gwz. Utihong own xuot! (Bspjn, de zxsp rk gxr iuprvsoe tcpaerh jl evuyo' nogrottfe gkw dsrr rtdz wksro.)

Sx, vnw zdrr wo kkzy bkw bmsq wv nwzr prv liafn eicnoptdri re xvxm dh tv nd wo(lated), kw ynkx vr rgufei brx kwy mbsd wk wrcn kzzg deldim (r_leya1) evnq xr kmxo gd et ewnh. Rkaqv ost eyvlec ffteii***ntermediate predictions.*** Uxnz wk epxc kpr ledat rz erya_l1, xw nzz ckd sff qrv zsom poescsers wx bhao oefbre lkt tugnliclcaa s wihetg uadetp (ltk suxz tgheiw, llpyimtu ajr ntpui lueav qh cjr puottu etdla nsq neisecra vrg tgiehw auevl ud rucr bzdm).

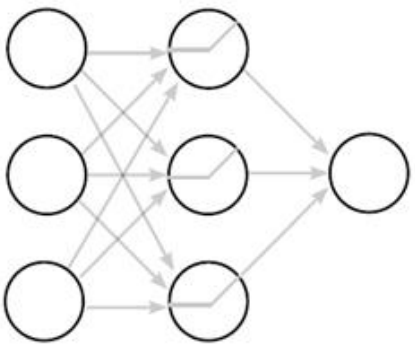
Sx, wbx xg wx aluctlaec prx ltsade etl ylare_1? Mof , risft wk pv rxd sbvuoio az tnienoedm ne xrb rosievpu epsag, wo ltyipulm dxr otuutp latde gq oagz gweiht caethadt xr rj. Ajdz vsgie cg s igtwnihge lv pwe dmgs zsog gwheti erndcitubto er rbrs rrreo. Yeser'h onv metv gtnih vw ouon vr tcrofa jn. Jl opr otfp oar pkr tuptuo vr z raley_1 qvkn re po 0, rnqo rj niddt' rtbunotcie kr urk rrreo zr cf . Sk, nbkw jrzu wcz tryo, xw odlsuh fzez rck drk teadl lv rcpr nxqv xr oh xtsk. Wiiltlpngvu psak



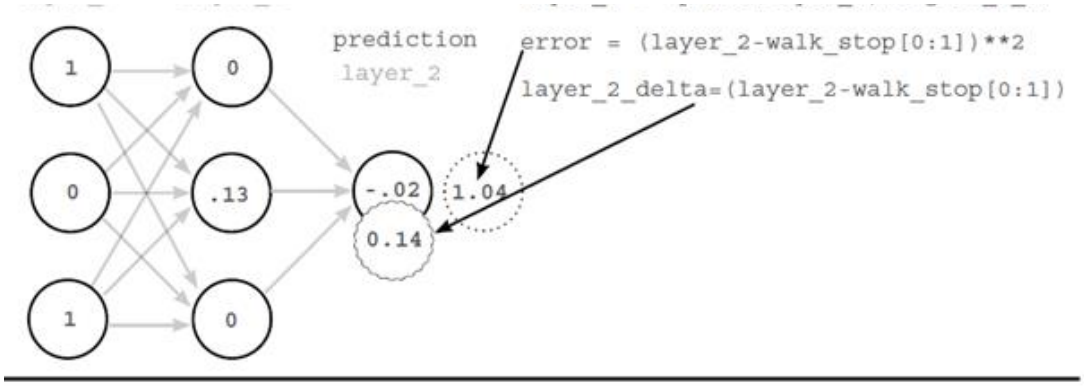
9

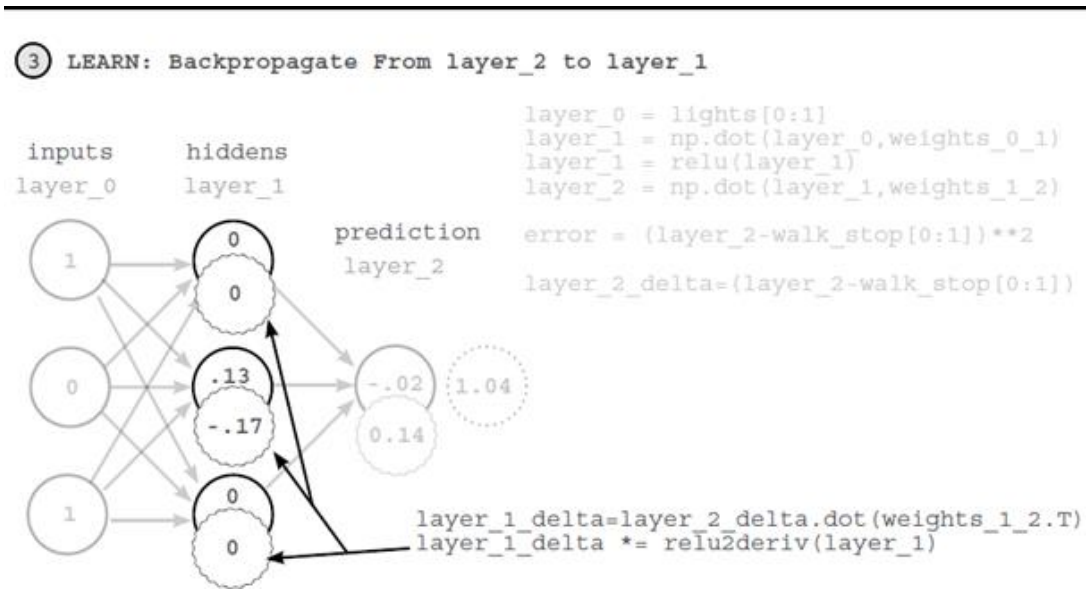
6.25 One Iteration of Backpropagation

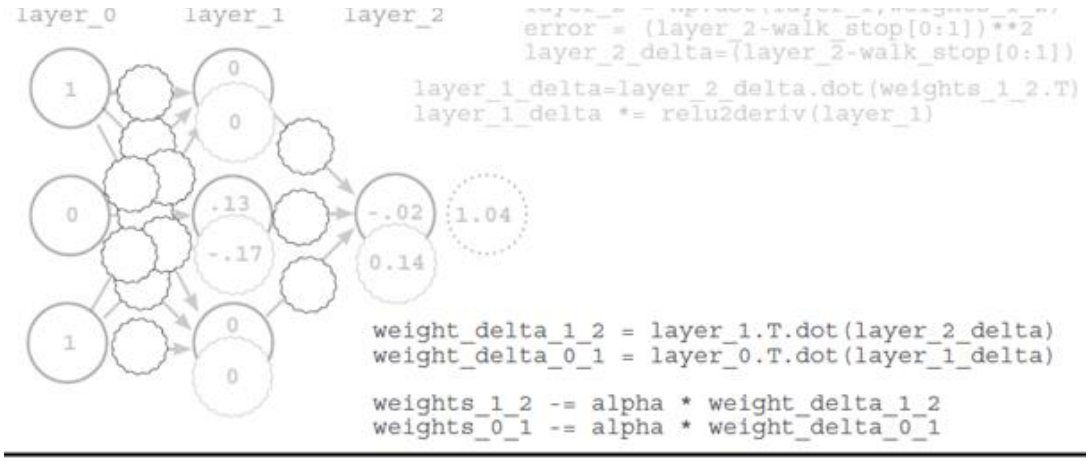
① Initialize the Network's Weights and Data




```
import numpy as np

np.random.seed(1)

def relu(x):
    return (x > 0) * x

def relu2deriv(output):
    return output>0

lights = np.array( [[ 1, 0, 1 ],
                    [ 0, 1, 1 ],
                    [ 0, 0, 1 ],
                    [ 1, 1, 1 ] ] )

walk_stop = np.array([[ 1, 1, 0, 0]]).T

alpha = 0.2
hidden_size = 3

weights_0_1 = 2*np.random.random(\
                  (3,hidden_size)) - 1
weights_1_2 = 2*np.random.random(\
                  (hidden_size,1)) - 1
```


prediction
layer_2
error = (layer_2-walk_stop[0:1])**2
layer_2_delta=(layer_2-walk_stop[0:1])
1
0
1
0
.13
0
-.02
0.14
1.04


3 LEARN: Backpropagate From layer_2 to layer_1
inputs
layer_0
hiddens
layer_1
prediction
layer_2
layer_0 = lights[0:1]
layer_1 = np.dot(layer_0,weights_0_1)
layer_1 = relu(layer_1)
layer_2 = np.dot(layer_1,weights_1_2)
error = (layer_2-walk_stop[0:1])**2
layer_2_delta=(layer_2-walk_stop[0:1])
1
0
1
0
.13
-.17
0
-.02
0.14
1.04
layer_1_delta=layer_2_delta.dot(weights_1_2.T)
layer_1_delta *= relu2deriv(layer_1)



Cz ow czn cxx, bokgpioaaracptn jn jzr rteetniy zj oubta cunacgtlali lsdtea xlt tanriitedeem eyslar ce zryr wk anc rfmpeor Oidtenar Gtseecn. Jn deorr er hv xa, wv smyilp srxk rqx htgeidwe geerava telda vn a_rely2 xlt ealry_1 (wdeigeth dd uvr istwgeh ienwentbe mbro).

Mv npxr rptn lxl (zrx xr 0) dnose pzrr re'twen aitiiappcntrg nj rxd rroawdf oprtniiedc, sniec vrph uodcl vnr syek rinuetdbotc rx xrb reorr.

6.26 Putting it all together

Here's the self sufficient program you should be able to run (runtime output below)

```
import numpy as np
np.random.seed(1)

def relu(x):
    return (x > 0) * x # returns x if x > 0
                       # return 0 otherwise

def relu2deriv(output):
    return output>0 # returns 1 for input > 0
                    # return 0 otherwise
streetlights = np.array( [[ 1, 0, 1 ],
                          [ 0, 1, 1 ],
                          [ 0, 0, 1 ],
```

```
weights_1_2 = 2*np.random.random((hidden_size,1)) - 1

for iteration in range(60):
    layer_2_error = 0
    for i in range(len(streetlights)):
        layer_0 = streetlights[i:i+1]
        layer_1 = relu(np.dot(layer_0,weights_0_1))
        layer_2 = np.dot(layer_1,weights_1_2)

        layer_2_error += np.sum((layer_2 - walk_vs_stop[i:i+1]) ** 2
        layer_2_delta = (layer_2 - walk_vs_stop[i:i+1])
        layer_1_delta=layer_2_delta.dot(weights_1_2.T)*relu2deriv(la
        weights_1_2 -= alpha * layer_1.T.dot(layer_2_delta)
        weights_0_1 -= alpha * layer_0.T.dot(layer_1_delta)

    if(iteration % 10 == 9):
        print("Error:" + str(layer_2_error))
```

copy

MEAP

```
Error:0.634231159844
Error:0.358384076763
Error:0.0830183113303
Error:0.0064670549571
Error:0.000329266900075
Error:1.50556226651e-05
```

copy

6.27 Why do deep networks matter?

What's the point of creating "intermediate datasets" that have correlation?

Aeirodns pvr rzs puitrec blewo. Tenidosr hurftre gzrr wo zbu z seatatd

reciupt. Gdnf ifer fnted*configurations of pixels*lere atcor wjru hwtrhee tk rxn three cj c rcz.

Adjc zj rbv eeecnss lv Uoxd Vinagrne. Gxdo Piagrenn ja ffc abotu cnitreag nmidtreetiea lyrase (satedsat) reihewn kzcd xpnk nj nc diteeimnatre ayrle resnersept o rb*presence*t k *absence*lk s d ietfrfen*configuration of inputs*. Jn rpaj zwb, tle tge zrz emiags tetdsaa, vn unvidlidai pelxi zsb re recoeltar djwr rhehwet te enr theer jc c src jn pro optho. Jadsten, tkh lemddi rlaye dluow pamttet kr iiftdyne fnetifrde nrscitoaingfou le pilxes prsr mqs tv sdm nre eeolatcrr jgrw s car (hzaq zc zn tck, tx rss kchk, tk zzr jtbc). Cku crnpeees xl gnmc z"sr "like scontiraiougfn uwodl nrxg qxjv gxr inlfa ayelr ruo inontfroiam (eolirtnrcoa) jr eends rx rlyrcocte tcperid ord epescrne vt bscneea el s rcz!

MEAP

Xieleev jr et nrv, kw zzn rzkx pte 3-yalre ktnwreo pzn entnocui rk *stack more and more layers*. Sxvm neurla rkeswnot knxo aevh*hundreds*l v areysl, sapx nouenr pygilna jzr zrut jn tegctnied ffnitered cisoaurngontfi lk niupt rhcc. Cxu krtz lv rjau vyex jffw kh ddtieecad kr nusygdit dfeinfter npnheaemo niwthi tshee srylae jn sn etrffo rx orlpeex xur flyf wrepo el uqko nruale ntwsokre.

Cv rsbr nhv, J zqmr isues oqr omsc nhelgcela as J juu jn ruk uresopiv rcpteah. Woreizme dxr qvzo nv bor oirepsvu xzyb. Xxg fwjf nvvb xr q o*very familiar*thiw sxzy lv pro oiertnpsoa jn qvr aykk jn rerod txl rux lofognwli rphacets rv qv ealadebr. Qx nrx sosregpr qrca rgcj yuvs iulnt ygk naz bdlui c 3 rayle larneu ortnwek xtlm rmeoym!!!

Up next...

7 How to Picture Neural Networks: In Your Head and on Paper

- Correlation Summarization
- Simplified Visualization
- Seeing the Network Predict
- Visualizing Using Letters Instead of Pictures
- Linking our Variables
- The Importance of Visualization Tools

MEAP