breast\_cancer (complex\_tanh\_sigmoid)

validations 300, epochs 200, batch size 128, step size 0.001 -> 0.0001 (100) -> 0.00001 (150), L2 decay 0

SGD momentum 0.9

|  |
| --- |
| class fc\_model(nn.Module):  def \_\_init\_\_(self):  super(fc\_model, self).\_\_init\_\_()  self.fc1 = nn.Linear(30, 128)  self.fc2 = nn.Linear(128, 64)  self.fc3 = nn.Linear(64, 32)  self.fc4 = nn.Linear(32, 1)  def forward(self, inputs):  fc1\_out = F.tanh(self.fc1(inputs))  fc2\_out = F.tanh(self.fc2(fc1\_out))  fc3\_out = F.tanh(self.fc3(fc2\_out))  fc4\_out = self.fc4(fc3\_out)  return fc4\_out  fc4\_out -> sigmoid -> cross entropy (BCE) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.00111+-0.00012 | 0.00122+-0.00015 | 0.96583+-0.00840 | 0.96047+-0.01414 |
| mixup alpha8.0 | 0.00225+-0.00016 | 0.00232+-0.00017 | 0.94617+-0.00971 | 0.94499+-0.01472 |
| mixup alpha4.0 | 0.00218+-0.00017 | 0.00224+-0.00016 | 0.94656+-0.01021 | 0.94507+-0.01412 |
| mixup alpha1.0 | 0.00184+-0.00014 | 0.00194+-0.00015 | 0.94998+-0.00957 | 0.94582+-0.01453 |
| mixup alpha0.5 | 0.00164+-0.00014 | 0.00173+-0.00016 | 0.95409+-0.00936 | 0.94991+-0.01541 |
| mixup alpha0.2 | 0.00140+-0.00013 | 0.00148+-0.00015 | 0.95847+-0.00904 | 0.95355+-0.01534 |
| mixup sc alpha8.0 | 0.00111+-0.00011 | 0.00122+-0.00019 | 0.95429+-0.00915 | 0.95035+-0.01551 |
| mixup sc alpha4.0 | 0.00111+-0.00011 | 0.00120+-0.00018 | 0.95469+-0.00959 | 0.95232+-0.01359 |
| mixup sc alpha1.0 | 0.00111+-0.00011 | 0.00120+-0.00017 | 0.95749+-0.00908 | 0.95378+-0.01455 |
| mixup sc alpha0.5 | 0.00110+-0.00011 | 0.00121+-0.00017 | 0.96107+-0.00922 | 0.95531+-0.01533 |
| mixup sc alpha0.2 | 0.00110+-0.00011 | 0.00121+-0.00017 | 0.96320+-0.00866 | 0.95709+-0.014353 |
| mixup nb0.5 alpha8.0 | 0.00121+-0.00012 | 0.00131+-0.00015 | 0.95744+-0.00929 | 0.95276+-0.01433 |
| mixup nb0.5 alpha4.0 | 0.00120+-0.00012 | 0.00132+-0.00016 | 0.95695+-0.00879 | 0.95095+-0.01537 |
| mixup nb0.5 alpha1.0 | 0.00121+-0.00012 | 0.00129+-0.00016 | 0.95686+-0.00913 | 0.95389+-0.01477 |
| mixup nb0.5 alpha0.5 | 0.00119+-0.00012 | 0.00127+-0.00016 | 0.95761+-0.00920 | 0.95471+-0.01433 |
| mixup nb0.5 alpha0.2 | 0.00117+-0.00012 | 0.00126+-0.00017 | 0.95870+-0.00888 | 0.95594+-0.01588 |
| mixup nb0.75 alpha8.0 | 0.00120+-0.00013 | 0.00129+-0.00016 | 0.95814+-0.00920 | 0.95430+-0.01486 |
| mixup nb0.75 alpha4.0 | 0.00120+-0.00011 | 0.00131+-0.00015 | 0.95770+-0.00902 | 0.95281+-0.01461 |
| mixup nb0.75 alpha1.0 | 0.00118+-0.00012 | 0.00127+-0.00016 | 0.95865+-0.00822 | 0.95456+-0.01502 |
| mixup nb0.75 alpha0.5 | 0.00117+-0.00012 | 0.00127+-0.00016 | 0.95905+-0.00995 | 0.95442+-0.01479 |
| mixup nb0.75 alpha0.2 | 0.00117+-0.00012 | 0.00127+-0.00015 | 0.95873+-0.00893 | 0.95433+-0.014879 |
| mixup nb1.0 alpha8.0 | 0.00118+-0.00012 | 0.00127+-0.00016 | 0.95975+-0.00925 | 0.95504+-0.01572 |
| mixup nb1.0 alpha4.0 | 0.00117+-0.00012 | 0.00129+-0.00016 | 0.95971+-0.00871 | 0.95380+-0.01507 |
| mixup nb1.0 alpha1.0 | 0.00118+-0.00011 | 0.00127+-0.00016 | 0.95950+-0.00999 | 0.95453+-0.01577 |
| mixup nb1.0 alpha0.5 | 0.00117+-0.00012 | 0.00126+-0.00016 | 0.95959+-0.00866 | 0.95518+-0.01463 |
| mixup nb1.0 alpha0.2 | 0.00116+-0.00012 | 0.00127+-0.00016 | 0.95954+-0.00902 | 0.95425+-0.01426 |
| gauss VRM std0.25 | 0.00113+-0.00013 | 0.00125+-0.00015 | 0.96393+-0.00873 | 0.95766+-0.01398 |
| gauss VRM std0.5 | 0.00117+-0.00012 | 0.001278+-0.00017 | 0.96053+-0.00855 | 0.95621+-0.01457 |
| gauss VRM std0.75 | 0.00123+-0.00014 | 0.00135+-0.00016 | 0.95725+-0.00929 | 0.95099+-0.01473 |

breast\_cancer (simple\_relu\_sigmoid)

validations 100, epochs 35, batch size 16, step size 0.001 -> 0.0001 (17) -> 0.00001 (26), L2 decay 0

SGD momentum 0.9

|  |
| --- |
| class fc\_model(nn.Module):  def \_\_init\_\_(self):  super(fc\_model, self).\_\_init\_\_()  self.fc1 = nn.Linear(30, 128)  self.fc2 = nn.Linear(128, 128)  self.fc3 = nn.Linear(128, 1)  def forward(self, inputs):  fc1\_out = F.relu(self.fc1(inputs))  fc2\_out = F.relu(self.fc2(fc1\_out))  fc3\_out = self.fc3(fc2\_out)  return fc3\_out  fc3\_out-> sigmoid -> cross entropy (BCE) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.01001+-0.00103 | 0.01098+-0.00102 | 0.95560+-0.00829 | 0.94877+-0.01120 |
| mixup alpha8.0 | 0.01793+-0.00125 | 0.01907+-0.00143 | 0.93701+-0.00971 | 0.93066+-0.01558 |
| mixup alpha4.0 | 0.01753+-0.00143 | 0.01833+-0.00172 | 0.93950+-0.00954 | 0.93667+-0.01652 |
| mixup alpha1.0 | 0.01515+-0.00113 | 0.01590+-0.00124 | 0.94352+-0.00884 | 0.94175+-0.01317 |
| mixup alpha0.5 | 0.01340+-0.00104 | 0.01440+-0.00133 | 0.94889+-0.00923 | 0.94535+-0.01309 |
| mixup alpha0.2 | 0.01179+-0.00116 | 0.012962+-0.00121 | 0.95238+-0.01004 | 0.94561+-0.01436 |
| mixup sc alpha8.0 | 0.00989+-0.00090 | 0.01047+-0.00143 | 0.94809+-0.00940 | 0.94526+-0.01258 |
| mixup sc alpha4.0 | 0.00964+-0.00089 | 0.01051+-0.00132 | 0.95018+-0.00922 | 0.94746+-0.01427 |
| mixup sc alpha1.0 | 0.00955+-0.00080 | 0.01060+-0.00138 | 0.95223+-0.00892 | 0.94675+-0.01496 |
| mixup sc alpha0.5 | 0.00971+-0.00076 | 0.01031+-0.00120 | 0.95390+-0.00821 | 0.95066+-0.01353 |
| mixup sc alpha0.2 | 0.00966+-0.00081 | 0.010567+-0.00118 | 0.95490+-0.00757 | 0.95031+-0.01212 |
| mixup nb0.5 alpha8.0 | 0.01205+-0.00081 | 0.01269+-0.00141 | 0.94361+-0.00881 | 0.94386+-0.01537 |
| mixup nb0.5 alpha4.0 | 0.01193+-0.00113 | 0.01265+-0.00149 | 0.94525+-0.0105 | 0.94289+-0.01495 |
| mixup nb0.5 alpha1.0 | 0.01144+-0.00095 | 0.01240+-0.00126 | 0.94848+-0.01006 | 0.94140+-0.01451 |
| mixup nb0.5 alpha0.5 | 0.01097+-0.00093 | 0.01208+-0.00114 | 0.95079+-0.00854 | 0.94335+-0.01305 |
| mixup nb0.5 alpha0.2 | 0.01087+-0.00109 | 0.01180+-0.00158 | 0.94903+-0.00928 | 0.94605+-0.01538 |
| mixup nb0.75 alpha8.0 | 0.01133+-0.00085 | 0.01230+-0.00130 | 0.09468+-0.00910 | 0.94342+-0.01570 |
| mixup nb0.75 alpha4.0 | 0.01131+-0.00084 | 0.01209+-0.00127 | 0.94628+-0.00873 | 0.94474+-0.01470 |
| mixup nb0.75 alpha1.0 | 0.01108+-0.00085 | 0.01216+-0.00134 | 0.94754+-0.00750 | 0.94316+-0.01142 |
| mixup nb0.75 alpha0.5 | 0.01069+-0.00098 | 0.01174+-0.00139 | 0.95032+-0.00879 | 0.94294+-0.01517 |
| mixup nb0.75 alpha0.2 | 0.01056+-0.00100 | 0.01145+-0.00136 | 0.95088+-0.00875 | 0.94658+-0.01372 |
| mixup nb1.0 alpha8.0 | 0.01109+-0.00095 | 0.01210+-0.00139 | 0.94771+-0.00881 | 0.94485+-0.01347 |
| mixup nb1.0 alpha4.0 | 0.01109+-0.00093 | 0.01190+-0.00134 | 0.94646+-0.00927 | 0.94581+-0.01505 |
| mixup nb1.0 alpha1.0 | 0.01066+-0.00091 | 0.01189+-0.00135 | 0.94991+-0.00925 | 0.94316+-0.01390 |
| mixup nb1.0 alpha0.5 | 0.01070+-0.00091 | 0.01151+-0.00115 | 0.94961+-0.00808 | 0.94662+-0.01229 |
| mixup nb1.0 alpha0.2 | 0.01057+-0.00093 | 0.01148+-0.00131 | 0.94985+-0.00957 | 0.94702+-0.01394 |
| gauss VRM std0.25 | 0.00989+-0.00094 | 0.01107+-0.00145 | 0.95639+-0.00895 | 0.95004+-0.01504 |
| gauss VRM std0.5 | 0.01023+-0.00089 | 0.01121+-0.00115 | 0.95437+-0.00821 | 0.95035+-0.01206 |
| gauss VRM std0.75 | 0.01088+-0.00097 | 0.01192+-0.00138 | 0.95167+-0.00912 | 0.94912+-0.01538 |
| gauss VRM std1.0 | 0.01144+-0.00086 | 0.012356+-0.00123 | 0.95282+-0.00739 | 0.95004+-0.01401 |
| gauss VRM std4.0 | 0.02653+-0.00092 | 0.02745+-0.00096 | 0.94792+-0.0107 | 0.94351+-0.01310 |

breast\_cancer (simple\_relu\_softmax)

validations 100, epochs 35, batch size 16, step size 0.001 -> 0.0001 (17) -> 0.00001 (26), L2 decay 0

SGD momentum 0.9

|  |
| --- |
| class fc\_model(nn.Module):  def \_\_init\_\_(self):  super(fc\_model, self).\_\_init\_\_()  self.fc1 = nn.Linear(30, 128)  self.fc2 = nn.Linear(128, 128)  self.fc3 = nn.Linear(128, 2)  def forward(self, inputs):  fc1\_out = F.relu(self.fc1(inputs))  fc2\_out = F.relu(self.fc2(fc1\_out))  fc3\_out = self.fc3(fc2\_out)  return fc3\_out  fc3\_out-> softmax -> cross entropy (CE) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.00738+-0.00071 | 0.00824+-0.00117 | 0.96806+-0.00699 | 0.95982+-0.01425 |
| mixup alpha8.0 | 0.01396+-0.00083 | 0.01507+-0.00119 | 0.94891+-0.00794 | 0.94175+-0.01440 |
| mixup alpha4.0 | 0.01353+-0.00103 | 0.01445+-0.00123 | 0.94780+-0.01058 | 0.94307+-0.01477 |
| mixup alpha1.0 | 0.01168+-0.00087 | 0.01256+-0.00115 | 0.95367+-0.00854 | 0.94864+-0.01344 |
| mixup alpha0.5 | 0.01054+-0.00073 | 0.01168+-0.00113 | 0.95625+-0.00718 | 0.95114+-0.01329 |
| mixup alpha0.2 | 0.00913+-0.00078 | 0.01006+-0.00103 | 0.95982+-0.00774 | 0.95487+-0.01318 |
| mixup sc alpha8.0 | 0.00769+-0.00075 | 0.00837+-0.00153 | 0.95592+-0.00784 | 0.95061+-0.01337 |
| mixup sc alpha4.0 | 0.00753+-0.00078 | 0.00853+-0.00177 | 0.95757+-0.00874 | 0.95184+-0.01614 |
| mixup sc alpha1.0 | 0.00747+-0.00079 | 0.00821+-0.00137 | 0.96023+-0.00969 | 0.95526+-0.01416 |
| mixup sc alpha0.5 | 0.00735+-0.00074 | 0.00821+-0.00143 | 0.96164+-0.00725 | 0.95561+-0.01526 |
| mixup sc alpha0.2 | 0.00741+-0.00069 | 0.00826+-0.00133 | 0.96469+-0.00776 | 0.95829+-0.01491 |
| gauss VRM std0.25 | 0.00739+-0.00080 | 0.00862+-0.00134 | 0.94566+-0.00841 | 0.95877+-0.01304 |
| gauss VRM std0.5 | 0.00774+-0.00071 | 0.00851+-0.00101 | 0.96501+-0.00719 | 0.95877+-0.01209 |
| gauss VRM std0.75 | 0.00823+-0.00085 | 0.00902+-0.00130 | 0.96346+-0.00881 | 0.95807+-0.01440 |
| gauss VRM std1.0 | 0.00880+-0.00077 | 0.00995+-0.00121 | 0.96129+-0.00737 | 0.95425+-0.01335 |
| gauss VRM std4.0 | 0.02547+-0.00092 | 0.02623+-0.00108 | 0.94941+-0.00897 | 0.94715+-0.01428 |

MNIST

validations 1, epochs 200, batch size 128, step size 0.01 -> 0.001 (100) -> 0.0001 (150), L2 decay 0

SGD momentum 0.9

ResNet18 No Pretrain; softmax -> cross entropy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.00000 | 0.00022 | 1.00000 | 0.99350 |
| mixup alpha8.0 | 0.00005 | 0.00016 | 0.99940 | 0.99520 |
| mixup alpha4.0 | 0.00004 | 0.00016 | 0.99960 | 0.99460 |
| mixup alpha1.0 | 0.00006 | 0.00018 | 0.99995 | 0.99540 |
| mixup alpha0.5 | 0.00003 | 0.00019 | 0.99997 | 0.99540 |
| mixup alpha0.2 | 0.00002 | 0.00016 | 1.00000 | 0.99550 |
| mixup sc alpha8.0 | 0.00001 | 0.00017 | 1.00000 | 0.99460 |
| mixup sc alpha4.0 | 0.00007 | 0.00019 | 1.00000 | 0.99460 |
| mixup sc alpha1.0 | 0.00002 | 0.00019 | 1.00000 | 0.99460 |
| mixup sc alpha0.5 | 0.00002 | 0.00019 | 1.00000 | 0.99440 |
| mixup sc alpha0.2 | 0.00002 | 0.00018 | 1.00000 | 0.99550 |
| mixup nb0.5 alpha8.0 | 0.00004 | 0.00018 | 0.99972 | 0.99510 |
| mixup nb0.5 alpha4.0 | 0.00003 | 0.00019 | 0.99982 | 0.99370 |
| mixup nb0.5 alpha1.0 | 0.00004 | 0.00020 | 1.00000 | 0.99340 |
| mixup nb0.5 alpha0.5 | 0.00004 | 0.00020 | 1.00000 | 0.99470 |
| mixup nb0.5 alpha0.2 | 0.00002 | 0.00016 | 1.00000 | 0.99470 |
| mixup nb0.75 alpha8.0 | 0.00004 | 0.00018 | 0.99973 | 0.99400 |
| mixup nb0.75 alpha4.0 | 0.00003 | 0.00018 | 0.99985 | 0.99420 |
| mixup nb0.75 alpha1.0 | 0.00002 | 0.00018 | 0.99998 | 0.99390 |
| mixup nb0.75 alpha0.5 | 0.00002 | 0.00020 | 1.00000 | 0.99380 |
| mixup nb0.75 alpha0.2 | 0.00001 | 0.00018 | 1.00000 | 0.99480 |
| mixup nb1.0 alpha8.0 | 0.00004 | 0.00018 | 0.99988 | 0.99440 |
| mixup nb1.0 alpha4.0 | 0.00003 | 0.00018 | 0.99988 | 0.99420 |
| mixup nb1.0 alpha1.0 | 0.00002 | 0.00019 | 1.00000 | 0.99420 |
| mixup nb1.0 alpha0.5 | 0.00001 | 0.00018 | 1.00000 | 0.99480 |
| mixup nb1.0 alpha0.2 | 0.00001 | 0.00015 | 1.00000 | 0.99520 |
| gauss VRM std0.25 | 0.00002 | 0.00020 | 1.00000 | 0.99520 |
| gauss VRM std0.5 | 0.00001 | 0.00014 | 1.00000 | 0.99620 |
| gauss VRM std0.75 | 0.00002 | 0.00014 | 1.00000 | 0.99530 |

cifar10

validations 1, epochs 200, batch size 128, step size 0.1 -> 0.01 (100) -> 0.001 (150), L2 decay 0

SGD momentum 0.9

ResNet18 No Pretrain; softmax -> cross entropy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.00000 | 0.00253 | 1.00000 | 0.93450 |
| mixup alpha8.0 | 0.00200 | 0.00277 | 0.98038 | 0.94480 |
| mixup alpha4.0 | 0.00173 | 0.00264 | 0.99104 | 0.95040 |
| mixup alpha1.0 | 0.00081 | 0.00189 | 0.99966 | 0.95220 |
| mixup alpha0.5 | 0.00042 | 0.00163 | 0.99998 | 0.95170 |
| mixup alpha0.2 | 0.00002 | 0.00164 | 1.00000 | 0.94780 |
| mixup sc alpha8.0 | 0.00020 | 0.00179 | 0.92214 | 0.94190 |
| mixup sc alpha4.0 | 0.00009 | 0.00190 | 0.99694 | 0.94130 |
| mixup sc alpha1.0 | 0.00001 | 0.00199 | 0.99982 | 0.94400 |
| mixup sc alpha0.5 | 0.00001 | 0.00200 | 0.99998 | 0.94290 |
| mixup sc alpha0.2 | 0.00000 | 0.00229 | 1.00000 | 0.94170 |
| mixup nb0.5 alpha8.0 | 0.00250 | 0.00302 | 0.97580 | 0.93810 |
| mixup nb0.5 alpha4.0 | 0.00261 | 0.00328 | 0.96394 | 0.92150 |
| mixup nb0.5 alpha1.0 | 0.00110 | 0.00211 | 0.99836 | 0.94760 |
| mixup nb0.5 alpha0.5 | 0.00032 | 0.00162 | 0.99928 | 0.94630 |
| mixup nb0.5 alpha0.2 | 0.00010 | 0.00177 | 0.99978 | 0.94100 |
| mixup nb0.75 alpha8.0 | 0.00249 | 0.00310 | 0.97640 | 0.93510 |
| mixup nb0.75 alpha4.0 | 0.00256 | 0.00323 | 0.96512 | 0.91710 |
| mixup nb0.75 alpha1.0 | 0.00087 | 0.00193 | 0.99752 | 0.94900 |
| mixup nb0.75 alpha0.5 | 0.00058 | 0.00186 | 0.99882 | 0.94280 |
| mixup nb0.75 alpha0.2 | 0.00010 | 0.00179 | 0.99972 | 0.93820 |
| mixup nb1.0 alpha8.0 | 0.00249 | 0.00307 | 0.97356 | 0.93190 |
| mixup nb1.0 alpha4.0 | 0.00217 | 0.00287 | 0.97776 | 0.93300 |
| mixup nb1.0 alpha1.0 | 0.00090 | 0.00194 | 0.99784 | 0.94830 |
| mixup nb1.0 alpha0.5 | 0.00103 | 0.00234 | 0.97828 | 0.91330 |
| mixup nb1.0 alpha0.2 | 0.00006 | 0.00171 | 0.99974 | 0.93780 |
| gauss VRM std0.25 | 0.00001 | 0.00292 | 0.99980 | 0.92120 |
| gauss VRM std0.5 | 0.00010 | 0.00382 | 0.99568 | 0.87480 |
| gauss VRM std0.75 | 0.00133 | 0.00549 | 0.94416 | 0.81020 |

spambase (complex\_tanh\_sigmoid)

validations 300, epochs 300, batch size 128, step size 0.001 -> 0.0001 (150) -> 0.00001 (225), L2 decay 0

SGD momentum 0.9

|  |
| --- |
| class fc\_model(nn.Module):  def \_\_init\_\_(self):  super(fc\_model, self).\_\_init\_\_()  self.fc1 = nn.Linear(57, 256)  self.fc2 = nn.Linear(256, 128)  self.fc3 = nn.Linear(128, 64)  self.fc4 = nn.Linear(64, 32)  self.fc5 = nn.Linear(32, 1)  def forward(self, inputs):  fc1\_out = F.tanh(self.fc1(inputs))  fc2\_out = F.tanh(self.fc2(fc1\_out))  fc3\_out = F.tanh(self.fc3(fc2\_out))  fc4\_out = F.tanh(self.fc4(fc3\_out))  fc5\_out = self.fc5(fc4\_out)  return fc5\_out  fc5\_out -> sigmoid -> cross entropy (BCE) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.00131+-0.00005 | 0.00160+-0.00010 | 0.94213+-0.00319 | 0.93217+-0.00479 |
| mixup alpha8.0 | 0.00222+-0.00004 | 0.00234+-0.00005 | 0.91925+-0.00349 | 0.91553+-0.00595 |
| mixup alpha4.0 | 0.00216+-0.00004 | 0.00228+-0.00005 | 0.92109+-0.00360 | 0.91710+-0.00576 |
| mixup alpha1.0 | 0.00193+-0.00005 | 0.00206+-0.00006 | 0.92723+-0.00328 | 0.92301+-0.00529 |
| mixup alpha0.5 | 0.00177+-0.00005 | 0.00192+-0.00006 | 0.93159+-0.00339 | 0.92630+-0.00517 |
| mixup alpha0.2 | 0.00157+-0.00005 | 0.00174+-0.00007 | 0.93596+-0.00321 | 0.92963+-0.00454 |
| mixup sc alpha8.0 | 0.00147+-0.00006 | 0.00177+-0.00014 | 0.93376+-0.00358 | 0.92741+-0.00552 |
| mixup sc alpha4.0 | 0.00145+-0.00006 | 0.00174+-0.00014 | 0.93408+-0.00341 | 0.92814+-0.00530 |
| mixup sc alpha1.0 | 0.00137+-0.00005 | 0.00169+-0.00012 | 0.93707+-0.00329 | 0.92891+-0.00515 |
| mixup sc alpha0.5 | 0.00134+-0.00006 | 0.00165+-0.00012 | 0.93887+-0.00321 | 0.93008+-0.00471 |
| mixup sc alpha0.2 | 0.00132+-0.00005 | 0.00162+-0.00010 | 0.94060+-0.00306 | 0.93101+-0.00458 |
| mixup nb0.5 alpha8.0 | 0.00157+-0.00005 | 0.00173+-0.00007 | 0.93498+-0.00345 | 0.92992+-0.00487 |
| mixup nb0.5 alpha4.0 | 0.00155+-0.00005 | 0.00173+-0.00008 | 0.93536+-0.00318 | 0.92997+-0.00487 |
| mixup nb0.5 alpha1.0 | 0.00150+-0.00005 | 0.00168+-0.00008 | 0.93632+-0.00319 | 0.93088+-0.00508 |
| mixup nb0.5 alpha0.5 | 0.00146+-0.00006 | 0.00166+-0.00009 | 0.93760+-0.00327 | 0.93144+-0.00473 |
| mixup nb0.5 alpha0.2 | 0.00142+-0.00005 | 0.00164+-0.00009 | 0.93850+-0.00293 | 0.93214+-0.00464 |
| mixup nb0.75 alpha8.0 | 0.00154+-0.00005 | 0.00171+-0.00008 | 0.93501+-0.00322 | 0.93012+-0.00456 |
| mixup nb0.75 alpha4.0 | 0.00153+-0.00005 | 0.00170+-0.00007 | 0.93582+-0.00317 | 0.93033+-0.00472 |
| mixup nb0.75 alpha1.0 | 0.00147+-0.00005 | 0.00168+-0.00008 | 0.93733+-0.00311 | 0.93032+-0.00450 |
| mixup nb0.75 alpha0.5 | 0.00145+-0.00005 | 0.00165+-0.00008 | 0.93760+-0.00287 | 0.93127+-0.00434 |
| mixup nb0.75 alpha0.2 | 0.00141+-0.00006 | 0.00165+-0.00010 | 0.93905+-0.00338 | 0.93171+-0.00486 |
| mixup nb1.0 alpha8.0 | 0.00152+-0.00005 | 0.00170+-0.00008 | 0.93569+-0.00316 | 0.93034+-0.00478 |
| mixup nb1.0 alpha4.0 | 0.00151+-0.00005 | 0.00169+-0.00008 | 0.93603+-0.00334 | 0.93039+-0.00505 |
| mixup nb1.0 alpha1.0 | 0.00146+-0.00005 | 0.00167+-0.00008 | 0.93701+-0.00328 | 0.93082+-0.00478 |
| mixup nb1.0 alpha0.5 | 0.00144+-0.00006 | 0.00165+-0.00009 | 0.93789+-0.00334 | 0.93136+-0.00495 |
| mixup nb1.0 alpha0.2 | 0.00141+-0.00006 | 0.00164+-0.00010 | 0.93908+-0.00333 | 0.93139+-0.00489 |
| gauss VRM std0.25 | 0.00145+-0.00005 | 0.00166+-0.00008 | 0.93684+-0.00315 | 0.93038+-0.00466 |
| gauss VRM std0.5 | 0.00168+-0.00005 | 0.00184+-0.00007 | 0.92627+-0.00301 | 0.92237+-0.00498 |
| gauss VRM std0.75 | 0.00189+-0.00005 | 0.00203+-0.00007 | 0.91814+-0.00371 | 0.91474+-0.00546 |

spambase (simple\_relu\_sigmoid)

validations 100, epochs 35, batch size 16, step size 0.001 -> 0.0001 (17) -> 0.00001 (26), L2 decay 0

SGD momentum 0.9

|  |
| --- |
| class fc\_model(nn.Module):  def \_\_init\_\_(self):  super(fc\_model, self).\_\_init\_\_()  self.fc1 = nn.Linear(57, 128)  self.fc2 = nn.Linear(128, 128)  self.fc3 = nn.Linear(128, 1)  def forward(self, inputs):  fc1\_out = F.relu(self.fc1(inputs))  fc2\_out = F.relu(self.fc2(fc1\_out))  fc3\_out = self.fc3(fc2\_out)  return fc3\_out  fc3\_out -> sigmoid -> cross entropy (BCE) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.00992+-0.00035 | 0.01220+-0.00086 | 0.94494+-0.00287 | 0.93470+-0.00485 |
| mixup alpha8.0 | 0.01684+-0.00041 | 0.01764+-0.00054 | 0.91856+-0.00446 | 0.91531+-0.00677 |
| mixup alpha4.0 | 0.01638+-0.00038 | 0.01742+-0.00051 | 0.92144+-0.00406 | 0.91505+-0.00556 |
| mixup alpha1.0 | 0.01475+-0.00034 | 0.01596+-0.00058 | 0.92801+-0.00355 | 0.92237+-0.00533 |
| mixup alpha0.5 | 0.01353+-0.00042 | 0.01477+-0.00062 | 0.93268+-0.00349 | 0.92582+-0.00583 |
| mixup alpha0.2 | 0.01208+-0.00040 | 0.01342+-0.00062 | 0.93794+-0.00318 | 0.93046+-0.00529 |
| mixup sc alpha8.0 | 0.01133+-0.00045 | 0.01372+-0.00124 | 0.93565+-0.00344 | 0.92858+-0.00490 |
| mixup sc alpha4.0 | 0.011110+-0.00040 | 0.01362+-0.00102 | 0.93629+-0.00321 | 0.92836+-0.00497 |
| mixup sc alpha1.0 | 0.01048+-0.00042 | 0.01299+-0.00105 | 0.93953+-0.00315 | 0.93087+-0.00486 |
| mixup sc alpha0.5 | 0.01032+-0.00037 | 0.01281+-0.00120 | 0.94089+-0.00293 | 0.93325+-0.00516 |
| mixup sc alpha0.2 | 0.01003+-0.00040 | 0.01250+-0.00101 | 0.94323+-0.00316 | 0.93310+-0.00464 |
| gauss VRM std0.25 | 0.01077+-0.00042 | 0.01283+-0.00077 | 0.94166+-0.00338 | 0.93320+-0.00469 |
| gauss VRM std0.5 | 0.01278+-0.00045 | 0.01445+-0.00077 | 0.93414+-0.00320 | 0.92747+-0.00441 |
| gauss VRM std0.75 | 0.01523+-0.00042 | 0.01652+-0.00054 | 0.92809+-0.00337 | 0.92287+-0.00541 |
| gauss VRM std1.0 | 0.01779+-0.00048 | 0.01887+-0.00050 | 0.92340+-0.00373 | 0.91900+-0.00590 |
| gauss VRM std4.0 | 0.03643+-0.00037 | 0.03669+-0.00034 | 0.88679+-0.02004 | 0.88370+-0.02084 |

spambase (simple\_relu\_softmax)

validations 100, epochs 35, batch size 16, step size 0.001 -> 0.0001 (17) -> 0.00001 (26), L2 decay 0

SGD momentum 0.9

|  |
| --- |
| class fc\_model(nn.Module):  def \_\_init\_\_(self):  super(fc\_model, self).\_\_init\_\_()  self.fc1 = nn.Linear(57, 128)  self.fc2 = nn.Linear(128, 128)  self.fc3 = nn.Linear(128, 2)  def forward(self, inputs):  fc1\_out = F.relu(self.fc1(inputs))  fc2\_out = F.relu(self.fc2(fc1\_out))  fc3\_out = self.fc3(fc2\_out)  return fc3\_out  fc3\_out -> softmax -> cross entropy (CE) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.00893+-0.00036 | 0.01188+-0.00085 | 0.94997+-0.00313 | 0.93670+-0.00455 |
| mixup alpha8.0 | 0.01627+-0.00032 | 0.017108+-0.00052 | 0.92292+-0.00345 | 0.91839+-0.00603 |
| mixup alpha4.0 | 0.01582+-0.00036 | 0.01673+-0.00046 | 0.92510+-0.00373 | 0.92063+-0.00555 |
| mixup alpha1.0 | 0.01409+-0.00032 | 0.01533+-0.00055 | 0.93234+-0.00326 | 0.92551+-0.00550 |
| mixup alpha0.5 | 0.01282+-0.00031 | 0.01415+-0.00048 | 0.93755+-0.00309 | 0.92968+-0.00491 |
| mixup alpha0.2 | 0.01128+-0.00028 | 0.01295+-0.00065 | 0.94257+-0.00285 | 0.93364+-0.00539 |
| mixup sc alpha8.0 | 0.01073+-0.00045 | 0.01323+-0.00112 | 0.93854+-0.00331 | 0.93032+-0.00519 |
| mixup sc alpha4.0 | 0.01057+-0.00036 | 0.01325+-0.00118 | 0.93921+-0.00294 | 0.93116+-0.00480 |
| mixup sc alpha1.0 | 0.00982+-0.00038 | 0.01278+-0.00124 | 0.94314+-0.00281 | 0.93250+-0.00474 |
| mixup sc alpha0.5 | 0.00954+-0.00040 | 0.01237+-0.00106 | 0.94558+-0.00301 | 0.93366+-0.00502 |
| mixup sc alpha0.2 | 0.00928+-0.00039 | 0.01184+-0.00088 | 0.94678+-0.00296 | 0.93630+-0.00442 |
| gauss VRM std0.25 | 0.01006+-0.00039 | 0.01233+-0.00079 | 0.94525+-0.00302 | 0.93512+-0.00485 |
| gauss VRM std0.5 | 0.01213+-0.00041 | 0.01420+-0.00076 | 0.93799+-0.00329 | 0.92825+-0.00478 |
| gauss VRM std0.75 | 0.01493+-0.00047 | 0.01615+-0.00059 | 0.92950+-0.00369 | 0.92450+-0.00510 |
| gauss VRM std1.0 | 0.01758+-0.00050 | 0.01858+-0.00054 | 0.92464+-0.00378 | 0.92100+-0.00642 |
| gauss VRM std4.0 | 0.03619+-0.00045 | 0.03644+-0.00043 | 0.88135+-0.02480 | 0.87889+-0.02704 |

abalone

validations 100, epochs 35, batch size 16, step size 0.01 -> 0.001 (17) -> 0.0001 (26), L2 decay 0

SGD momentum 0.9

|  |
| --- |
| class fc\_model(nn.Module):  def \_\_init\_\_(self):  super(fc\_model, self).\_\_init\_\_()  self.fc1 = nn.Linear(10, 128)  self.fc2 = nn.Linear(128, 128)  self.fc3 = nn.Linear(128, 28)  def forward(self, inputs):  fc1\_out = F.relu(self.fc1(inputs))  fc2\_out = F.relu(self.fc2(fc1\_out))  fc3\_out = self.fc3(fc2\_out)  return fc3\_out  fc3\_out -> softmax -> cross entropy (CE) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.11392+-0.00073 | 0.12203+-0.00128 | 0.31953+-0.00619 | 0.28263+-0.00783 |
| mixup alpha8.0 | 0.12061+-0.00066 | 0.12449+-0.00103 | 0.29679+-0.00569 | 0.27717+-0.00764 |
| mixup alpha4.0 | 0.12004+-0.00079 | 0.12410+-0.00110 | 0.29942+-0.00575 | 0.27829+-0.00752 |
| mixup alpha1.0 | 0.11798+-0.00077 | 0.12304+-0.00126 | 0.30629+-0.00584 | 0.28084+-0.00794 |
| mixup alpha0.5 | 0.11667+-0.00073 | 0.12247+-0.00115 | 0.31071+-0.00627 | 0.28161+-0.00735 |
| mixup alpha0.2 | 0.11533+-0.00074 | 0.12220+-0.00114 | 0.31481+-0.00599 | 0.28061+-0.00073 |
| gauss VRM std0.25 | 0.11814+-0.00078 | 0.12422+-0.00140 | 0.30358+-0.00573 | 0.27990+-0.00821 |
| gauss VRM std0.5 | 0.12251+-0.00074 | 0.12713+-0.00131 | 0.29225+-0.00559 | 0.27674+-0.00800 |
| gauss VRM std0.75 | 0.12565+-0.00071 | 0.12914+-0.00106 | 0.28499+-0.00545 | 0.27240+-0.00840 |
| gauss VRM std1.0 | 0.12800+-0.00076 | 0.13089+-0.00114 | 0.27901+-0.00585 | 0.26888+-0.00805 |
| gauss VRM std4.0 | 0.14691+-0.00092 | 0.14798+-0.00088 | 0.23013+-0.00885 | 0.22750+-0.01034 |

iris

validations 100, epochs 35, batch size 16, step size 0.01 -> 0.001 (17) -> 0.0001 (26), L2 decay 0

SGD momentum 0.9

|  |
| --- |
| class fc\_model(nn.Module):  def \_\_init\_\_(self):  super(fc\_model, self).\_\_init\_\_()  self.fc1 = nn.Linear(4, 128)  self.fc2 = nn.Linear(128, 128)  self.fc3 = nn.Linear(128, 3)  def forward(self, inputs):  fc1\_out = F.relu(self.fc1(inputs))  fc2\_out = F.relu(self.fc2(fc1\_out))  fc3\_out = self.fc3(fc2\_out)  return fc3\_out  fc3\_out -> softmax -> cross entropy (CE) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.00696+-0.00131 | 0.00963+-0.00256 | 0.97389+-0.01364 | 0.95600+-0.02316 |
| mixup alpha8.0 | 0.02093+-0.00143 | 0.02284+-0.00259 | 0.90000+-0.02155 | 0.89683+-0.04282 |
| mixup alpha4.0 | 0.02050+-0.00120 | 0.02269+-0.00297 | 0.90478+-0.01854 | 0.89450+-0.03852 |
| mixup alpha1.0 | 0.01681+-0.00147 | 0.01943+-0.00331 | 0.93078+-0.01852 | 0.91150+-0.04234 |
| mixup alpha0.5 | 0.01428+-0.00150 | 0.01752+-0.00347 | 0.94456+-0.01809 | 0.92050+-0.04062 |
| mixup alpha0.2 | 0.01111+-0.00139 | 0.01419+-0.00314 | 0.96244+-0.01676 | 0.93617+-0.03249 |
| gauss VRM std0.25 | 0.00862+-0.00139 | 0.01114+-0.00299 | 0.97067+-0.01222 | 0.95617+-0.02883 |
| gauss VRM std0.5 | 0.01203+-0.00137 | 0.01390+-0.00239 | 0.95944+-0.01574 | 0.94833+-0.03010 |
| gauss VRM std0.75 | 0.01658+-0.00141 | 0.01738+-0.00215 | 0.94056+-0.01787 | 0.93483+-0.03473 |
| gauss VRM std1.0 | 0.02072+-0.00139 | 0.02121+-0.00230 | 0.92511+-0.02129 | 0.91917+-0.04037 |
| gauss VRM std4.0 | 0.05824+-0.00189 | 0.05859+-0.00276 | 0.88167+-0.05330 | 0.85767+-0.06869 |

wine

validations 100, epochs 35, batch size 16, step size 0.01 -> 0.001 (17) -> 0.0001 (26), L2 decay 0

SGD momentum 0.9

|  |
| --- |
| class fc\_model(nn.Module):  def \_\_init\_\_(self):  super(fc\_model, self).\_\_init\_\_()  self.fc1 = nn.Linear(13, 128)  self.fc2 = nn.Linear(128, 128)  self.fc3 = nn.Linear(128, 3)  def forward(self, inputs):  fc1\_out = F.relu(self.fc1(inputs))  fc2\_out = F.relu(self.fc2(fc1\_out))  fc3\_out = self.fc3(fc2\_out)  return fc3\_out  fc3\_out -> softmax -> cross entropy (CE) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.00106+-0.00020 | 0.00472+-0.00256 | 1.00000+-0.00000 | 0.98267+-0.01658 |
| mixup alpha8.0 | 0.00714+-0.00067 | 0.00990+-0.00150 | 0.99566+-0.00489 | 0.98569+-0.01324 |
| mixup alpha4.0 | 0.00699+-0.00063 | 0.00982+-0.00141 | 0.99698+-0.00479 | 0.98514+-0.01277 |
| mixup alpha1.0 | 0.00576+-0.00058 | 0.00889+-0.00171 | 0.99868+-0.00327 | 0.98361+-0.01380 |
| mixup alpha0.5 | 0.00475+-0.00057 | 0.00798+-0.00184 | 0.99877+-0.00317 | 0.98458+-0.01193 |
| mixup alpha0.2 | 0.00311+-0.00039 | 0.00639+-0.00210 | 0.99953+-0.00206 | 0.98361+-0.01514 |
| gauss VRM std0.25 | 0.00106+-0.00025 | 0.00448+-0.00188 | 0.99991+-0.00094 | 0.98333+-0.01375 |
| gauss VRM std0.5 | 0.00141+-0.00036 | 0.00447+-0.00193 | 0.99915+-0.00270 | 0.98569+-0.01381 |
| gauss VRM std0.75 | 0.00239+-0.00052 | 0.00514+-0.00190 | 0.99708+-0.00436 | 0.98639+-0.01402 |
| gauss VRM std1.0 | 0.00448+-0.00076 | 0.00681+-0.00149 | 0.99538+-0.00472 | 0.98625+-0.01250 |
| gauss VRM std4.0 | 0.05095+-0.00170 | 0.05445+-0.00187 | 0.97151+-0.01399 | 0.96306+-0.02301 |

abalone

validations 100, epochs 10, batch size 16, step size 0.01 -> 0.001 (5) -> 0.0001 (7), L2 decay 0

SGD momentum 0.9

|  |
| --- |
| class fc\_model(nn.Module):  def \_\_init\_\_(self):  super(fc\_model, self).\_\_init\_\_()  self.fc1 = nn.Linear(10, 128)  self.fc2 = nn.Linear(128, 128)  self.fc3 = nn.Linear(128, 28)  def forward(self, inputs):  fc1\_out = F.relu(self.fc1(inputs))  fc2\_out = F.relu(self.fc2(fc1\_out))  fc3\_out = self.fc3(fc2\_out)  return fc3\_out  fc3\_out -> softmax -> cross entropy (CE) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.11957+-0.00077 | 0.12317+-0.00132 | 0.29664+-0.00505 | 0.27690+-0.00792 |
| mixup alpha8.0 | 0.12557+-0.00090 | 0.12804+-0.00127 | 0.27941+-0.00700 | 0.26557+-0.00971 |
| mixup alpha4.0 | 0.12500+-0.00084 | 0.12765+-0.00122 | 0.28166+-0.00640 | 0.26746+-0.00874 |
| mixup alpha1.0 | 0.12308+-0.00066 | 0.12604+-0.00114 | 0.28889+-0.00498 | 0.27172+-0.00844 |
| mixup alpha0.5 | 0.12196+-0.00068 | 0.12497+-0.00114 | 0.29162+-0.00623 | 0.27508+-0.00895 |
| mixup alpha0.2 | 0.12072+-0.00084 | 0.12415+-0.00135 | 0.29439+-0.00650 | 0.27514+-0.00916 |
| gauss VRM std0.25 | 0.12176+-0.00078 | 0.12526+-0.00125 | 0.29065+-0.00549 | 0.27479+-0.00816 |
| gauss VRM std0.5 | 0.12534+-0.00074 | 0.12848+-0.00113 | 0.28510+-0.00618 | 0.26874+-0.00891 |
| gauss VRM std0.75 | 0.12837+-0.00078 | 0.13094+-0.00108 | 0.27619+-0.00610 | 0.26529+-0.00947 |
| gauss VRM std1.0 | 0.13117+-0.00083 | 0.13328+-0.00094 | 0.27187+-0.00618 | 0.26232+-0.00916 |
| gauss VRM std4.0 | 0.15619+-0.00116 | 0.15716+-0.00097 | 0.22586+-0.0109 | 0.22134+-0.01198 |

iris

validations 100, epochs 10, batch size 16, step size 0.01 -> 0.001 (5) -> 0.0001 (7), L2 decay 0

SGD momentum 0.9

|  |
| --- |
| class fc\_model(nn.Module):  def \_\_init\_\_(self):  super(fc\_model, self).\_\_init\_\_()  self.fc1 = nn.Linear(4, 128)  self.fc2 = nn.Linear(128, 128)  self.fc3 = nn.Linear(128, 3)  def forward(self, inputs):  fc1\_out = F.relu(self.fc1(inputs))  fc2\_out = F.relu(self.fc2(fc1\_out))  fc3\_out = self.fc3(fc2\_out)  return fc3\_out  fc3\_out -> softmax -> cross entropy (CE) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.02320+-0.00181 | 0.02549+-0.00384 | 0.87123+-0.02965 | 0.85845+-0.06082 |
| mixup alpha8.0 | 0.032979+-0.00246 | 0.03546+-0.00353 | 0.82433+-0.04776 | 0.79917+-0.08361 |
| mixup alpha4.0 | 0.03275+-0.00244 | 0.03426+-0.00395 | 0.83256+-0.04622 | 0.81900+-0.07962 |
| mixup alpha1.0 | 0.02922+-0.00216 | 0.03148+-0.00397 | 0.84744+-0.03688 | 0.82067+-0.07517 |
| mixup alpha0.5 | 0.02796+-0.00222 | 0.02942+-0.00387 | 0.84856+-0.03770 | 0.83483+-0.0723 |
| mixup alpha0.2 | 0.02532+-0.00226 | 0.02740+-0.00326 | 0.86000+-0.03439 | 0.84133+-0.06044 |
| gauss VRM std0.25 | 0.02348+-0.00203 | 0.02579+-0.00322 | 0.86611+-0.02852 | 0.85200+-0.05101 |
| gauss VRM std0.5 | 0.02459+-0.00197 | 0.02589+-0.00324 | 0.86878+-0.02858 | 0.85750+-0.05464 |
| gauss VRM std0.75 | 0.02583+-0.00204 | 0.02709+-0.00247 | 0.86356+-0.03228 | 0.86217+-0.05405 |
| gauss VRM std1.0 | 0.02778+-0.00197 | 0.02856+-0.00292 | 0.86011+-0.02749 | 0.84983+-0.04947 |
| gauss VRM std4.0 | 0.05874+-0.00200 | 0.05889+-0.00227 | 0.82322+-0.06358 | 0.81250+-0.07797 |

wine

validations 100, epochs 10, batch size 16, step size 0.01 -> 0.001 (5) -> 0.0001 (7), L2 decay 0

SGD momentum 0.9

|  |
| --- |
| class fc\_model(nn.Module):  def \_\_init\_\_(self):  super(fc\_model, self).\_\_init\_\_()  self.fc1 = nn.Linear(13, 128)  self.fc2 = nn.Linear(128, 128)  self.fc3 = nn.Linear(128, 3)  def forward(self, inputs):  fc1\_out = F.relu(self.fc1(inputs))  fc2\_out = F.relu(self.fc2(fc1\_out))  fc3\_out = self.fc3(fc2\_out)  return fc3\_out  fc3\_out -> softmax -> cross entropy (CE) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.01296+-0.00228 | 0.01595+-0.00266 | 0.98320+-0.01161 | 0.97127+-0.01886 |
| mixup alpha8.0 | 0.03158+-0.00330 | 0.03522+-0.00362 | 0.96491+-0.02327 | 0.94861+-0.03307 |
| mixup alpha4.0 | 0.03003+-0.00343 | 0.03357+-0.00419 | 0.96840+-0.01976 | 0.95167+-0.03371 |
| mixup alpha1.0 | 0.02424+-0.00324 | 0.027908+-0.00389 | 0.97509+-0.01471 | 0.95653+-0.02803 |
| mixup alpha0.5 | 0.02039+-0.00300 | 0.02313+-0.00363 | 0.97679+-0.01344 | 0.96694+-0.02220 |
| mixup alpha0.2 | 0.01665+-0.00287 | 0.01990+-0.00338 | 0.98094+-0.01124 | 0.96694+-0.02506 |
| gauss VRM std0.25 | 0.01324+-0.00226 | 0.01566+-0.00272 | 0.98255+-0.00975 | 0.97056+-0.02118 |
| gauss VRM std0.5 | 0.01348+-0.00235 | 0.01678+-0.00262 | 0.98462+-0.00882 | 0.96958+-0.01839 |
| gauss VRM std0.75 | 0.01498+-0.00235 | 0.01760+-0.00287 | 0.97972+-0.01200 | 0.97083+-0.02000 |
| gauss VRM std1.0 | 0.01703+-0.00241 | 0.01989+-0.00260 | 0.97792+-0.01274 | 0.96653+-0.01885 |
| gauss VRM std4.0 | 0.05201+-0.00230 | 0.05538+-0.00268 | 0.95274+-0.02388 | 0.94667+-0.03256 |

FashionMNIST

validations 1, epochs 30, batch size 128, step size 0.01 -> 0.001 (15) -> 0.0001 (22), L2 decay 0

SGD momentum 0.9

ResNet18 No Pretrain; softmax -> cross entropy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train loss | test loss | train accuracy | test accuracy |
| standard | 0.00001 | 0.00225 | 0.99998 | 0.91840 |
| mixup alpha8.0 | 0.00174 | 0.00225 | 0.93593 | 0.90890 |
| mixup alpha4.0 | 0.00169 | 0.00224 | 0.94270 | 0.91410 |
| mixup alpha1.0 | 0.00122 | 0.00221 | 0.97115 | 0.91900 |
| mixup alpha0.5 | 0.00076 | 0.00207 | 0.98838 | 0.92320 |
| mixup alpha0.2 | 0.00020 | 0.00205 | 0.99768 | 0.92290 |
| gauss VRM std0.25 | 0.00027 | 0.00218 | 0.98753 | 0.91630 |
| gauss VRM std0.5 | 0.00106 | 0.00235 | 0.94475 | 0.89860 |
| gauss VRM std0.75 | 0.00167 | 0.00256 | 0.91362 | 0.88050 |
| gauss VRM std1.0 | 0.00218 | 0.00285 | 0.89080 | 0.86840 |
| gauss VRM std4.0 | 0.00610 | 0.00628 | 0.71918 | 0.70740 |