Anil Poonai

Github link for files and code: https://github.com/DevonARP/DeepLearning_A1

Problem 1: Part A

F(x) = $||X||_2^2$

F(x) = $((\sum_{i=1}^{n} x_i^2)^{1/2})2$

F(x) = $\sum_{i=1}^{n} x_i^2$

Can ignore the summation as n is 1.

$\frac{\partial}{\partial x}$ = 2x

Problem 1: Part B

$\sum_{i=1}^{n} ||x_i - \mu||_2^2$

$\frac{\partial}{\partial x}$ = 2(x- $\mu$) Got the derivative, could just put it in the answer for Part A but accounting for $\mu$. This works because this is the derivative at each point, and we have the summation of all of the points already in the expression.

Now we set the derivative with summation $= 0$

$\sum_{i=1}^{n} 2(x_i - \mu) = 0$

Do some algebra and we end up with:

$\mu = (\sum_{i=1}^{n} x_i)/n$

Problem 2: Part A

L(w) = $||x||_1$

L(w) = $||Xw - y||_1$

Sizes: X(data) -> n*c, w(weights) -> c*1, y(labels) -> n*1

Problem 2: Part B

No, because L1 norm isn't differentiable at zero therefore gradient optimization cannot be used. The explanation is partly explained in Part C as well.

Problem 2: Part C

Part A was straightforward since we just need the difference from the estimate and actual value. Since L1 is the summation of ||x||, we can just replace Xw-y for x. For part B, it wouldn't have a value at zero if we took the derivative as it's non-convex there, there would be no unique global minimum so we can't minimize the loss function. This would mostly be concerned with the loss function part of the 3-step recipe as we are calculating the losses from the prediction regarding the label.

Problem 3: Part A

The code for this question is below, the answers I got for the parameters in this order $W_{h1}$, $W_{h2}$, $W_{h3}$, $W_{h4}$, $b_{h1}$, $b_{h2}$, $b_{h3}$, $b_{h4}$, $W_{o1}$, $W_{o2}$, $W_{o3}$, $W_{o4}$, $b_{0.}$

For Dataset 1:
([-7.82455998e-01, -5.16665284e-01, 1.32120087e+00, -3.55189485e+00, 6.25964798e+00, 3.09863545e+00, -1.05696070e+01, 1.42169031e+01, 2.55605428e+00, -7.76233237e+00, 1.51377436e+00, 1.12912342e+00, 2.69958647e-11]

For Dataset 2:
[ -0.35614338, 0.05176055, 0.04679073, 4.26739128, 1.06854914, 0.69438611, 0.33182292, -21.3355306 , -5.61759265, -20.45828289, 1.24519253, 0.46882705, 19.79543786]

I also made an excel solver to calculate these values out for a more insightful look into the process.

For Dataset 1:

| Parameters | | | Input | Z1 | Z2 | Z3 | Z4 | ReLu1 | ReLu2 | ReLu3 | ReLu4 | Prediction | | Value for D1 | Value for D2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| w1 | = | -7.82E-01 | 0 | 6.25964798 | 3.09863545 | -10.569607 | 14.2169031 | 6.25964798 | 3.09863545 | 0 | 14.2169031 | 8.000000004 | = | 8 | 0 |
| w2 | = | -5.17E-01 | 1 | 5.477191982 | 2.581970166 | -9.24840613 | 10.66500825 | 5.477191982 | 2.581970166 | 0 | 10.66500825 | 6 | = | 6 | 1 |
| w3 | = | 1.32E+00 | 2 | 4.694735984 | 2.065304882 | -7.92720526 | 7.1131134 | 4.694735984 | 2.065304882 | 0 | 7.1131134 | 3.999999995 | = | 4 | 2 |
| w4 | = | -3.55E+00 | 3 | 3.912279986 | 1.548639598 | -6.60600439 | 3.56121855 | 3.912279986 | 1.548639598 | 0 | 3.56121855 | 1.99999999 | = | 2 | 3 |
| b1 | = | 6.26E+00 | 4 | 3.129823988 | 1.031974314 | -5.28480352 | 0.0093237 | 3.129823988 | 1.031974314 | 0 | 0.0093237 | -1.4339E-08 | = | 0 | 2 |
| b2 | = | 3.10E+00 | 5 | 2.34736799 | 0.51530903 | -3.96360265 | -3.54257115 | 2.34736799 | 0.51530903 | 0 | 0 | 2.000000033 | = | 2 | 1 |
| b3 | = | -1.06E+01 | 6 | 1.564911992 | -0.00135625 | -2.64240178 | -7.094466 | 1.564911992 | 0 | 0 | 0 | 3.999999995 | = | 4 | 2 |
| b4 | = | 1.42E+01 | 7 | 0.782455994 | -0.51802154 | -1.32120091 | -10.6463609 | 0.782455994 | 0 | 0 | 0 | 1.999999992 | = | 2 | 3 |
| w5 | = | 2.56E+00 | 8 | -4E-09 | -1.03468682 | -4E-08 | -14.1982557 | 0 | 0 | 0 | 0 | 2.69959E-11 | = | 0 | 4 |
| w6 | = | -7.76E+00 | 9 | -0.782456 | -1.55135211 | 1.32120083 | -17.7501506 | 0 | 0 | 1.32120083 | 0 | 1.999999941 | = | 2 | 5 |
| w7 | = | 1.51E+00 | 10 | -1.564912 | -2.06801739 | 2.6424017 | -21.3020454 | 0 | 0 | 2.6424017 | 0 | 3.999999942 | = | 4 | 6 |
| w8 | = | 1.13E+00 | 11 | -2.347368 | -2.58468267 | 3.96360257 | -24.8539403 | 0 | 0 | 3.96360257 | 0 | 5.999999944 | = | 6 | |
| b0 | = | 2.70E-11 | 12 | -3.129824 | -3.10134796 | 5.28480344 | -28.4058351 | 0 | 0 | 5.28480344 | 0 | 7.999999945 | = | 8 | |

For Dataset 2:

| Parameters | | | Input | Z1 | Z2 | Z3 | Z4 | ReLu1 | ReLu2 | ReLu3 | ReLu4 | Prediction | | Value for D1 | Value for D2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| w1 | = | -0.35614338 | 0 | 1.06854914 | 0.69438611 | 0.33182292 | -21.3355306 | 1.06854914 | 0.69438611 | 0.33182292 | 0 | 1.29723E-08 | = | 8 | 0 |
| w2 | = | 0.05176055 | 1 | 0.71240576 | 0.74614666 | 0.37861365 | -17.0681393 | 0.71240576 | 0.74614666 | 0.37861365 | 0 | 0.99999994 | = | 6 | 1 |
| w3 | = | 0.04679073 | 2 | 0.35626238 | 0.79790721 | 0.42540438 | -12.800748 | 0.35626238 | 0.79790721 | 0.42540438 | 0 | 1.999999867 | = | 4 | 2 |
| w4 | = | 4.26739128 | 3 | 0.000119 | 0.84966776 | 0.47219511 | -8.53335676 | 0.000119 | 0.84966776 | 0.47219511 | 0 | 2.999999794 | = | 2 | 3 |
| b1 | = | 1.06854914 | 4 | -0.35602438 | 0.90142831 | 0.51898584 | -4.26596548 | 0 | 0.90142831 | 0.51898584 | 0 | 1.99999978 | = | 0 | 2 |
| b2 | = | 0.69438611 | 5 | -0.71216776 | 0.95318886 | 0.56577657 | 0.0014258 | 0 | 0.95318886 | 0.56577657 | 0.0014258 | 0.999999727 | = | 2 | 1 |
| b3 | = | 0.33182292 | 6 | -1.06831114 | 1.00494941 | 0.6125673 | 4.26881708 | 0 | 1.00494941 | 0.6125673 | 4.26881708 | 1.999999685 | = | 4 | 2 |
| b4 | = | -21.3355306 | 7 | -1.42445452 | 1.05670996 | 0.65935803 | 8.53620836 | 0 | 1.05670996 | 0.65935803 | 8.53620836 | 2.999999643 | = | 2 | 3 |
| w5 | = | -5.61759265 | 8 | -1.7805979 | 1.10847051 | 0.70614876 | 12.80359964 | 0 | 1.10847051 | 0.70614876 | 12.80359964 | 3.999999601 | = | 0 | 4 |
| w6 | = | -20.4582829 | 9 | -2.13674128 | 1.16023106 | 0.75293949 | 17.07099092 | 0 | 1.16023106 | 0.75293949 | 17.07099092 | 4.999999559 | = | 2 | 5 |
| w7 | = | 1.24519253 | 10 | -2.49288466 | 1.21199161 | 0.79973022 | 21.3383822 | 0 | 1.21199161 | 0.79973022 | 21.3383822 | 5.999999517 | = | 4 | 6 |
| w8 | = | 0.46882705 | 11 | -2.84902804 | 1.26375216 | 0.84652095 | 25.60577348 | 0 | 1.26375216 | 0.84652095 | 25.60577348 | 6.999999475 | = | 6 | |
| b0 | = | 19.79543786 | 12 | -3.20517142 | 1.31551271 | 0.89331168 | 29.87316476 | 0 | 1.31551271 | 0.89331168 | 29.87316476 | 7.999999433 | = | 8 | |

The predicted values are really close to the actual values on all accounts.

Problem 3: Part B

This is just mean squares, so the derivative is straightforward:

$L(\theta^{\rightarrow}) = \sum_{i=1}^{n} (y_i - f(x_i, \Theta^{\rightarrow}))^2$

$\nabla L(\theta^{\rightarrow}) = -2 \sum_{i=1}^{n} ((y_i - f(x_i, \Theta^{\rightarrow})) * \nabla f(x_i, \Theta^{\rightarrow}))$

I left it in terms of $\nabla f(x_i, \Theta^{\rightarrow})$ since that's what the question asked for.

Problem 3: Part C

First layer values after plugging in the input(x) and parameters.

$Z_1 = -1$

$Z_2 = 3$

$Z_3 = 1$

$Z_4 = -1$

After ReLu

$Z_1 = 0$

$Z_2 = 3$

$Z_3 = 1$

$Z_4 = 0$

After summation in output layer

-4

After adding final bias

-3

Did this both manually and on the excel calculator I made:

P3. C: $\bar{z}_i = w_{hi} x + b_{hi}$

relu

$Y = b_0 + \sum_{i=1}^{4} w_{0i} z_i$

$x = 2$

$\bar{z}_1 = -1 \rightarrow 0 \rightarrow 0$

$\bar{z}_2 = 3 \rightarrow 3 \rightarrow -3$

$\bar{z}_3 = 1 \rightarrow 1 \rightarrow -1$

$\bar{z}_4 = -1 \rightarrow 0 \rightarrow 0$

$-4$

$\boxed{-3}$

| | | Parameters | Input | Z1 | Z2 | Z3 | Z4 | ReLu1 | ReLu2 | ReLu3 | ReLu4 | Prediction |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| w1 | = | -1 | 0 | 1 | 1 | -1 | 1 | 1 | 1 | 0 | 1 | 0 |
| w2 | = | 1 | 1 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | -1 |
| w3 | = | 1 | 2 | -1 | 3 | 1 | -1 | 0 | 3 | 1 | 0 | -3 |
| w4 | = | -1 | 3 | -2 | 4 | 2 | -2 | 0 | 4 | 2 | 0 | -5 |

Problem 3: Part D

I have all the formulas written out but am not sure what the parameters are supposed to be. But once I'm given those, I can figure out what the derivative at x=2 is with no problem.

$\frac{\partial L}{\partial w0} = \frac{\partial L}{\partial y} * \frac{\partial y}{\partial w0}$

$\frac{\partial y}{\partial w0} = \sum_{i=1}^{4} z_i$

$\frac{\partial L}{\partial y} = 2 \sum_{i=1}^{n} (y0 - f(x, \theta))$

$\frac{\partial L}{\partial \mathrm{w0}}$ = 2 * $\sum_{i=1}^{4} z$ * $\sum_{i=1}^{n}(y0 - f(x,\theta))$

------------------------------------------

$\frac{\partial L}{\partial b0}$ = $\frac{\partial L}{\partial \mathrm{y}}$ * $\frac{\partial y}{\partial \mathrm{b0}}$

$\frac{\partial y}{\partial \mathrm{b0}}$ = 1

$\frac{\partial L}{\partial \mathrm{y}}$ = 2 $\sum_{i=1}^{n}(y0 - f(x,\theta))$

$\frac{\partial L}{\partial b0}$ = 2 $\sum_{i=1}^{n}(y0 - f(x,\theta))$

------------------------------------------

$\frac{\partial L}{\partial \mathrm{w1}}$ = $\frac{\partial L}{\partial \mathrm{y}}$ * $\frac{\partial y}{\partial \mathrm{z}}$ * max(0,z) * $\frac{\partial z}{\partial \mathrm{wo}}$

$\frac{\partial z}{\partial \mathrm{wo}}$ = x -> max(0,z) = max(0,x)

$\frac{\partial y}{\partial \mathrm{z}}$ = $\sum_{i=1}^{4} wo_i$

$\frac{\partial L}{\partial \mathrm{y}}$ = 2 $\sum_{i=1}^{n}(y0 - f(x,\theta))$

$\frac{\partial L}{\partial \mathrm{w1}}$ 2 * $\sum_{i=1}^{4} wo$ * max(0,x) * $\sum_{i=1}^{n}(y0 - f(x,\theta))$

------------------------------------------

$\frac{\partial L}{\partial \mathrm{b1}}$ = $\frac{\partial L}{\partial \mathrm{y}}$ * $\frac{\partial y}{\partial \mathrm{z}}$ * max(0,z) * $\frac{\partial z}{\partial \mathrm{bo}}$

$\frac{\partial z}{\partial \mathrm{bo}}$ = 1 -> max(0,z) = 1

$\frac{\partial y}{\partial \mathrm{z}}$ = $\sum_{i=1}^{4} wo_i$

$\frac{\partial L}{\partial \mathrm{y}}$ = 2 $\sum_{i=1}^{n}(y0 - f(x,\theta))$

$\frac{\partial L}{\partial \mathrm{w1}}$ 2 * $\sum_{i=1}^{4} wo$ * $\sum_{i=1}^{n}(y0 - f(x,\theta))$