**Honor Agreement**

This is an **individual** take-home exam. You will do this alone and without anyone else. By signing here, you agree not to engage in any form of exchange, with respect to this exam, with any other humans. You also agree that if any exchange is reported or discovered, and proved, that you accept an F for the course as well as any punishment imposed by the CU Boulder Board of Ethics (which will remain on your permanent record).

I, \_\_\_\_Xingyu Chen\_\_, agree to complete this exam individually and without any form of exchange with any other human(s). I also agree to the statements and penalties noted above. I understand that **it is OK** to use the Web, Book-style resources, **resources and code posted by Dr. Gates**, etc. I also understand that I am NOT permitted to discuss, share, exchange, compare, or engage in assistance (giving or receiving) from any humans (in the class or not). Finally, if I am confused or unsure about whether it is acceptable to use a resource, I will simply not use that resource.

All exams may be run through Turnitin.  
Student Signature and Date: Please print and sign and date here.

Xingyu Chen \_\_10/03/2022\_\_\_\_\_\_\_\_\_\_\_\_\_

**Part 1:**

Suppose you have a NN such that X is n by 4, H (the one hidden layer) has 3 hidden units, and there is one output, Y^. The hidden layer activation function is the Sigmoid. The output activation function is the Sigmoid. The Loss function is SUM 1/2 (yhat - y)^2.

Write out the complete derivative set for **dL/dw11** for n = 3 X rows. For example,

[ [x11, x12, x13, x14]

[x21, x22, x23, x24],

[x31, x32, x33, x34] ]

Show the chain of derivatives, the derivative values, and other details that illustrate your understanding of the concepts. We are not using “numbers” here – so there will be no “numeric” values.

**Answer:**

We have X = [ [x11, x12, x13, x14]. Y = [[y1],[y2],[y3]] W1 = [[w11,w12,w13],  
 [x21, x22, x23, x24], [w21,w22,w23],  
 [x31, x32, x33, x34] ] [w31,w32,w33],

[w41,w42,w43]]

W2 = [w1, w2, w3]

Z1 = X @ W1 + b

[**z11** = x11**w11** + x12w21 + x13w31 + x14w41 + b1 z12 z13

**z21**= x21**w11** + x22w21 + x23w31 + x24w41 + b1 z22 z23

**z31** = x31**w11** + x32w21+x33w31+x34w41+b1 z32 z33] shape n by h

H1 = [**h11** = sig(**z11**) h12 = Sig(z12) h13 = Sig(z13)

**h21** = sig(**z21**) h22 = Sig(z22) h23 = Sig(z23)

**h31** = sig(**z31**) h32 = Sig(z32) h33 = Sig(z33)

Z2 = [ z(2)1= **h11**w1 + h12w2 + h13w3 + c Y^ = [ y^1 = sig(z(2)1)

z(2)2 = **h21**w1 + h22w2 + h23w3 + c y^2 = sig(z(2)2)

z(2)3 = **h31**w1 + h32w2 + h33w3 + c] y^3 = sig(z(2)3)]

dL/dw11 = dz11/dw11 \* dh11/dz11 \* dz(2)1/dh11 \* dy^1/dz(2)1\* dL/dy^1 +

dL/dw11 = dz21/dw11 \* dh21/dz21 \* dz(2)2/dh21 \* dy^2/dz(2)2\* dL/dy^2 +

dL/dw11 = dz31/dw11 \* dh31/dz31 \* dz(2)3/dh31 \* dy^3/dz(2)3\* dL/dy^3

x11 \* (h11)(1 – h11) \* w1 \* (y^1)(1 – y^1) \* (y^1– y1) +

x21 \* (h21)(1 – h21) \* w1 \* (y^2)(1 – y^2) \* (y^2– y2) +

x31 \* (h31)(1 – h31) \* w1 \* (y^3)(1 – y^3) \* (y^3– y3)

**Part 2**:

Suppose you have the following NN details: X is n by 4, H (the number of hidden units in the one hidden layer) is 2, and each output, y^ has  3 values  (y^1, y^2, and y^3). The activation function for H is the Sigmoid and the activation function for the output layer is Softmax. The labels, Y, are one-hot-encoded. The Loss function is Categorical Crossentropy.

**Answer these questions:**

1. Suppose Y is

 [[3]

 [1]

 [2]

 [3]

 [3] ]

 What would Y look like once it was one-hot-encoded?

**Answer**:

[[0,0,1]

[1,0,0]

[0,1,0]

[0,0,1]

[0,0,1]]

1. Suppose the NN produces this output for one of the inputs:

[  -.2,   1.4,   .89]

What would this output be once Softmax is applied?

Why is this softmax result a probability distribution?

**Answer**:

The softmax formula: e^x / Σe^x\_i

[0.112, 0.555, 0.333]

The normalized exponential function converts a vector of k real numbers into a probability distribution of k probabilities that all sum to 1, which means give a probability of each output

1. Suppose you read only one row of X ([ [x11, x12, x13, x14] ]) into this network. Show what the following would all be:

X = [ [x11, x12, x13, x14] ]

Y =  [0, 1, 0] (Hint: one hot encoded –you can pretend that the category is a 2. )

W1 = [[w11, w12]

[w21, w22]

[w31, w32]

[w41, w42]]

Z1 = [z11 = x11w11 + x12w21 + x13w31 + x14w41 + b1 z12 = x11w12 + x12w22 + x13w32 + x14w42 + b2]

H = [h1 = sig(z11) h2 = sig(z12)]

Z2 = [z2 = h1w1 + h2w2 + c]

Y^ = [softmax(z2)] (such that y^ is the final output after softmax activation)

1. Write out the chain of derivatives for dL/dw11 as well as what the derivatives would resolve into.

 Hint: When using one-hot-encoding, softmax, and categorical cross entropy, as you are using here, the last two derivatives (**dy^/dz2 and dL/dy^ )** can be expressed together as dL/dz and can be resolved as one value (as we proved in class).

dL/dw11 = dz11/dw11 \* dh11/dz11 \* dz(2)1/dh11 \* dy^1/dz(2)1\* dL/dy^1

= dz11/dw11 \* dh11/dz11 \* dz(2)1/dh11 \* dL/dz

= dz11/dw11 \* dh11/dz11 \* dz(2)1/dh11 \* (y^1 – y1)

= x11 \* (h11)(1-h11) \* w1 \* (y^1 – y1)