Quiz 2: Differentiation

Question 1

1 / 1 pts

Suppose $y = tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$, where $z = 2x_1 - 3x_2,\ x_1 = t^2,\ x_2 = 2t$. Use the chain rule to determine which of the following is $\frac{dy}{dt}$:

A: $-4y^2$

B: $(1 - y^2)(4t - 6)$

C: $-4(1 - y^2)$

D: $(4t - 3)\ (1 - y^2)$

- A
- B
- C
- D

$y = tanh(z)$, $y^{'} = 1 - y^2$

$\frac{dy}{dt} = \frac{dy}{dz}\frac{dz}{dx} = (1 - y^2)(4t - 6)$

Question 2

1 / 1 pts

Let $\sigma(z) = \frac{1}{1+e^{-z}}$, $\sigma(z) = 0.6.$, what is the derivative $\sigma'(z)$?

A: 0.24

B: -0.24

C: 0.6

D: 0.16

- (selected) A
- B
- C
- D

$\sigma(z) = \frac{1}{1+e^{-z}}$, $\sigma^{'}(z) = \sigma(z)(1 - \sigma(z)) = 0.6 * (1 - 0.6) = 0.24$

Question 3

1 / 1 pts

Let $C = (\sigma(z) - y)^2$, $\sigma(z) = \frac{1}{1+e^{-z}}$, $z = w_1x_1 - w_2x_2 + b.$ At some point of your model training, you have $\sigma(z) = 0.8,\ y = 1,\ x_1 = -2,\ x_2 = 1.$ What is the gradient of $\frac{\partial C}{\partial w_2}$?

- - 0.32
- (selected) 0.064
- -0.16
- 0.128

$$\frac{\partial C}{\partial w_2} = \frac{\partial C}{\partial \sigma(z)} \frac{\partial \sigma(z)}{\partial z} \frac{\partial z}{\partial w_2} = 2(\sigma(z) - y) * \sigma(z)(1 - \sigma(z)) * (-x_2) = 0.064$$

Question 4

1 / 1 pts

Let $C = (\sigma(z) - y)^2,\ \sigma(z) = \frac{1}{1+e^{-z}},\ z = w_1 x_1 - w_2 x_2 + b.$ At some point of your model training, you have $\sigma(z) = 0.8,\ y = 1,\ x_1 = -2,\ x_2 = 1.$ We would like to minimize C by adjusting w_1 and w_2. Which of the following is correct?

A: w_1 should be increased

B: w_2 should be increased

C: w_1 should be decreased

D: w_2 should be decreased

- [ ] A
- [ ] B
- [x] C
- [x] D

Like Q3,

$\frac{\partial C}{\partial w_2} = 0.064$, and $\frac{\partial C}{\partial w_1} = 0.128$

And the method we update w is:

$w = w - \alpha \frac{\partial C}{\partial w}$

So: both w_1 and w_2 should be decreased.

Question 5

1 / 1 pts

Let $L = -[y * \ln \hat{y} + (1-y) * ln(1-\hat{y})], \hat{y} = \sigma(z) = \frac{1}{1+e^{-z}}$, what is $\frac{dL}{dz}$?

A: $-\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}$

B: $-\frac{y\ \sigma'(z)}{\hat{y}} + \frac{(1-y)\ \sigma'(z)}{1-\hat{y}}$

C: $-y\ (1-\sigma(z)) + (1-y)\ \sigma(z)$

D: $\hat{y} - y$

- [ ] A
- [x] B
- [x] C
- [x] D

$\frac{dL}{dz} = -(\frac{y}{\hat{y}} - \frac{1-y}{1-\hat{y}})(\hat{y}(1-\hat{y}))$ which is equal to B.

$= -y(1-\hat{y}) + \hat{y}(1-y)$ which is equal to C.

$= \hat{y} - y$ which is equal to D.

Question 6

1 / 1 pts

$f = (y-z)^2, z = 2x_1 + 3x_2^2 + 1, y = 5, x_1 = -2, x_2 = 1. [\frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}, \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}]$ =?

```
A: [10, -10, -20, -60]

B: [10, -10, 20, 60]

C: [10, 10, -20, -30]

D: [-10, -10, -20, -60]
```

- ◉ A
- ○ B
- ○ C
- ○ D

$z = 2x_1 + 3x_2^2 + 1 = 0$

$\frac{\partial f}{\partial y} = 2(y - z) = 10$

$\frac{\partial f}{\partial z} = 2(z - y) = -10$

$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial x_1} = -10 * 2 = -20$

$\frac{\partial f}{\partial x_2} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial x_2} = -10 * 6 = -60$

Question 7

1 / 1 pts

$l = (y-z)^T(y-z)$, where $z = Wx + b, x \in R^8, W \in R^{3\times 8}, z, y, b \in R^3$

What is $\frac{\partial z_2}{\partial W_{12}}$? Note $z_2 = W_{2,:}x$, and $W_{2,:}$ is the 2nd row of W.

A: x_1

B: x_2

C: 0

D: x

- A
- B
- C
- D

$z_1 = w_{11}x_1 + w_{12}x_2 + \ldots + w_{18}x_8$

$z_2 = w_{21}x_1 + w_{22}x_2 + \ldots + w_{28}x_8$

$z_3 = w_{31}x_1 + w_{32}x_2 + \ldots + w_{38}x_8$

You don't need to use w_{12} when you calculate z_2. So $\frac{\partial z_2}{\partial w_{12}} = 0$.

Question 8

1 / 1 pts

$l = (y - z)^T (y - z)$, where $z = Wx + b, x \in R^8, W \in R^{3\times 8}, z, y, b \in R^3$

What is the dimension/size of the Jacobian matrix $\frac{\partial z}{\partial W}$?

- (3, 8)
- (3, 3)
- (3,3,3)
- (3,3,8)

Question 9

1 / 1 pts

Which of the statements about Backpropagation algorithm are correct?

A. Backpropagation tunes the parameters for a neural network and returns the best parameter values
B. Backpropagation calculates gradients based on the derivative chain rule
C. To calculate gradients, a feedforward step is needed to compute the output and the total loss using the current parameter values, and then a backpropagation step is carried out to estimate the impact of each weight or bias parameter on the total loss.
D. The goal of backpropagation is to compute the partial derivatives ∂C/∂w and ∂C/∂b of the cost function C with respect to any weight w or bias b in the network

- A
- B
- C
- D

Question 10

1 / 1 pts

Let $z_2 = f_1(w_1 z_1 + b_1), z_3 = f_2(w_2 z_2 + b_2), z_4 = f_3(w_3 z_3 + b_3), L = (y - z_4)^2$, where $f_1, f_2,$ and f_3 are functions. Let

$$y - z_4 = 0.25, f_1' = \frac{df_1(x)}{dx} = 0.6, f_2' = \frac{df_2(x)}{dx} = 0.5, f_3' = \frac{df_3(x)}{dx} = 0.1,$$

$$z_1 = 0.25, z_2 = 0.2, z_3 = 0.1$$

$$w_1 = 0.1, w_2 = -0.3, w_3 = 0.2$$

Which of the statements are correct:

A. The gradient $\frac{dz_4}{dw_2} = 0.02$

B. The gradient $\frac{dz_4}{dz_2} = -0.003$

C. The gradient $\frac{dL}{dw_1} = -2 * 0.25 * 0.1 * 0.2 * 0.5 * (-0.3) * 0.6 * 0.25$, which is quite small, i.e. vanishing gradient.

D. Regarding the absolute values of gradients, $|\frac{dL}{dw_3}| < |\frac{dL}{dw_2}|$ and $|\frac{dL}{dw_3}| < |\frac{dL}{dw_1}|$

- [ ] A
- [x] B
- [x] C
- [ ] D

$$\frac{dz_4}{dw_2} = 0.1 * 0.2 * 0.5 * 0.2 = 0.002$$

$$\frac{dz_4}{dz_2} = 0.1 * 0.2 * 0.5 * (-0.3) = -0.003$$

$$\frac{dL}{dw_3} = -2 * 0.25 * 0.1 * 0.1 = -0.005$$

$$\frac{dL}{dw_2} = -2 * 0.25 * 0.1 * 0.2 * 0.5 * 0.2 = -0.001$$

$$\frac{dL}{dw_1} = -2 * 0.25 * 0.1 * 0.2 * 0.5 * (-0.3) * 0.6 * 0.25 = 0.000225$$