CS5242 August 2023, Assignment 1

Due week 6, 19 September 2023, 1800hours

Question 1

1. The MNIST data set has 60,000 training samples and 10,000 validations. MNIST is grayscale with 28x28x1 pixels. Suppose we consider MNIST as binary images, only black and white pixels after thresholding. How many mathematically possible binary images are there for a 28x28x1 image? Let k be the number you calculated. What is $(60,000 + 10,000)/k$?

2. Suppose the average mobile phone has 1000 images, and there are 8 billion mobile phones worldwide. How many images are there on mobile phones in the world? We suppose that this number is roughly equal to all the pictures there are in the whole world. Let n be the number you calculated. Compute the ratio n/k.

3. Now consider that mobile phone images are much larger than 28x28x1. Say they are 1000x1000x3 binary pixels. How many mathematically possible images are there? Let this number be m. Calculate the ratio n/m.

4. What is the significance of the ratios computed? What does it mean when the ratio is small/large? How does this relate to the curse of dimensionality?

Question 2

Consider the function, $f : [-5, 5] \mapsto \mathbb{R}$,

$$f(x) = \sum_{j,k=-3}^{3} c_{jk}\psi(2^j(x-k)) \tag{1}$$

With unknown c_{jk} and ψ being the "bump" function,

$$\psi(x) = (1 - \frac{1}{1+\exp(-(10x-2))}) * (1 - \frac{1}{1+\exp(10x+2)}) \tag{2}$$

1. Given the dataset containing pairs of inputs x (2-1-x.txt) and outputs $f(x)$ (2-1-f_x.txt), determine c_{jk}. This produces Figure 1 (left).

2. Given the dataset containing pairs of inputs x (2-2-x.txt) and outputs $f(x)$ (2-2-f_x.txt), determine c_{jk}. This produces Figure 1 (right).

3. How do the parameters j, k, and c impact the bump function?

Question 3

For the training data points, use the following 4 points:
$[0.1, 0.19], [0.2, 0.18], [0.3, 0.36], [0.4, 0.33]$. Plot these 4 data points and use your eyes to look at their trend. Do they fall approximately in a linear curve?

Perform polynomial fit on the training data for the polynomial degree of $1, 2, 4, 8, 16, 32, 64$ using mean square error loss on the 4 data points,

$$L = \sum_{i=1}^{4} (\hat{f}(x_i) - y_i)^2 \tag{3}$$

$\hat{f}$ is the polynomial fit result. For each polynomial fit, repeat the fitting 10 times to get 10 models for each polynomial degree. Therefore you will get a total of 10×7 models. For each model, compute the following test loss,

$$L = \int_{-5}^{5} (\hat{f}(x) - g(x))^2 dx \tag{4}$$

where $g(x) = 0.6x + 0.1$. You will need to do numerical integration with bin size $\Delta x = 0.01$ and loss

$$L = \sum_{i=0}^{1000} (\hat{f}(i\Delta x - 5) - g(i\Delta x - 5))^2 \Delta x \tag{5}$$

1. Make a plot of training loss versus the degree of polynomial. What do you observe? Explain the observation.

2. Make a plot of testing loss versus the degree of polynomial. Include the standard deviation of the test loss in the graph. What do you observe? Explain the observation.

Quesiton 4

A series of compute trees are given. Make the code that generates these trees. For checking, you may use the following code

```
from torchviz import make_dot
def print_compute_tree(name,node):
    dot = make_dot(node)
    dot.render(name)
```

For the operations that require a constant, use the value of 5.

1. Make code that generates Figure 2. Additionally, when both the inputs are $[1, 2, 3]$, the output is $[20.8, 161.6, 542.4]$.

2. Make code that generates Figure 3. Additionally, when both the inputs are $[1, 2, 3]$, the output is $[0.0035, 0.0026, 0.0021]$.

3. Make code that generates Figure 4. Additionally, when both the inputs are $[1, 2, 3]$, the output is $[-104, -1188, -6468]$.

Submit the output of the operations when all the inputs are $[4, 5, 6]$.

Quesiton 5

Consider the network shown in Fig.5. Given a data $x \in \mathbb{R}$ and label $y \in \mathbb{R}$. And the activations are given by

$$\begin{aligned} z_1 &= w_1 x + b_1 && (6) \\ a_1 &= \sigma(z_1) && (7) \\ z_2 &= w_2 a_1 + b_2 && (8) \\ a_2 &= \sigma(z_2) && (9) \\ z_3 &= w_3 a_2 + b_3 && (10) \\ a_3 &= \sigma(z_3) && (11) \end{aligned}$$

Using mean square error to minimize loss,

$$L = (a_3 - y)^2 \tag{12}$$

Compute the following in symbolic form in terms of a, y, and x,

$$\frac{\partial L}{\partial w_i} \quad \text{for} \quad i = 1, 2, 3 \tag{13}$$

$$\frac{\partial L}{\partial b_i} \quad \text{for} \quad i = 1, 2, 3 \tag{14}$$

Now let $x = 1.0$, $w_1 = 0.1$, $b_1 = 0$, $w_2 = -0.2$, $b_2 = 0.1$, $w_3 = -0.1$, $b_3 = 0.2$, $y = 1$. Let σ be the sigmoid function.

1. Compute a_i, for $i = 1, 2, 3$.

2. Compute L.

3. Compute the gradients $\partial L/\partial w_i$ and $\partial L/\partial b_i$, $i = 1, 2, 3$

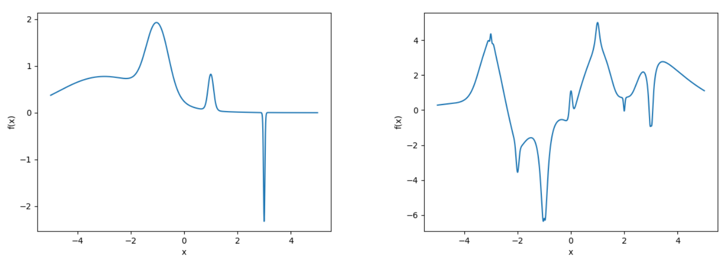

Figure 1: Function output for Question 2.1 (left) and Question 2.2 (right)

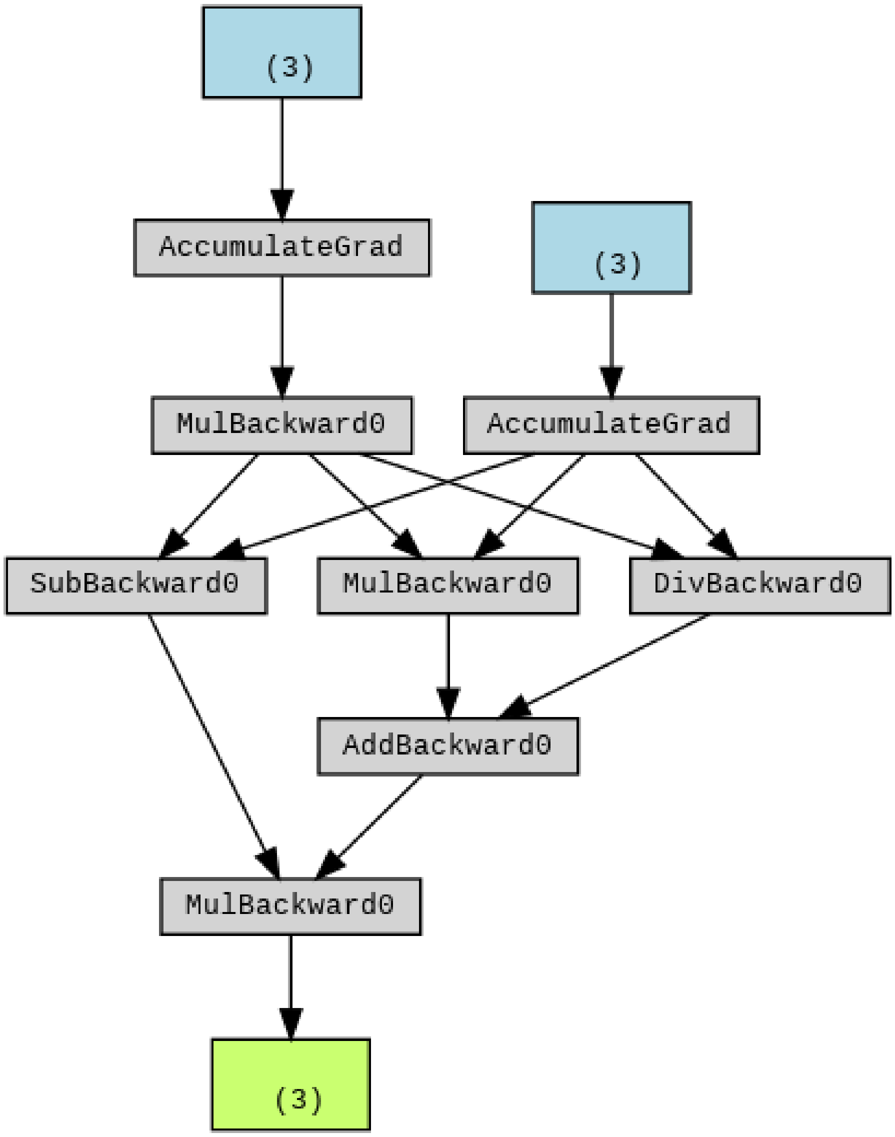

Figure 2: Compute tree for Question 4.1

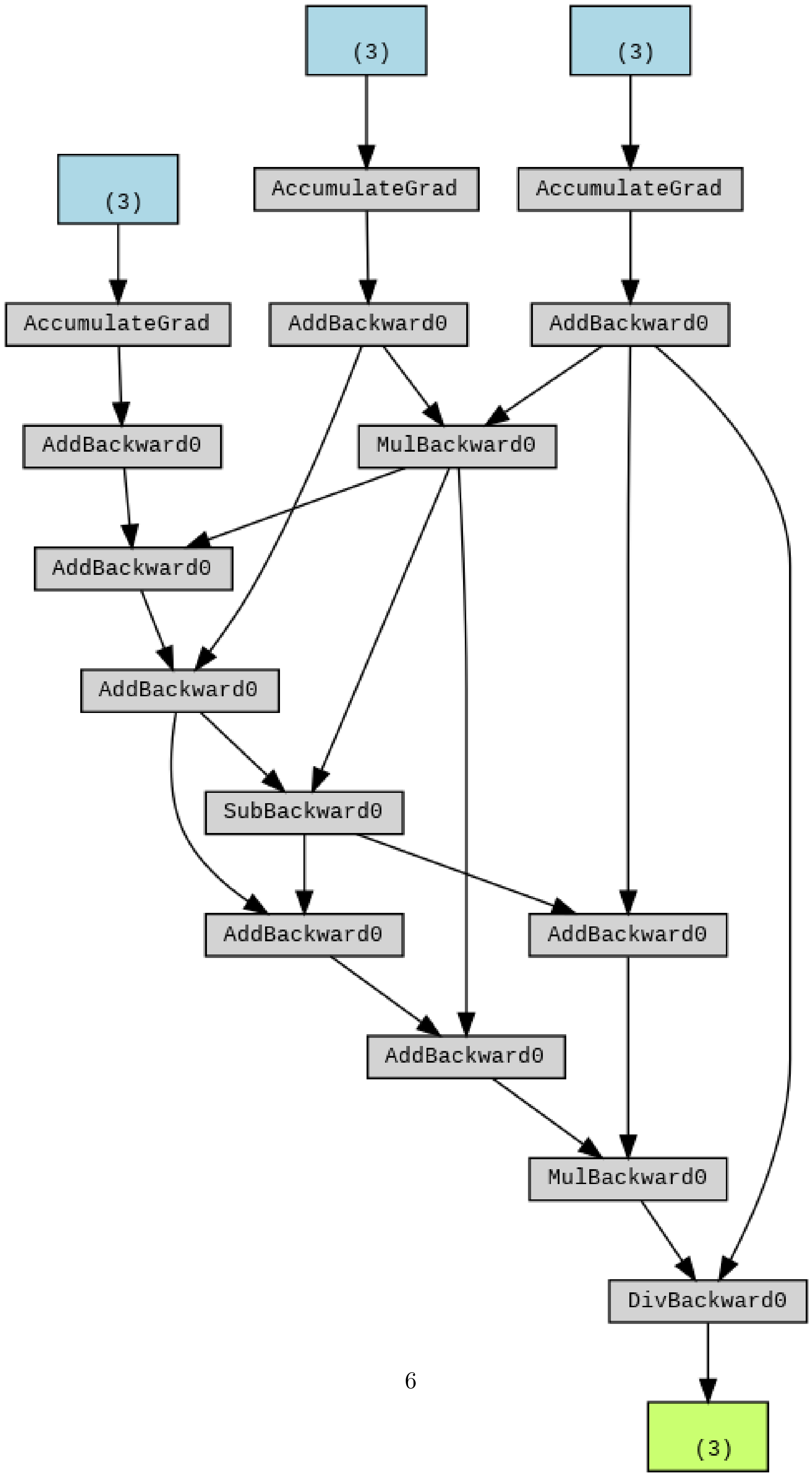

Figure 3: Compute tree for Question 4.2

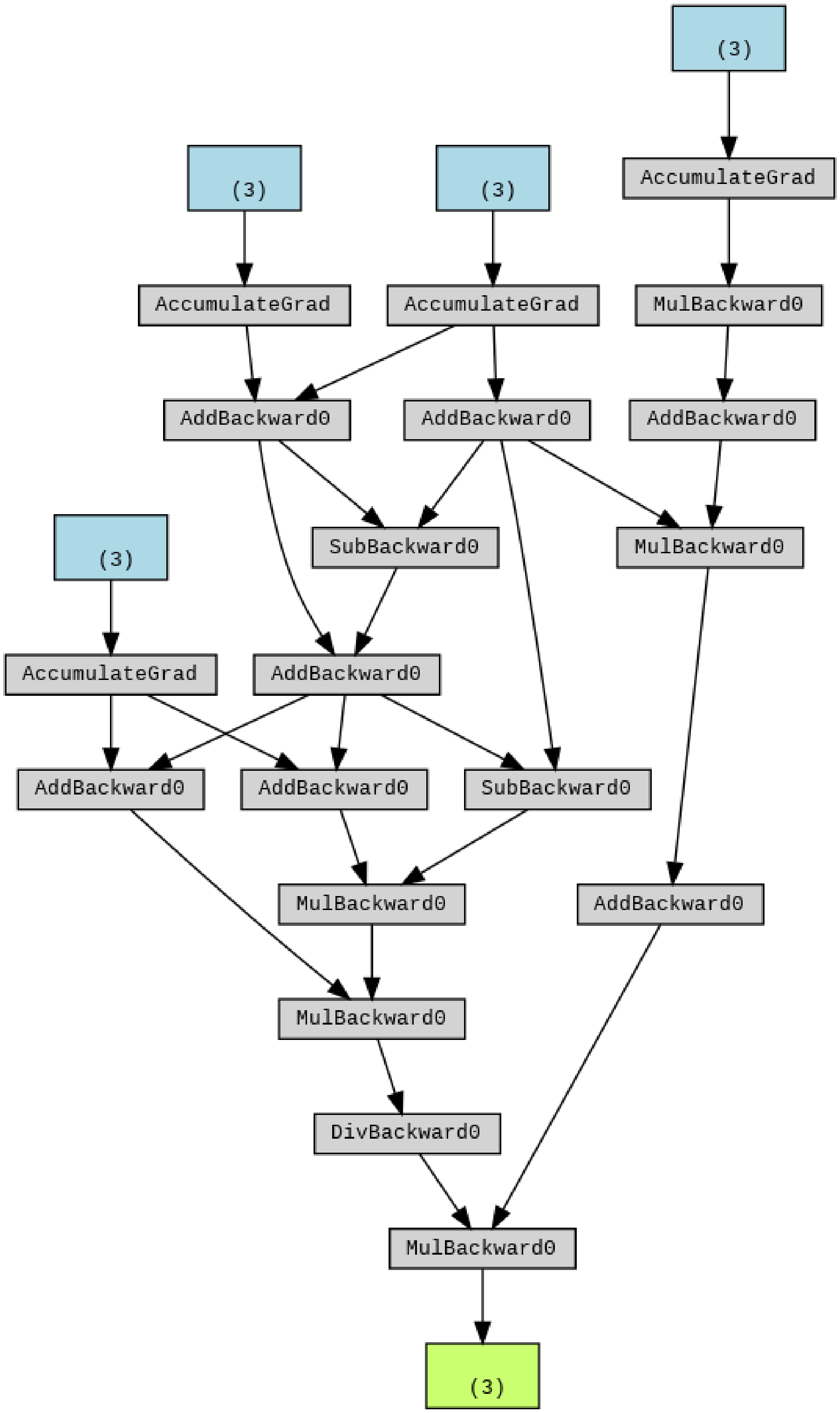

Figure 4: Compute tree for Question 4.3

Figure 5: Network for Question 5