Quiz Submissions - Quiz 10

×

Yuelin Liu (username: yuelin.liu@mail.mcgill.ca)

Attempt 1

Written: Dec 7, 2020 8:20 AM - Dec 7, 2020 1:44 PM

Submission View

Released: Dec 9, 2020 1:00 AM

Gradient Computation and AD

Question 1 1 / 1 point

Choose the correct statement about gradient computation from the options given below

- ◯ Numerical differentiation is generally faster and more accurate than automated algorithmic differentiation.
- ✔ ◯ Numerical differentiation is useful to compute derivatives of black box cost functions.
- ◯ Automated algorithmic differentiation is generally used to check the correctness of gradient function from Numerical differentiation.
- ◯ All of the above

Question 2 1 / 1 point

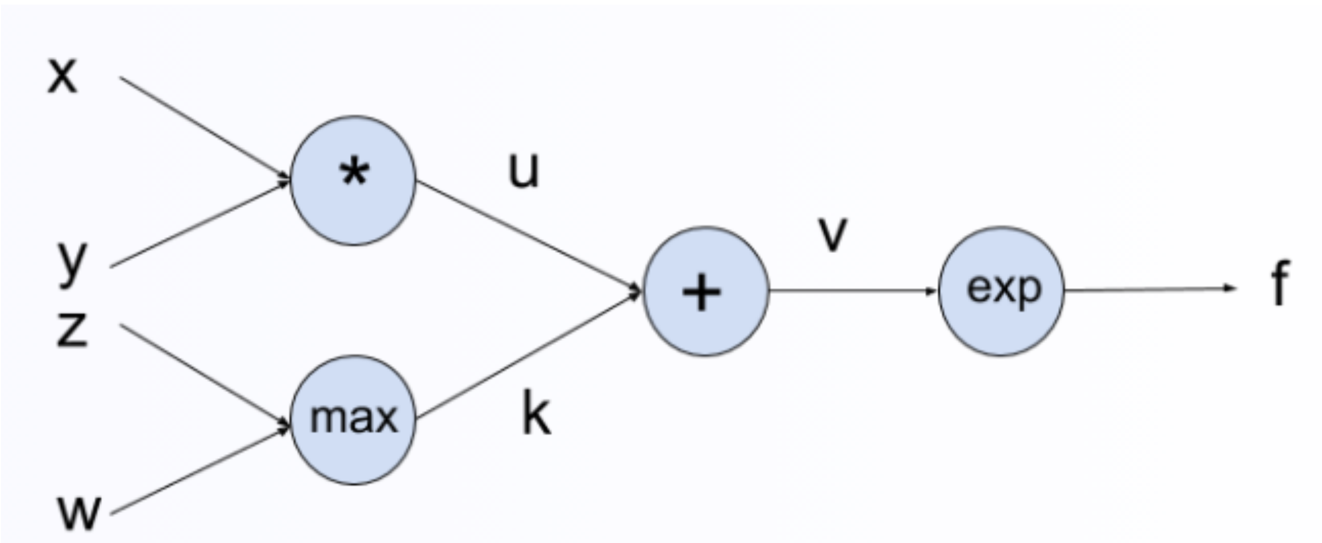
In backpropagation the z=max(x,y) node in a computation graph distributes the incoming gradients in backward pass to paths from x and y.

- ○ True
- ✓ ○ False

Question 3

1 / 1 point

Consider the computation graph given below. x, y and z are input to the graph to compute f. (Note that exp denotes the exponential function)



Suppose you are given that x=1, y=-2, z=1 and w=2. After the forward pass and backpropagation what are the values of:

$$\frac{\partial f}{\partial u}, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial w}, \frac{\partial f}{\partial z}$$

Answer for blank # 1: 1 ✓(25 %)

Answer for blank # 2: -2 ✓(25 %)

Answer for blank # 3: 1 ✓(25 %)

Answer for blank # 4: 0 ✓(25 %)

▼ Hide Feedback

$$x = 1, y = -2, z = 1, w = 2$$
$$u = xy = -2$$
$$k = \max(z, w) = w = 2$$
$$v = u + k = 0$$
$$f = \exp(v) = 1$$
$$\frac{\partial f}{\partial u} = \frac{\partial f}{\partial v}\frac{\partial v}{\partial u} = \exp(v) = 1$$
$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial v}\frac{\partial v}{\partial u}\frac{\partial u}{\partial x} = \exp(v)y = -2$$
$$\frac{\partial f}{\partial w} = \frac{\partial f}{\partial v}\frac{\partial v}{\partial k}\frac{\partial k}{\partial w} = \exp(v) = 1$$
$$\frac{\partial f}{\partial z} = 0$$

Question 4 — 2.5 / 3 points

Consider a two layered multi layered perceptron model given by:

$$y = \sigma(V ReLu(u)),$$
$$u = Wx,$$
$$\sigma(x) = \frac{1}{1 + e^{-x}},$$
$$ReLu(x) = \max(0, x)$$
$$L = |y - y_T|$$
$$x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}$$
$$W = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$$
$$V = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$$

The output of this network is clearly a scalar value between 0 to 1. The loss function is defined by L where y_T is the target output.

Now consider the following inputs, weights and target output.

$$x = \begin{bmatrix} 2 & -1 \end{bmatrix}$$
$$W = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}$$
$$V = \begin{bmatrix} 1 & 2 \end{bmatrix}$$
$$y_T = 0$$

You need to compute the following: (Note if the answer is not an integer then round it to one decimal place eg. 0.387 -> 0.4)

1. $\frac{\partial L}{\partial w_{11}}$
2. $\frac{\partial L}{\partial w_{12}}$
3. $\frac{\partial L}{\partial w_{21}}$
4. $\frac{\partial L}{\partial w_{22}}$
5. $\frac{\partial L}{\partial v_1}$
6. $\frac{\partial L}{\partial v_2}$

Answer for blank # 1: 0.4 ✔(16.67 %)

Answer for blank # 2: 0 ✖ (-0.2)

Answer for blank # 3: 0 ✔(16.67 %)

Answer for blank # 4: 0 ✔(16.67 %)

Answer for blank # 5: 0.2 ✔(16.67 %)

Answer for blank # 6: 0 ✔(16.67 %)

▼ Hide Feedback

$$y = \sigma(z),$$

$$z = V ReLu(u) = \begin{bmatrix} 1 & 2 \end{bmatrix} ReLu(\begin{bmatrix} 1 \\ -2 \end{bmatrix}) = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1$$

$$u = Wx = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial L}{\partial y} \frac{\partial y}{\partial z} \frac{\partial z}{\partial u} \frac{\partial u}{\partial w_{11}}$$

$$\frac{\partial L}{\partial y} = 1$$

$$\frac{\partial y}{\partial z} = \sigma(z)(1-\sigma(z)) = \frac{e}{e+1}(1-\frac{e}{e+1}) = \frac{e}{(e+1)^2} = 0.2$$

$$\frac{\partial z}{\partial u} = V\begin{bmatrix}1\\0\end{bmatrix} = 1$$

$$\frac{\partial u}{\partial w_{11}} = \nabla_{w_{11}}\Big(\begin{bmatrix}w_{11} & w_{12}\\ w_{21} & w_{22}\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}\Big) = \nabla_{w_{11}}\Big(\begin{bmatrix}w_{11}x_1 + w_{12}x_2\\ w_{21}x_1 + w_{22}x_2\end{bmatrix}\Big) = \begin{bmatrix}x_1\\0\end{bmatrix}$$

$$\frac{\partial L}{\partial w_{11}} = 0.2x_1 = 0.4$$

$$\frac{\partial L}{\partial w_{12}} = 0.2x_2 = -0.2$$

$$\frac{\partial L}{\partial w_{21}} = \frac{\partial L}{\partial w_{22}} = 0(\text{ because } u_2 = 0)$$

$$\frac{\partial L}{\partial v_1} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial z}\frac{\partial z}{\partial v_1} = 0.2\begin{bmatrix}\nabla v_1 & 0\end{bmatrix} ReLu(u) = 0.2\begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} = 0.2$$

$$\frac{\partial L}{\partial v_2} = \frac{\partial L}{\partial y}\frac{\partial y}{\partial z}\frac{\partial z}{\partial v_2} = 0.2\begin{bmatrix}0 & \nabla v_2\end{bmatrix} ReLu(u) = 0.2\begin{bmatrix}0 & 1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} = 0$$

Question 5 1 / 1 point

Consider a function

$$f : \mathbb{R}^M \to \mathbb{R}^N,\ \text{where } M >> N$$

Suppose we are trying to compute the Jacobian, then reverse mode of automatic differentiation is more efficient.

- [x] True
- [ ] False

CNN

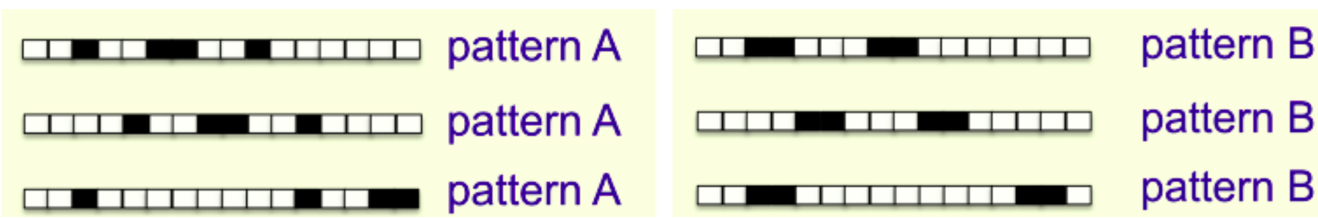
Question 6 — 1 / 1 point

For all neural networks, setting the weights initialization to 0 when optimizing using stochastic gradient descent method is acceptable?

- [ ] True
- [x] False

Question 7 — 0 / 1 point

Consider the following binary classification problem which is impossible to solve using linear classifiers (classes are not linearly separable).



Now lets say the training data consists of all possible translation of patterns A and B. Consider a model with a single 1D convolution layer (without any activation function) followed by a fully connected layer with logistic sigmoid activation. Answer True/False. **This CNN would perfectly classify all the training data.**

- [ ] True ✗
- [ ] False ➡

Hide Feedback

Because convolution layers are linear and composition of linear layers results into linear. And we know that problem is not linearly separable. Therefore, not solvable by the above CNN.

Question 8 1 / 1 point

Consider a binary classification task as dog vs not-dog problem. We use a CNN with a single output neuron and let it be “z”. Let the final output be y described as:

$$y = \sigma(ReLu(z))$$

We classify input with final value y>=0.5 as dog image. Answer Yes/No. The above network would classify all the images as a dog image?

✔ ◯ Yes

◯ No

▼ Hide Feedback

ReLu would output either 0 or positive value. Therefore, sigmoid function would produce a value which is always greater or equal than 0.5 leading to all images with dog labels.

Question 9 0.75 / 1 point

Given the input and the kernel,

$$\text{INPUT} = \begin{bmatrix} 3 & 3 & 2 & 1 \\ 0 & 0 & 1 & 1 \\ 3 & 1 & 2 & 2 \end{bmatrix} \quad \text{KERNEL} = \begin{bmatrix} 0 & 1 \\ 2 & 2 \end{bmatrix}$$

consider 2D convolution without kernel flipping. The output is as follows:

$$\text{OUTPUT} = \begin{bmatrix} o_{11} & 4 & o_{13} \\ o_{21} & o_{22} & 11 \end{bmatrix}$$

What are the values for:

1. o_{11}
2. o_{13}
3. o_{21}
4. o_{22}

Answer for blank # 1: 3 ✔(25 %)

Answer for blank # 2: 5 ✖ **(9)**

Answer for blank # 3: 8 ✔(25 %)

Answer for blank # 4: 7 ✔(25 %)

▼ Hide Feedback

$$\text{OUTPUT} = \begin{bmatrix} 3 & 4 & 9 \\ 8 & 7 & 11 \end{bmatrix}$$

O_{11} = 3*0 + 3*1 + 0*2 + 0*2 = 3

Question 10

1 / 1 point

You have input of dimension 34 X 34. After convolving the input with 6 X 6 kernel with 0 padding and stride 2. The dimension of resulting volume is M x M. What is M?

Answer: 15 ✔

Hide Feedback

Output length: (Input - K)/stride +1 = (34 - 6) /2 +1

Attempt Score: 10.25 / 12 - 85.42 %

Overall Grade (highest attempt): 10.25 / 12 - 85.42 %

Done