# S4851/6851 IDL: Homework 3 March 14, 2022

Note: All coding problems to be submitted with GitHub Link. Do not Upload the files/folder. Use git commands only.

Note: this is the distribution of questions:

1. Question 1 to Question 2: Required for everyone.
2. Question 3 to Question 4: Bonus question for both Graduate Students and Undergraduate Students

# Problem 1 (30 points)

Object detection and Object Segmentation are two very important tasks in computer vision. There are many real-world applications of these tasks. Take one of these tasks and explain how you would implement it. For example, Image segmentation is used in medical imaging.

Image Segmentation and autonomous driving. The process of image segmentation breaks up a large image into their individual components. For a self-driving car to navigate it will need to break up the entire front-view of the road into individual pieces and label them accordingly. Start by separating the pedestrians as non-drivers so they are avoided, all the signs that appear when one is on the road: stop sign, yield sign, crosswalk, etc., the lanes that differentiate the flow of traffic: two lanes vs one lane, crossing the marker between traffic coming and going, the lanes that correspond to the turning lanes on the perpendicular side of the road.

# Problem 2 (30 points)

how does YOLO (You only Look once) works? What is the difference between YOLO and Faster-RCNN?

YOLO is a real time object detection algorithm that runs at 45 fps, while the faster one runs with over 150 fps. It splits the input image into an S x S grid, if the center of the object falls into a grid cell, that grid cell is responsible for detecting that object. It uses 24 convolutional layers followed by 2 fully connected layers. We use 1 x 1 reduction layers followed by 3 x 3 convolutional layers. It predicts multiple bounding boxes per grid cell. During training we only want one bounding box predictor to be responsible for each object. We assign one predictor based on the which one has the highest IOU with the ground truth.

YOLO is better when it comes to detecting small objects, the speed of YOLO is faster than Faster-RCNN, results are cleaner with YOLO than Faster-RCNN with little to no overlapping boxes.

Bonus for both undergraduates and graduates beyond this line.

# Problem 3 (40 points)

In classification problem, cross entropy typically exhibits a faster convergence time than the alternatives. Consider a binary classification problem with cross binary as a loss. Note, in this case the labels are *y* 0*,* 1 and network output of a single logistic sigmoid unit represents the conditional distribution of *k* = 1 class given data *P* (*k* = 1 |*x*). Assume the data is i.i.d and write down the negative log likelihood that corresponds to the case when data can be mislabeled with probability *ϵ*. Make sure that your function reduces to the convolutional cross entropy when *ϵ* = 0. Discuss the robustness properties if any of your new loss.

*∈*

*|*

# Problem 4 (40 points)

Consider the computational graph in [1](#_bookmark0). Note, that this is a simplified version of **ResNet** that we studied in class. Write down reverse mode trace for this

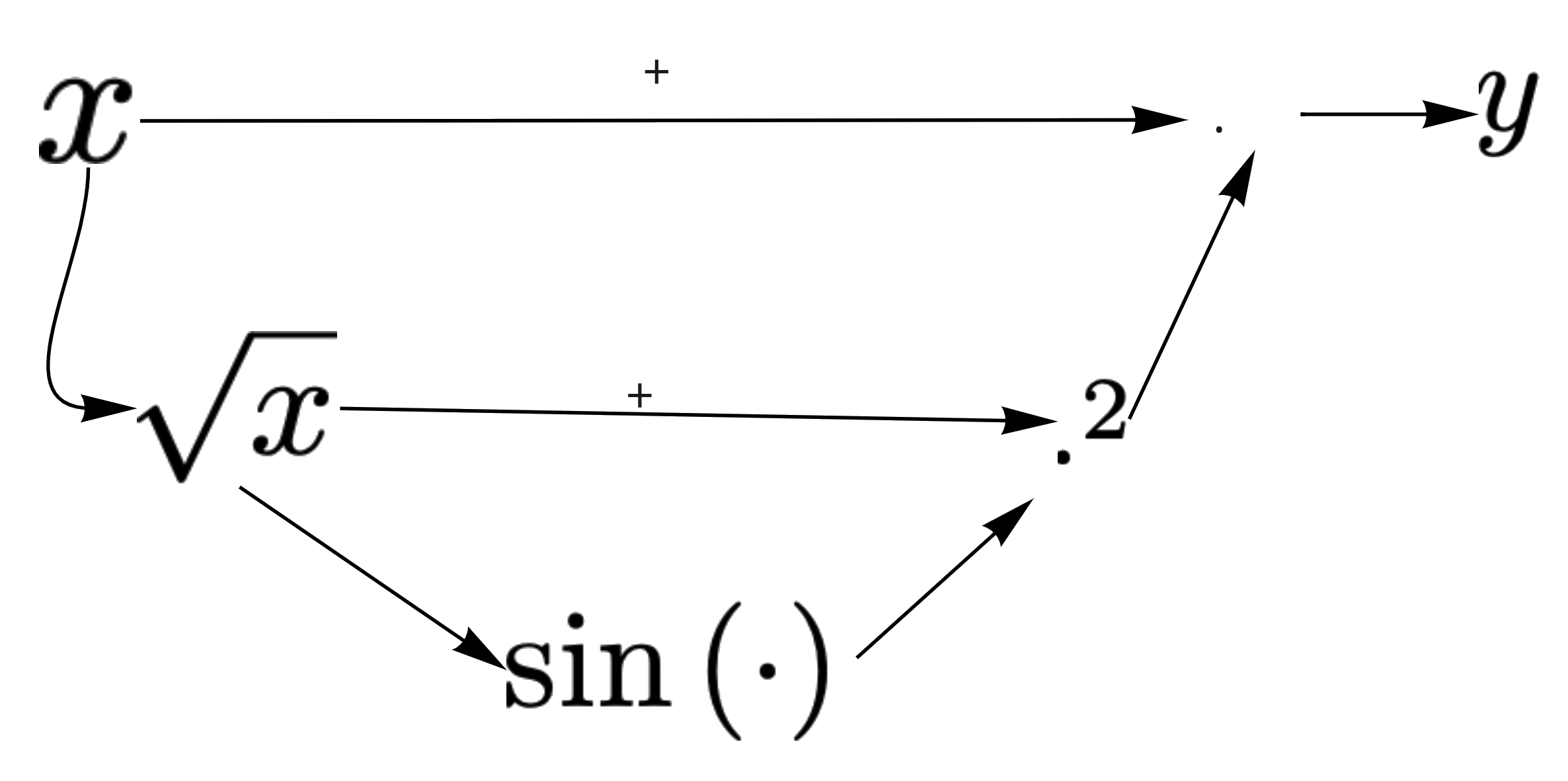


Figure 1: Resnet

computation without evaluating it at a point to find dy/dx. Describe your ob- servation of the most pronounced property of the final expression. Evaluate the gradient at x = 0. What would happen to the gradient if there was no skip connection from x to y. This exercise is to understand how gradients flow in a skip connection network and how we can encounter vanishing gradient issues if any.