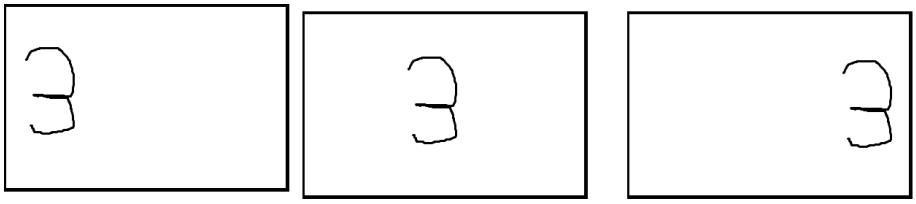
Homework 2: Multi-Layer & Convolutional Neural Networks

CS 1470/2470

Due October 05, 2020 at 11:59pm AoE

1 Conceptual Questions

1. Consider the three following 23x23 images of the digit 3.



Which neural net is better suited for identifying the digit in each image: a convolutional neural net or a feed-forward neural network? Explain your reasoning. (2-3 sentences)

2. In lecture, you were introduced to the concepts of *translational invariance* and *translational equivariance*.

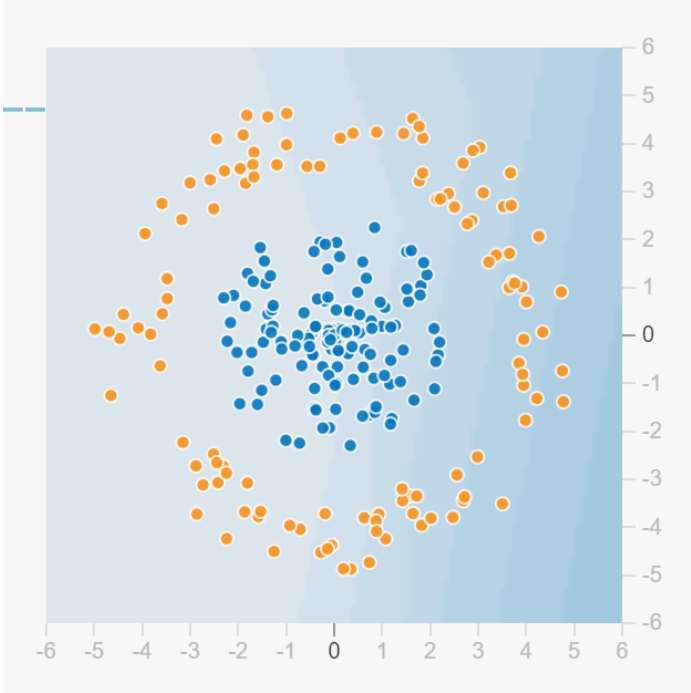
   (a) What is the difference between these two concepts? *(We expect to see some function notation and text.)*

   (b) What part(s) of a convolutional neural network (CNN) with max-pooling, if any, are translationally equivariant, and why? (1-2 sentences)

   (c) What part(s) of a CNN with max-pooling, if any, are translationally invariant, and why? (1-2 sentences)

   (d) Consider the images of the digit 3 in the previous question. Will a CNN with standard max-pooling (e.g 2x2 pooling) produce the same or different outputs for all of them? Why/why not? How does this relate to translational invariance/equivariance? (hint: remember that the image is 23x23) (2-4 sentences)

3. Consider the dataset shown in this scatterplot:



The orange points are labeled with class label 0, and the blue points are labeled with class label 1. Write out a mathematical expression in terms of the inputs, using linear layers and ReLUs, that will correctly classify all of these points. *(We expect something like output* $= ....x_1 + ...x_2$*. Hint: Use* https://tinyurl.com/y5gayl5b *.)*

4. (Optional) Have feedback for this assignment? Found something confusing? We'd love to hear from you!

2 Ethical Implications

Suppose you want to build an app designed for visually-impaired people to help them gain more information about what's around them; they can take a picture of something, and your app will tell them what's "in the picture".

1. What type of data would be best to train this model? How can we obtain this data in an ethical and privacy-preserving manner? (3-6 sentences)

2. What workarounds can you implement to overcome the accessibility challenges for visually-impaired users? (i.e. How specifically will your app tell

the user what is in the picture? Does your app account for users with different levels of visual impairment?) For more background on accessible design, please read this article:
https://www.w3.org/WAI/fundamentals/accessibility-usability-inclusion/.
(4 sentences minimum)

3. A given user is visually-impaired and would heavily rely on this technology. With your design, what would happen if a user is depending on your app and it returns an incorrect response? Given this potential issue, should this technology be built? (4 sentences minimum)

3 CS2470-only Questions

1. Prove that convolution is equivariant under translation (it's fine to do this just for 1D convolution).

2. Suppose you have a CNN that begins by taking an input image of size 28x28x3 and passing through a convolution layer that convolves the image using 3 filters of dimensions 2x2x3 with valid padding.

   (a) How many learnable parameters does this convolution layer have?

   (b) Suppose that you instead decided to use a fully connected layer to replicate the behavior of this convolutional layer. How many parameters would that fully connected layer have?

3. Read this paper about cutout: https://arxiv.org/pdf/1708.04552.pdf

   (a) What is cutout? Why is it useful?

   (b) What are some similar methods to cutout? What makes them similar?

   (c) What were the cutout sizes for CIFAR-10 and CIFAR-100? How did the researchers decide on their cutout size? Why do you think the cutout size differed for CIFAR-10 vs CIFAR-100?