Test Person 2021-09-14

Email: testperson@example.com
Class: E1a
Teacher: Professor Teacher

2021-10-19

DLAI (14 September 2021)

Word count: 502

YOUR_NAME = ...

Question 1 (4 points)

How many learnable parameters has a standard self-attention module with 7 heads, working on a 16-long sequence of 10-dimensional vectors?

Question 2 (6 points)

Consider a MLP with a standard mean squared error loss. If you initialize all the network parameters with the same constant value α, do you expect the training process to lead to a desirable state? Do you expect the resulting network to work well on test data, and why?

Question 3 (6 points)

Imagine you trained a deep ReLU network, and by the end of the training, you realize the resulting network has sparse weights. However, this was not happening just a few trainings ago. What has happened? You remember that you changed the following parameters, but don't exactly know which one is responsible for the curious behavior: learning_rate, batch_size, a different weight on L2 regularization. State what you think is the parameter that induced the sparse weights, and explain why.

Question 4 (8 points)
Suppose you have trained a fully convolutional NN on 64x64 images, which produces a single *k*-dimensional vector (with constant *k*) as output. Can you apply this CNN to a larger image at test time? If no, which operation is the one that fails? If yes, what would you obtain as output? And what happens if the test image is instead smaller than 64x64?

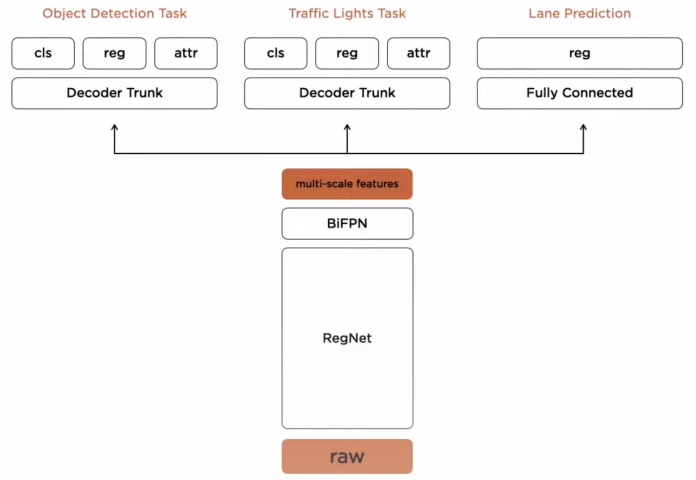
Question 5 (4 points)
Consider the CNN from the previous exercise. Suppose you want to fine-tune it for a classification task where the model should discriminate between 64x64 images of 10 dog breeds. You decide to do so by adding an extra layer, and training only this one. Which of the two approaches would you prefer, and why?

1- You add a 1x1 convolutional layer with 10 filters.

2- You add a fully connected layer with 10 neurons as output.

"Hi YOUR_NAME, and thank you for applying for the role of machine-learning engineer here at Tesla! Let's start with the interview."

Question 6 (6 points)
As you know, we are developing a a deep-learning-based driverless car system using only camera images as input. Such a system requires the solution of many tasks, such as the detection of different objects (cars, pedestrians, ...), the identification of traffic lights, the segmentation of traffic lines, etc. As a general design choice, we had to choose between creating a single model for every task, or a single model with many heads like the one pictured below:



Please discuss what are the pros of the many-heads approach, both during development and once in production.

Question 7 (6 points)

The model I showed you above is a visual model that takes as input single frames, yet many of the predictions we want our driverless model to make can not be assessed from a single frame. Can you make 3-4 examples? As a naive solution we could take as input many consecutive frames, yet this solution is computationally expensive -- can you think of something smarter?