Test Person 2021-10-19

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

DLAI (19 October 2021)

Word count: 333

Question 1 (3 points)

How many learnable parameters has a 3-layer MLP with bias (input → hidden1 → hidden2 → output) with 10-20-10-20 neurons, with batch normalization in the first hidden layer, and dropout (p=0.1) in the second hidden layer?

Question 2 (8 points)

Consider a targeted adversarial attack on an image. First, write down the optimization problem that should be solved to find the adversarial attack. Now, how would you modify the energy, such that it leads to a *sparse* attack that only perturbs a few pixels?

Question 3 (7 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 learning_rate parameter, but how could that cause the sparse weights?

Question 4 (4 points)

What is the difference between translation equivariance and translation invariance? Is convolution translation-equivariant or invariant?

Question 5 (6 points)
Imagine you need to build a graph autoencoder, where the representation at the input and at the output is the adjacency matrix of a graph. How would you define the reconstruction loss? What is the main difficulty we have to deal with, that we don't have when we build an autoencoder for images?

Question 6 (4 points)
What do we mean when we say that convolution is a *linear* operation? Try to be as formal as possible in your answer, using formulas if needed.

Question 7 (8 points)
Today, we often hear the expressions "explainable AI" and "interpretable machine learning". What do you think are the main issues these expressions try to capture? As a possible topic of discussion, imagine you have a tool that allows you to explore and visualize all the learned features of a deep neural network. Would you find it useful? If you had a magic wand, what tool would you like to have to really achieve "explainability" as you mean it?