Deep Structured Learning (IST, Fall 2019)

Homework 3

Instructors: André Martins and Vlad Niculae
TAs: Gonçalo Correia and Ben Peters

Deadline: Friday, November 22, 2019.

Please turn in the answers to the questions below, **in English**, together with the code you implemented to solve them (when applicable). Please email your solutions in **electronic format** (a single zip file) with the subject "Homework 3" to:

`deep-structured-learning-instructors@googlegroups.com`

Hard copies will not be accepted.

Question 1

OCR with a convolutional neural network. In homework 1, you implemented an optical character recognition system, first with independent pixel values as features, then using pairwise pixel values (feature engineering), and then using a multi-layer perceptron. You will now try out convolutional networks. For this exercise, we recommend you use a deep learning framework with automatic differentiation (suggested: Pytorch).

1. (10 points) What kind of invariances can you capture with convolutional layers that you cannot with pairwise features?

2. (30 points) Implement a simple convolutional network with the following structure:

   - A first convolutional layer with 16 channels and filters of size 5x5, stride of 1, and padding chosen to preserve the original image size.
   - A relu activation applied to the end of this layer.
   - Max pooling of size 2x2, with stride of 2.
   - A second convolutional layer with 32 channels and filters of size 7x7, stride of 1, and padding chosen such that the size of the output of this layer is identical to the size of its input (the size of the output of the previous pooling layer).
   - A relu activation applied to the end of this layer.
   - Max pooling of size 2x2, with stride of 2.
   - An affine transformation followed by an output softmax layer.

   Regularize and optimize at your will. Plot the training loss and validation accuracies over epoch number. Report the final test accuracies. Hint: if you're using Pytorch, use the functions `nn.Conv2d` and `nn.MaxPool2d`.

3. (10 points) Randomly pick 3 channels in the first layer and 3 channels in the second layer and plot the correspond filters. What are these representations capturing?

4. (10 points (bonus)) Increase the number of convolutional layers and play a bit with the architecture. Report the results.

Question 2

Sequential OCR with RNNs. In homework 2, you solved the OCR problem with structured prediction, using a linear sequential model. You will now try out a recurrent neural network, more precisely a BILSTM tagger.

1. (25 points) Exploit the sequential structure of the characters (as they form words). Use as input a feedforward layer shared by all characters, followed by a bidirectional LSTM. Then, append an affine transformation followed by an output softmax layer to map this to character predictions. Regularize and optimize at your will. Plot the training loss and validation accuracies over epoch number. Report the final test accuracies. Hint: if you're using Pytorch, use the function `nn.LSTM` for this exercise.

2. (10 points) What are the advantages and disadvantages of this BILSTM compared to the CRF you implemented in the previous class? Can these two approaches be combined? How?

3. (15 points) Replace the feedforward layer above by a convolutional layer similar to the one you implemented in the first exercise. Report the results.