Information Mining - winter semester 2020

Exercise sheet 8

Exercise 1: Dropout

Randomly drop neurons from the network during training
To prevent overfitting

What is Dropout and why do we need it?

Exercise 2: Weight initialisation

What is weight initialisation in neural networks and what techniques do you know? Why is zero initialisation not a recommended technique?

set weight at beginning. random initialization ,Xavier initialization.
they will make the exact same decisions!

Exercise 3: Hyperparameters

What are hyperparameters, provide some examples.

parameter for model itself. e.g. k-nn : k .

Exercise 4: Embedding

What is a sparse and what is a dense vector? Can you motivate the use of embeddings.

sparse vector has many zeros. for dense,it is not.
embeddings : Representation of words in continuous space , it is dense vector.

Exercise 5: RNN

What is the motivation of RNNs?

work for sequential information.earlier input should have influence on current output.

Exercise 6: LSTM

it has a longer memory about earlier input.
can be used for longer sentences. e.g. deutsch.

What is the advantage of LSTM over simple RNNs?

Exercise 7: Deeper understanding questions

- Assume you have been using a neural network with the ReLU activation function within the hidden layers. Now you are replacing it with a linear activation function. Note a linear activation function passes the inputs as they are without doing anything on them. Would this new neural network be able to approximate an XNOR function? Why? no.it is a linear model .
- Given an n-character word, we want to predict which character would be the n+1th character in the sequence. For example, our input is "predictio" (which is a 9 character word) and we have to predict what would be the 10th character. Which neural network architecture from below would be suitable to complete this task?
  - A) Fully-Connected Neural Network with two hidden layers
  - B) Fully-Connected Neural Network with 10 hidden layers

- C) Recurrent Neural Network
- D) LSTMs D
- E) Fully-Connected Neural Network with 3 hidden layers and various Dropouts

Exercise 8: CNN basics

- What is the motivation of CNNs?
- What is a filter or a kernel? we slide filter over the entire image
- What is an activation map? the result we use filter to get.
- What is a stride? the number of pixels to shift at each step

Exercise 9: CNN - Paddings

to maintain image size in the map
valid : only padding when filter cant be used.
same : to get same size with map
full : to get largest map size.

- What do you understand under padding in terms of image processing? Why do you we use this? What is valid, same and full padding?
- Assume you have an 5x5 image. You apply a 3x3 kernel with stride 1. You decide for same padding. What is the output of the resulting feature map? 5x5
- You repeat the above sliding with the valid padding, what is now the size of the resulting feature map? 3x3 2x2 ???

Exercise 10: CNN Filters

- What is the shape of a CNN filter for images? What is the shape of a CNN filter for text? 2-D , 1-D . ?????
- The input image has been converted into a matrix of size 28 X 28 and a kernel/filter of size 7 X 7 with a stride of 1. What will be the size of the convoluted matrix? Note, no padding has been applied.

22x22

Exercise 11: Pooling

- What is pooling? What is the motivation/role of pooling? What techniques to you know for pooling? subsampling. max-pooling/mean-pooling
- Suppose an input to Max-Pooling layer is given. The pooling size of neurons in the layer is (3, 3). What is the output when the following input is given:

9

Table 1:

3	3	5
4	2	6
8	9	7