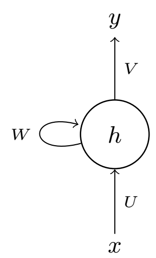
Ludwig-Maximilians-Universitaet Muenchen 28.05.2021
Institute for Informatics
Prof. Dr. Volker Tresp
Rajat Koner
Andreas Lohrer
Sebastian Schmoll

Machine Learning
Summer 2021
Exercise Sheet 6

Exercise 6-1 RNNs

Consider the following RNN in which each hidden state h_t is given by $h_t = \sigma(Wh_{t-1} + Ux_t)$ where $\sigma(z) = \frac{1}{1+e^{-z}}$, $h_t, h_{t-1}, x_t \in \mathbb{R}^d$, and $W, U \in \mathbb{R}^{d \times d}$. Assume that $h_0 = 0$ is given as initial hidden state.



(a) Draw the unfolded RNN for three timesteps $t = 1..3$. Do not forget to label the nodes and edges.

(b) What is the main advantage of RNNs compared to simple feed forward neural networks? What are they commonly used for?

(c) Explain the idea of Encoder-Decoder Networks!

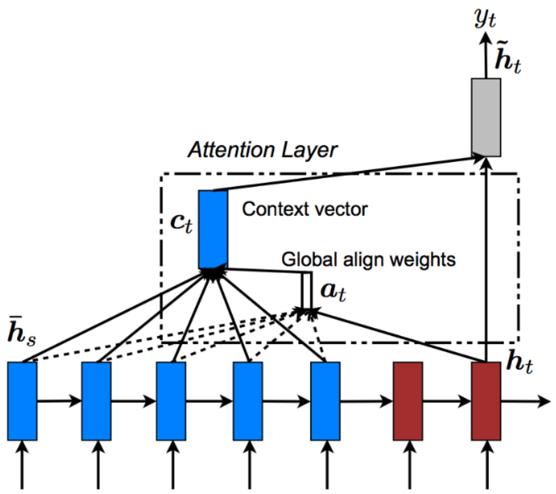
Exercise 6-2 Attention

The Encoder-Decoder model has found to be limited when it comes to very long sequences:

"A potential issue with this encoder-decoder approach is that a neural network needs to be able to compress all the necessary information of a source sentence into a fixed-length vector. This may make it difficult for the neural network to cope with long sentences, especially those that are longer than the sentences in the training corpus."(Neural Machine Translation by Jointly Learning to Align and Translate, 2015.)

(a) How does the attention mechanism overcome this problem?

(b) Consider an Encoder-Decoder model with global attention as shown in the figure below. The blue hidden units correspond to the encoder and the red ones to the Decoder.



The depiction of Global Attention in an Encoder-Decoder Recurrent Neural Network.
Taken from "Effective Approaches to Attention-based Neural Machine Translation."

Assume we want to translate the sentence $x =$ "This is German" into "Dies ist deutsch".

Let the hidden units corresponding to the three input words in the Encoder be given as:

$\bar{h}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$, $\bar{h}_2 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$ and $\bar{h}_3 = \begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix}$. Moreover, let the current target state of the Decoder corresponding to the output word "deutsch" be $h_t = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$.

- Calculate the attention scores (alignments) $a_t(\bar{h}_i)$ for $i = 1..3$ using the formula introduced in the lecture. How can the result be interpreted?
- Calculate the resulting context vector c_t as weighted sum of the hidden units in the Encoder.

Exercise 6-3 PyTorch - Early Stopping and Save/Load Models

On the course website you find an ipython notebook leading you through the former implementation of a CNN in PyTorch. We will extend this model by early stopping and learn how we can save/store the model's parameters.

Exercise 6-4 Autoencoders

(a) What is the main idea of an autoencoder? What components does it consist of and how is it trained?

(b) Name at least three possible applications of an autoencoder.

(c) Consider a linear autoencoder (without regularization or denoising). What happens if the number of dimensions of the latent space is set larger than the dimensionality of the data space?

Exercise 6-5 Generative Adversarial Networks (GANs)

(a) Provide a short and intuitive, non-formal description of a GAN. What is the main idea? What are the individual components of the model and how do they interact?

(b) What are possible applications for a GAN?

(c) Name at least two different variations of the GAN model and provide a possible application for each.