

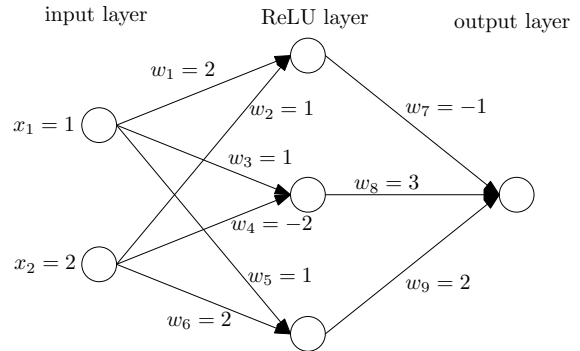
# Deep Learning – NPFL114 – Exam

Write your answers to the following questions. Each question is worth 20 points and the required number of points (including the maximum of 40 surplus points from the practicals) to obtain grades 1, 2, 3 are 85, 70 and 55, respectively.

You can get a hint(s) for any question, but you will then get less points for such questions.

## 1) Training Neural Network

Assume the artificial neural network on the right, with mean square error loss and gold output of 3. Compute the values of all weights  $w_i$  after performing an SGD update with learning rate 0.1.



## 2) Adagrad

Write down the AdaGrad algorithm and show that it internally decays learning rate by a factor of  $1/\sqrt{t}$  in step  $t$ . Compare this to the RMSProp algorithm.

## 3) Convolution

Write down equations of how convolution of a given image is computed. Assume the input is an image  $I$  of size  $H \times W$  with  $C$  channels, the kernel  $K$  has size  $N \times M$ , the stride is  $T \times S$ , the operation performed is in fact cross-correlation (as usual in convolutional neural networks) and that  $O$  output channels are computed. Explain both *SAME* and *VALID* padding schemes and write down output size of the operation for both these padding schemes.

## 4) Neural Machine Translation and BPE

Draw/write how an encoder-decoder architecture is used for machine translation, both during training and during inference, including attention. Furthermore, elaborate on how subword units are used to reduce out-of-vocabulary problem and sketch BPE algorithm for constructing fixed number of subword units.

## 5) Variational Autoencoders

Describe deep generative modelling using variational autoencoders – show VAE architecture, devise training algorithm, write training loss, and propose sampling procedure.