UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Exam:	**INF 5860 / INF 9860 – Machine learning for image analysis**
Date:	**Thursday June 8, 2017**
Exam hours:	**9.00-13.00 (4 hours)**
Number of pages:	**5 pages**
Enclosures:	**None**
Allowed aid:	**Calculator**

- Read the entire exercise text before you start solving the exercises. Please check that the exam paper is complete. If you lack information in the exam text or think that some information is missing, you may make your own assumptions, as long as they are not contradictory to the "spirit" of the exercise. In such a case, you should make it clear what assumptions you have made.

- You should spend your time in such a manner that you get to answer all exercises shortly. If you get stuck on one question, move on to the next question.

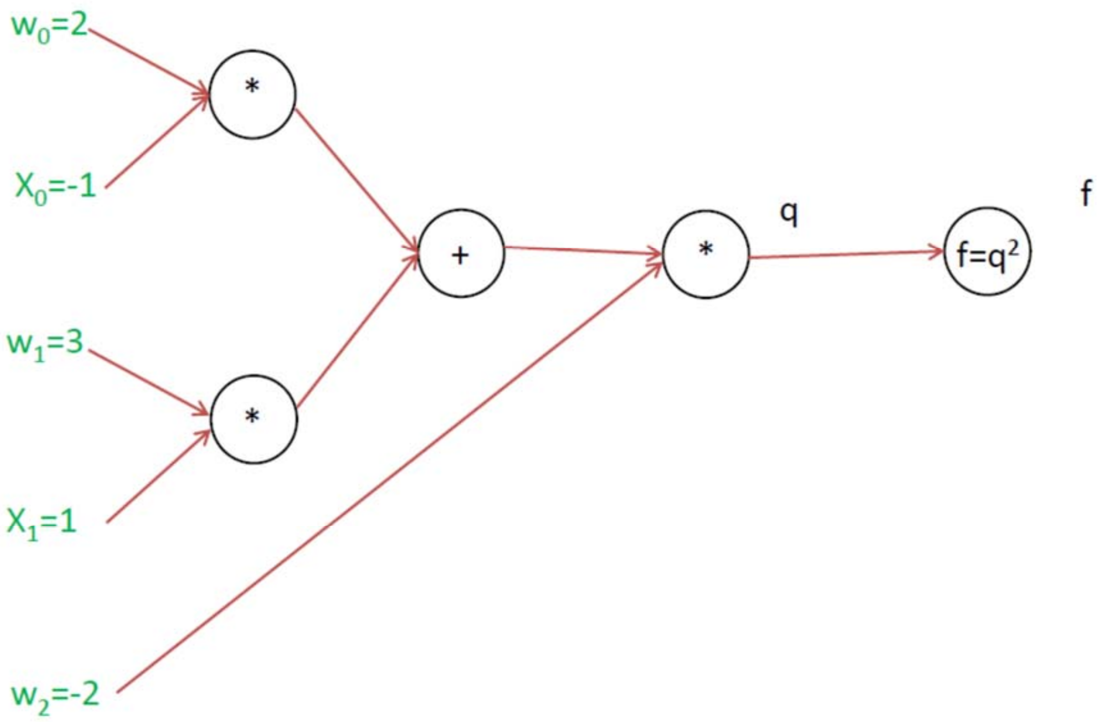
- Each task gives 1 point unless otherwise stated in the exercise text.

- PhD students should solve all exercises, including the one marked PhD only.

- Your answers should be **short**, typically a few sentences and / or a sketch should be sufficient.

Good luck!!

Exercise 1: Backpropagation



a) Perform backpropagation on this example and compute one value for each arrow. Make a sketch of the net and draw you answers on the corresponding arrow. (2 points)

Exercise 2: Basic networks

a) Assume that all weights in a net are initialized to the same random number. Discuss briefly if this is a robust initialization.

b) Discuss briefly if the ReLU activation function has some shortcomings.

c) Explain briefly the main principle with dropout, and what is it used for.

d) Given a cost function $J(\theta)$. Explain how gradient checking is performed for one parameter θ.

e) Explain how gradient descent with momentum updates works.

f) Why should the range or standard deviations of weights in a neural network depend on the size of the input to a layer?

Exercise 3: Convolutional neural nets

You have a convolutional neural network (CNN) with the following kernel sizes 3x3, 5x5 and 7x7. This is a 3-layer CNN with filters in that order. Consider the output activation map of the last layer. *By field-of-view we mean the number of pixels in each dimension influencing the output activation map.*

a) What will be the theoretical field-of-view for that network, if all layers use a stride of 1?

b) What will be the theoretical field-of-view be if the first layer has a stride of 2 and the following layers have a stride of 1?

c) The input image has 3 channels, and each layer has 2 filters. Excluding biases, how many parameters are in this model?

d) Dilated convolutions are a common way to increase the field-of-view of a convolutional network without reducing the spatial size. How can you use dilated convolutions to make the field of view as large as possible, without getting “holes” in the view?

e) You have raw audio data for speech recognition and want to detect different words. Why can a convolutional neural network be a better architecture for this application than a standard feed forward neural network?

Exercise 4:

a) Explain the difference between residual networks and standard feed forward networks.

b) List two techniques for training large deep neural networks that are suitable when you don't have much labelled training data.

Exercise 5:

When running *occlusion experiments* for visualization, you often get large responses for regions that does not belong to the target category. For example, the output for the category car, can also give a large output for regions of road and street signs.

a) What do we mean by occlusion experiments?

b) Why do we often get large responses for other regions, even though they look very different from the target object? Give some different examples.

Exercise 6:

a) You know the architecture and the weights of a trained neural network, with good performance on a test set with human made labels. Discuss how you could construct images that you are almost certain would be classified differently by humans and the neural network. Both the humans and the neural network should also be very confident in their decision.

b) t-SNE is optimized with gradient descent. What are you differentiating with respect to?

Exercise for INF9860/PhD students only:

a) Give one example of data that deep learning will probably work well for, and one example of data where other machine learning techniques may work better. Briefly point out the differences between the two.

b) In reinforcement learning, what does the output of a Q-network represent? Why do you only need two steps to update a Q-network?.

c) An example of *hard attention* could be that one part of the network selects a sub-region of an image, and another part of the network does classification with the sub-region as input.
Why do we often use reinforcement learning to optimize networks using hard attention?

Thank You for Your Attention!