UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in: IN 54000/IN 9400 — Machine Learning for Image Analysis

Day of examination: 31th May 2019

Examination hours: 14:30 – 18:30

This exercise set consists of 6 pages.

Appendices: None

Permitted aids: Certified 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.

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

Every subtask has equal weight in the evaluation.

(Continued on page 2.)

Exercise 1 Computational graph for a simple RNN

You are given a simple RNN network as illustrated in the computational graph. Assume that we use identity functions as activation functions.

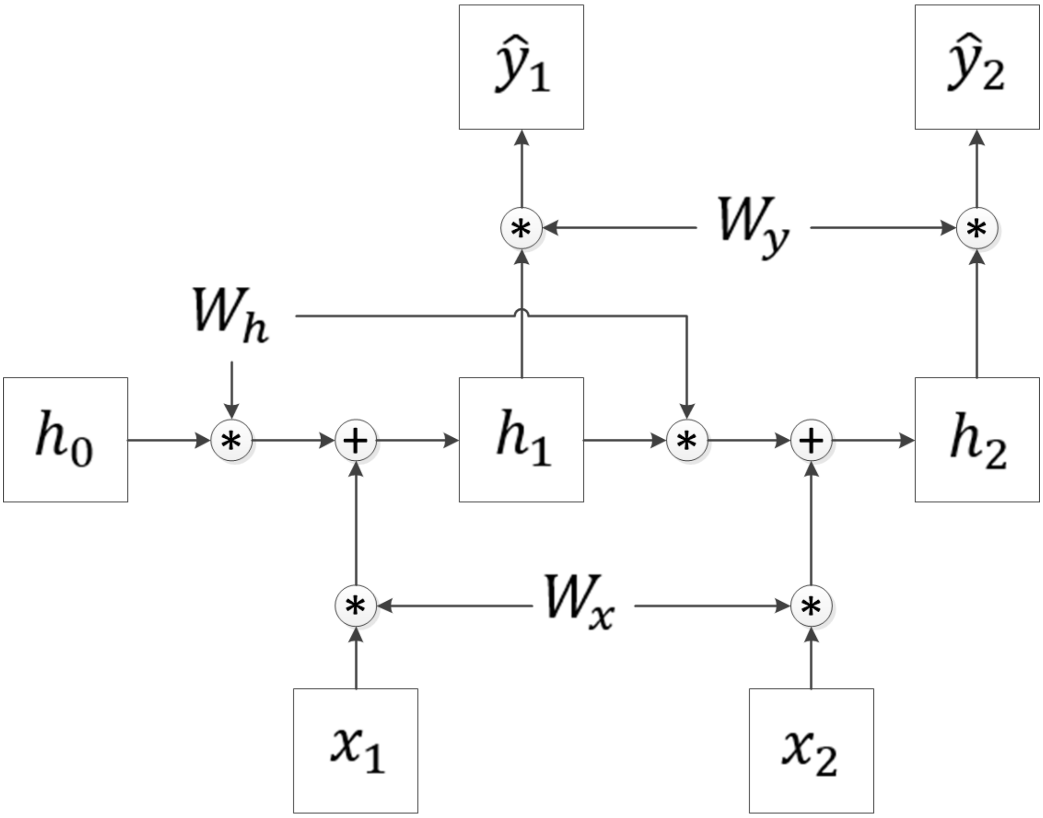

Figure 1: The structure of a very simple RNN network

Assume that the input at a given time, the hidden state, and the output at a given time are scalar.
Let $h_0 = 1$, $x_1 = 10$, $x_2 = 10$, $y_1 = 5$, $y_2 = 5$.
We assume the initial values of the weights are: $W_h = 1$, $W_x = 0.1$, $W_y = 2$

1a

Compute the predicted value $\hat{y}_2$.

1b

If we use quadratic loss, the loss at a given time t is $L_t = (\hat{y}_t - y_t)^2$. Compute the total loss given the values for the weights and inputs given above.

(Continued on page 3.)

1c

Compute the derivative of the total loss with respect to h_1, $\frac{\partial L}{\partial h_1}$.

1d

Compute the derivative of the total loss with respect to W_h, $\frac{\partial L}{\partial W_h}$.

Exercise 2 Convolutional neural networks, pooling and segmentation

2a

Which network architecture would you consider to be most parameter efficient on image data: convolutional neural network or dense neural network? Justify your answer.

2b

Discuss briefly whether convolutional neural networks build view invariance or not.

2c

Explain the different hyperparameters that define a convolutional layer.

2d

Give three reasons to use max pooling.

2e

Give two arguments against using max pooling.

2f

PhD students only

Give an example on how you could design a semantic segmentation network. Discuss which loss function you would use.

Exercise 3 Training a neural network

3a

Describe the algorithm for gradient descent with momentum updates.

(Continued on page 4.)

3b

PhD students only

Make a small sketch that illustrates Nesterov momentum updates

3c

Gradients that are close to zero can be a problem when training a network. Suggest three different things that can reduce the problem.

Exercise 4 Generalization and regularization

4a

Give an example of how you could use transfer learning.

4b

Describe briefly a few data augmentation techniques.

4c

Explain region (1) and region (2) in the figure below. Explain also the double arrow (3).

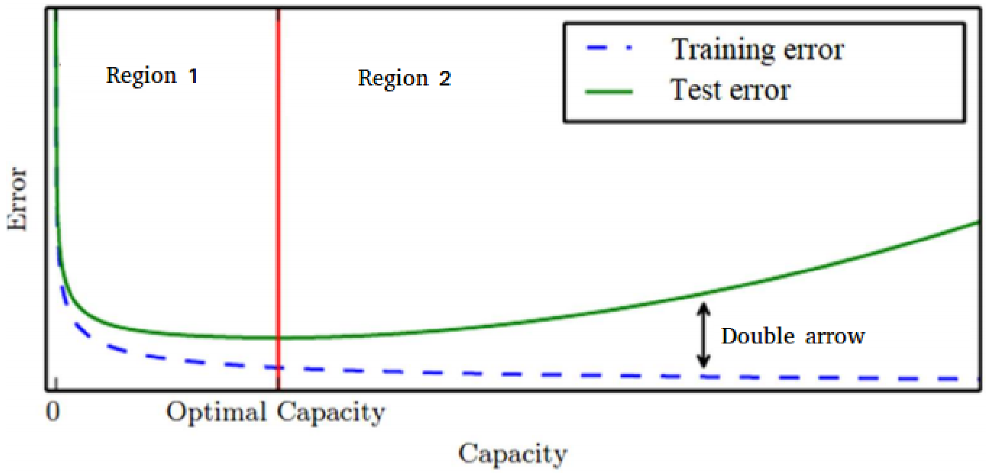

Figure 2

4d

Given the dense layer shown in figure 3. What would the **maximum** value of the output be assuming a dropout probability of p=0.5?

(Continued on page 5.)

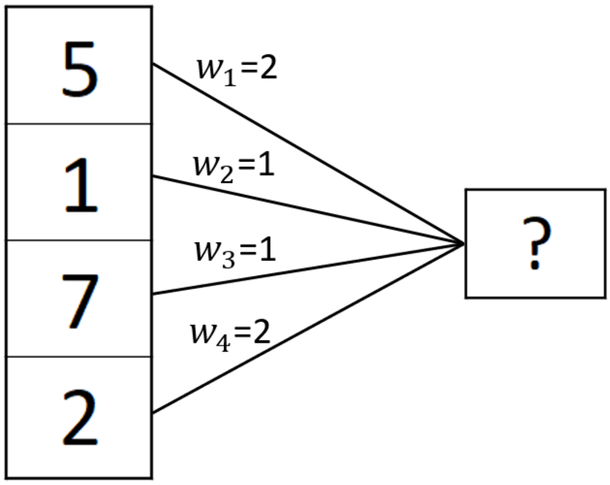

Figure 3

4e

Given the dense layer shown in figure 3. What would the **minimum** value of the output be assuming a dropout probability of p=0.5?

Exercise 5 Visualization and adverserial fooling

5a

Consider the last layer before the softmax layer in a convolutional network. Suggest a method for visualizing the information learned by the set of features or feature vectors in this layer.

5b

Explain shortly how we can use backpropagation with respect to the input image to visualize a filter.

5c

Explain briefly what an adverserial image is.

5d

Describe briefly the fast gradient sign method for generating adverserial examples.

(Continued on page 6.)

Exercise 6 Unsupervised learning

6a

t-SNE is often used to vizualize high dimentional data by embedding the data to a lower dimention (2D/3D). How would you inculde a new sample in the t-SNE?

6b

How does t-SNE mitigate the crowding problem?