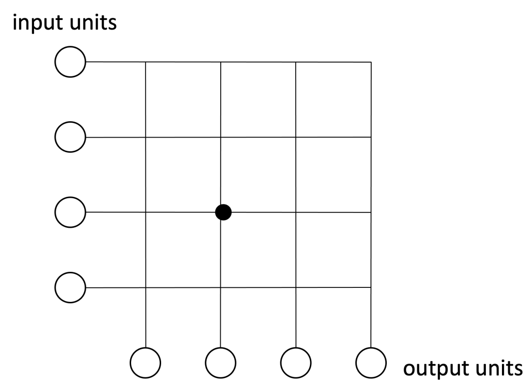
7COM1033: Neural Networks and Machine Learning
7th May 2021, 10:00 - 11:30
Time: 90 minutes
Total marks: 82, 5 questions, 4 pages of questions
Version 1001

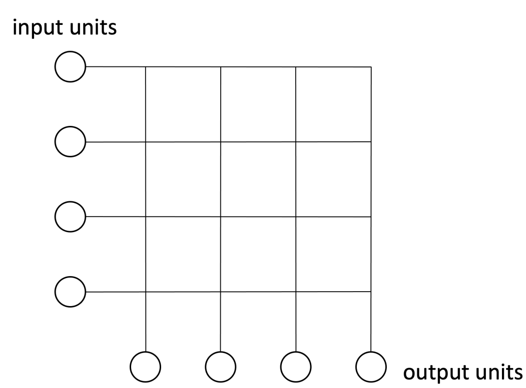
1. Associative Networks (TOTAL: 21 marks)

   (a) (4 points) Give an example of an auto-associative memory task you might encounter in real life.

   (b) (4 points) The following diagram shows an associative (Willshaw) net with four input units and four output units. In this network, 15 of the 16 synapses are switched off. The only synapse that is switched on is marked by a black dot. The association between which input pattern and which output pattern has been stored?



   (c) (7 points) Store the following input-output pattern pairs in the net: input pattern 1 = $(0, 1, 1, 0)$, output pattern 1 = $(1, 0, 1, 0)$; input pattern 2= $(0, 1, 0, 1)$, output pattern 2 = $(0, 1, 0, 1)$. Redraw the net and label the synapses that have been switched on.



   (d) (6 points) Present input pattern 1 and try to recall the correct output pattern. What are the input sum, threshold and resulting activation of each of the four output units?

2. Perceptrons (TOTAL: 15 marks)

   (a) (4 points) In a simple threshold linear unit (perceptron) with two inputs x_1 and x_2 and weights w_1 and w_2, how is the output y calculated?

   (b) (5 points) b) Draw the input space (including all four input patterns with an indication of their target outputs and a potential decision line) for the following classification task:

x_1	x_2	target output
0	1	0
0.5	0	0
0	2	1
1.5	0	1

   (c) (3 points) Give the equation that describes your decision surface .

   (d) (3 points) Give potential weights w_1 and w_2 and a potential threshold θ for the decision surface

3. Unsupervised Learning (TOTAL: 10 marks)
   Assume a Kohonen Network (Self-Organising Map) with two input units and a physical grid of $3 \times 3 = 9$ output units a, b, c, d, e, f, g, h, i. The weight vectors of the output units are initially given by (shown by their physical arrangement in output space):

$\mathbf{w}_a = (1, 5)$	$\mathbf{w}_a = (2, 3)$	$\mathbf{w}_c = (4, 1)$
$\mathbf{w}_d = (5, 1)$	$\mathbf{w}_e = (3, 2)$	$\mathbf{w}_f = (1, 4)$
$\mathbf{w}_g = (5, 5)$	$\mathbf{w}_h = (3, 3)$	$\mathbf{w}_i = (1, 1)$

   (a) (4 points) The network is presented with an input vector $\mathbf{x} = (7, 1)$. Which output unit will be the winner and why?

   (b) (2 points) Assuming a learning rate $\alpha = 0.5$, what will the weight vector of the winning output be changed to after a single presentation of the input vector $\mathbf{x} = (7, 1)$?

   (c) (4 points) The network from (a) is presented with a large number of inputs where x_1 and x_2 are chosen randomly from values between 0 and 10. Draw the final state of the network both in (mathematical) input space and in (physical) output space.

4. Deep Learning (TOTAL: 15 points)

(a) Consider an RGB input image with shape $50 \times 1000 \times 3$.

i. (3 points) If you flatten this image and connect it to 30 hidden units in the next layer of a feedforward neural network, how many parameters are needed to describe the relationship between these two layers (assuming a bias vector)? Explain how you arrive at your answer.

ii. (3 points) If instead you run this image in a convolutional layer with filters of size 8×8, assuming a stride length of 1, how many times can you slide the filter in the vertical and horizontal directions? Finally, given your previous answers, what is the size of the resulting filter maps?

(b) Let A, the 5×5 matrix below, represent the pixel values of a simplified input image in a convolutional neural network.

$$A = \begin{pmatrix} 20 & 7 & 20 & 20 & 20 \\ 0 & 0 & 0 & 20 & 0 \\ 0 & 0 & 20 & 0 & 0 \\ 0 & 20 & 0 & 0 & 0 \\ 10 & 10 & 10 & 10 & 20 \end{pmatrix}$$

Assuming no padding, a stride length of 1 and a bias of 0, calculate the feature maps resulting from the following filters:

i. (2 points) $F_1 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

ii. (2 points) $F_2 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$

iii. (2 points) $F_3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$

iv. (3 points) Interpret the above results (i.e. describe the kinds of features in an image each filter emphasizes, e.g. corners).

5. Deep Learning (TOTAL: 21 points)

(a) (2 points) Find the derivative $\sigma'(z)$ of the sigmoid function given below:

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Given your answer in terms of e and z.

(b) (1 point) Which of the following is true of $\sigma(z)$ when z becomes very large and negative?

i. $\sigma(z)$ approaches ∞.

ii. $\sigma(z)$ approaches 0.

iii. $\sigma(z)$ approaches 1.

(c) (3 points) What happens to $\sigma'(z)$ when $z = 0$? Explain your answer.

(d) A very simple neural network is given by the following set of equations:

$$\begin{aligned} z_1 &= w_1 x + b \\ a_1 &= ReLU(z_1) \\ z_2 &= w_2 a_1 \\ y &= \sigma(z_2) \end{aligned}$$

The loss function is given by $C = \frac{1}{2}(y-t)^2$, where y is the prediction and t is the target. The ReLU function is defined as $ReLU(z) = \max(0, z)$ for $z \in \mathbb{R}$.

i. (4 points) How many neurons and layers are there in this network? What are the activation functions?

ii. (2 points) What is the output of the network, y, if the input $x = -1$, and the weights and biases are $w_1 = 2$,$w_2 = 2$,$b = -1$?

iii. (1 point) What is the loss for the input values given in part ii) and $t = 1$? (You may give your answer to 2 significant figures or leave it as a fraction).

iv. (5 points) Use the Chain rule to give an expression for $\frac{\partial C}{\partial w_1}$.

v. (3 points) Name and describe a problem often encountered in neural networks that occurs when many partial derivatives are multiplied together. How is training affected?