Resit Exam Neural Networks

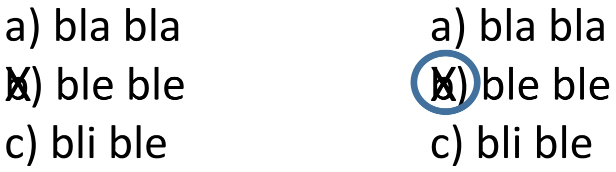
Wojtek Kowalczyk
w.j.kowalczyk@liacs.leidenuniv.nl

03.07.2019

It is a closed book exam: you are not allowed to use any notes, books, calculators, smartphones, etc. The number of points attached to each question reflects the (subjective) level of question's difficulty. In total you may get 100 points. The final grade for the exam is the total number of points you receive divided by 10.

The exam consists of a number of questions with a "single choice answer". It means that for each question you should select exactly one answer. For every correct choice you get some points; for an incorrect choice or no choice you get 0 points. In very few cases you are asked to write down some formulas.

Mark your choices by crossing the selected option. In case you want to "undo" your choice put a circle around the cross. For example, on the left side the option **b** is selected; on the right side nothing is select – the selection of **b** is "undone":



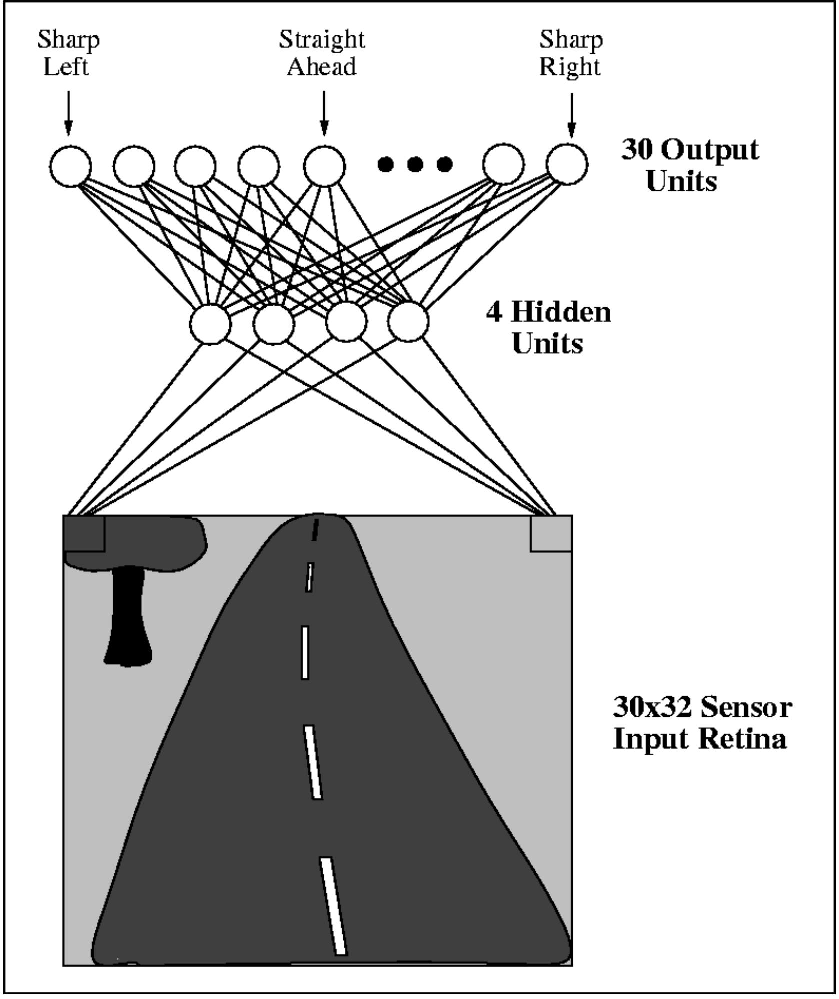
If you think that your marking is no longer readable, put your final choice(s) on the left margin (e.g., by writing "a" if you want to select "a"). Finally, you are free to add to your answers your comments (in a free space). Your comments may help us to adjust the exam grade (up or down – depending on the comment).

Before starting answering the questions, fill in the following entries:

Name:

Student number:

Study type (ICT, Astronomy, …):

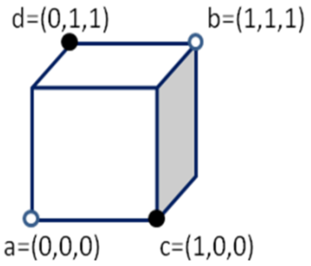
10 pts	**Q1: The Alvinn System**
Question	In 1989 Pomerleau presented a neural network system that could learn how to drive a car (see below). Not shown on the picture are bias parameters. To calculate, for a given input image, the output of the network a number of multiplications (#M) and additions (#A) is required (we ignore other operations like computing the values of logistic, sigmoid, softmax, etc. functions).  Sharp Left Straight Ahead Sharp Right 30 Output Units 4 Hidden Units 30x32 Sensor Input Retina
Answer options	1. What is the value of #M? Write down an expression. a) #M = 2. What can you say about the relation between #M and #A? [✓ b) #M>#A] c) #M=#A d) #M<#A e) None of the above

10 pts	**Q2: Bayes' Rule**
Question	Let us suppose that the chance of having a lung cancer (C) is 0.0001 and that a new system for detecting long cancer is almost perfect: when applied to a healthy patient, the test result is negative in 99% of cases; when applied to a patient with lung cancer the test result is positive in 99% of cases. Symbolically, P(C)=0.0001, P(No|NoC)=0.99, P(Yes|C)=0.99. What is P(C|Yes)? In other words, what is the probability that a patient really has lung cancer when the result of the test is positive?
Answer options	a) About 0.99 b) About 0.1 c) About 0.01 d) About 0.001 e) About 0.0001 f) None of the above If you have problems with finding actual values (you are not allowed to use a calculator), just write down a derivation of the final formula that expresses this probability:

3 pts	**Q3: Discriminant function**
Question	Let us suppose that f(x) is a discriminant function for sets A and B. Which of the following functions is not necessarily a discriminant for A and B?
Answer options	a) –f(x) b) $f(x)^3$ c) exp(f(x)) d) sigmoid(f(x)) e) none of the above

3 pts	**Q4: Multi-class linear separability**
Question	Which definition of the *multi-class linear separability* concept (restricted here to 3 classes) is correct: Sets A, B, C are linearly separable if and only if:
Answer options	a) Any two of them are linearly separable. b) Any one of them is linearly separable from the union of the remaining two sets. c) A single layer perceptron (with 3 nodes) can be trained to separate these sets. d) None of the above definitions is correct.

5 pts	**Q5: Cover’s Theorem**
Question	Let us consider a collection of 1500 color images of size 16x16, where each pixel is a randomly generated triplet (r,g,b) of numbers between 0 and 1. Suppose, that these images are split into two classes A and B, at random. What is the probability that classes A and B are linearly separable?
Answer options	a) About 0.0 b) About 0.5 [✓ c) About 1.0] d) None of them

5 pts	**Q6: Recognizing corners of a 3D cube**
Question	Consider the following set of 4 points in 3-dimensional space: *a=(0,0,0), b=(1,1,1), c=(1,0,0), d=(0,1,1)* that are split into two sets *A={a, b}* and *B={c, d}*: 
Answer options	1. Are the sets A and B linearly separable? [✓ a. Yes] b. No 2. Can the sets A and B be separated with help of a generalized perceptron: a network that consists of 2 perceptrons such that one of them returns bigger values than the other one on points from A, and smaller ones than the other one on points from B? a. Yes [✓ b. No] 3. Can sets A and B separated with help of a simple network with 3 input nodes (representing x, y, z coordinates), two hidden nodes and one output node? [✓ a. Yes] b. No

5 pts	**Q7: ReLU**
Question	Let us consider a Rectified Linear Unit that computes a function f(x). Which of the following statements is *false* (select one):
Answer options	1. f(x) is unbounded (i.e., may take arbitrary big values) [✓ 2. The derivative of f(x) is either 0 or 1] 3. The derivative of f(x) is either 0 or x 4. ReLU's reduce the problem of vanishing or exploding gradients

7 pts	**Q8: Gradient descent**
Question	Let us suppose that to find a minimum of the function f(x) = x^2 the gradient descent algorithm, with the learning rate set to 1/2, was run for a number of iterations, starting at $x_0 = 1$ and stopping as soon as an x was found such that f(x) < 10^{-6}. After how many iterations the algorithm stopped? Select the answer which is closest to your estimate:
Answer options	a) 2 b) 5 c) 10 d) 25 e) 50

4 pts	**Q9: Gradient descent with momentum**
Question	Suppose that while searching for a minimum of a function *f(x)*, after *k* iterations of the algorithm gradient descent with momentum, the algorithm reached a point x_k. What update rule should be applied to find x_{k+1}?
Answer options	*a)* $x_{k+1} = x_k + \alpha g(x_k) + \beta d(x_k)$ [✓ *b)* $x_{k+1} = x_k - \alpha g(x_k) + \beta d(x_k)$] *c)* $x_{k+1} = x_k + \alpha g(x_k) - \beta d(x_k)$ *d)* $x_{k+1} = x_k - \alpha g(x_k) - \beta d(x_k)$ *e) none of the above* Here α denotes the learning rate, β denotes the momentum, *g(x)* denotes the gradient of the function being optimized (i.e., the vector of partial derivatives of *f* at *x*), and *d(x)* denotes the direction of the last step, i.e., $d(x_k) = x_k - x_{k-1}$.

10 pts	**Q10: Plain Recurrent Neural Networks**
Question	Let us consider a simple (vanilla) recurrent network which has I input nodes, H hidden nodes, O output nodes. No bias parameters are used. What was the total number of trainable weights used by this network?
Answer options	Write down a formula that involves O, I and H: #parameters=

5 pts	**Q11: Contrastive Divergence Algorithm**
Question	How many multiplications are needed by the Contrastive Divergence algorithm to update weights for a single input vector x, assuming that two steps of Gibbs sampling are used? Assume that the network has M input nodes, N hidden nodes, and no biases.
Answer options	a) 3*M*N b) 5*M*N c) 7*M*N d) 8*M*N e) None of the above

2 pts	**Q12: AlphaGo**
Question	What network architecture was used by the AlphaGo program? Select the most specific (correct) answer.
Answer options	a) a multi-layer perceptron b) a convolutional network c) a recurrent network d) an LSTM network e) a residual network (resnet) f) none of the above

2 pts	**Q13: Word2Vec**
Question	What network architecture was used generate the “word to vector” mapping? Select the most specific (correct) answer.
Answer options	a) a multi-layer perceptron b) a convolutional network c) a recurrent network d) an LSTM network e) a residual network (resnet) f) none of the above

2 pts	**Q14: DeepDream**
Question	What network architecture was used generate the Google DeepDream video(s)?
Answer options	a) a multi-layer perceptron b) a convolutional network c) a recurrent network d) an LSTM network e) a residual network (resnet) f) none of the above

3*5 pts	**LeNet5**
Question	The first convolutional layer of LeNet5 consists of 6 feature maps, each of size 28x28. Each map is determined by a convolutional filter of size 5x5 which is applied to the input layer of size 32x32, with padding=0 and stride=1. Moreover, each filter uses a bias term.
Answer options	a) How many trainable parameters define this convolutional layer? Write down a formula! b) How many connections exist between the input and the first convolutional layer? Write down a formula! c) How many connections would exist if the input and the first convolutional layer were fully connected? Write down a formula!

5 pts	**Invariance with respect to input permutations**
Question	It is well-known that convolutional networks outperform simple multi-layer perceptrons on image recognition task, such as ImageNet collection of images. Let us suppose that the original ImageNet data would be “randomized”: a fixed permutation of pixel positions would be applied to all images making them practically unreadable (at least for humans). How would it affect the learning capabilities of CNNs and MLPs?
Answer options	a) Both networks would suffer a lot: they wouldn’t be able to learn to classify “randomized” images. [✓ b) Both networks would be slightly affected: they would be able to learn to classify “randomized” images with a similar accuracy as on the original set.] c) MLP wouldn’t be affected; CNN would suffer. d) CNN wouldn’t be affected, MLP would suffer.