UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

INF3490/4490 — Biologically Inspired Computing
November 30th, 2017
Exam hours: 09:00 – 13:00
Permitted materials: None
The course teachers will visit the exam room at least once during the exam.

The exam text consists of problems 1-40 (multiple choice questions) to be answered by selecting true or false for each statement. If you think a statement could be either true or false, consider the most likely use/case. Problems 41-42 are answered by entering text (preferably in English language). Problems 1-40 have a total weight of 80%, while problems 41-42 have a weight of 20%.

Scoring in multiple choice questions
Each problem has a variable number of true statements, but there is always one true and one false statement for each problem.

0.5 point is given for each correctly marked statement. Further, an incorrectly marked statement or an unmarked statement(s) results in 0 point. The maximum score for a question is 2 points and the minimum is 0. Since it is possible to get a positive score just by random answering, the final grading thresholds will be adjusted accordingly.

Problem 1

Machine learning	A	Can make technology adapt by learning	X
	B	Is applicable to classification problems	X
	C	Biology-inspired methods are not applicable	
	D	Combines gradient ascent and descent	

Problem 2

Gradient ascent	A	The starting point does not affect whether a global optimum is found	
	B	Is applicable when the problem is about optimizing for a benefit	X
	C	Gradient ascent will always find a local or global optimum	X
	D	The step size is proportional to the gradient	X

Problem 3

Which of the following is/are continuous optimization problems?	A	Control engineering	X
	B	Travelling salesman problem	
	C	Routing tracks during chip layout design	
	D	Product recommendations in web shopping	

Problem 4

Exploration and exploitation	A	Exploration consists of trying out local variations of a currently known good solution	
	B	Exploitation is most important to find the global best solution	
	C	Exploration and exploitation are combined in evolutionary algorithms	X
	D	Exploration is important for problems with many local optima	X

Problem 5

Which is/are typically *not* a part of an evolutionary algorithm (EA)	A	A population	
	B	A fitness function	
	C	A temperature control	X
	D	A termination criterion	

Problem 6

Which is/are not EA variation operator(s)?	A	Mutation	
	B	Hillclimbing	X
	C	Fitness sharing	X
	D	Recombination	

Problem 7

What property/ies of the elements of a solution do we try to conserve when doing mutation and crossover on a permutation-based genotype?	A	Order	X
	B	Adjacency	X
	C	Sum	
	D	Concurrency	

Problem 8

Which operator(s) are suitable for a permutation representation?	A	Creep mutation	
	B	Uniform mutation	
	C	Order crossover	X
	D	Edge recombination	X

Problem 9

What affects the selection pressure in an evolutionary algorithm?	A	The population size	X
	B	The selection operator	X
	C	The variation operator	
	D	Whether elitism is applied	X

Problem 10

Fitness sharing	A	Is dependent on a niche size	X
	B	Does not allow mating between very different individuals	
	C	Affects the selection pressure	X
	D	Depends on a way to measure the distance between individuals	X

Problem 11

Island model EAs	A	Are more parallellizable than standard EAs	X
	B	Never allow individuals to move between islands	
	C	Allow an individual on one island to mate with an individual on a different island	
	D	Help preserve diversity	X

Problem 12

Which is a type of evolutionary algorithm?	A	Support vector machine	
	B	Genetic programming	X
	C	Greedy search	
	D	Evolution strategies	X

Problem 13

In evolution strategies	A	(μ,λ)–selection can result in losing the best candidate	X
	B	Self-adaptation of parameters is common	X
	C	Offspring may have more than 2 parents	X
	D	$(\mu+\lambda)$ is the best selection strategy for leaving local optima	

Problem 14

Genetic Programming	A	Uses trees as genotypes	X
	B	Does not apply variation operators	
	C	Can be used to evolve computer programs	X
	D	Chromosomes may grow or shrink during evolution	X

Problem 15

Which is/are example(s) of a design problem?	A	Planning the University of Oslo's annual budget	X
	B	Optimize the best use of a fleet of self-driving cars	
	C	Restructuring the codebase of Facebook to increase security	X
	D	Selecting the most appropriate car to pick up a customer needing a taxi	

Problem 16

Which is/are on-line (that is, used during the run of the algorithm) performance measure(s) for EAs?	A	Mean best fitness at termination	
	B	Success rate	
	C	Population distribution	X
	D	Improvement per generation	X

Problem 17

Hybrid evolutionary algorithms	A	May apply domain-knowledge in the search for a solution	X
	B	Cannot apply random mutations	
	C	May apply a local search	X
	D	May apply seeding	X

Problem 18

Multiobjective evolutionary algorithms	A	Depend on a way to scalarize conflicting objectives	
	B	Aim to find a good estimate of the Pareto Front	X
	C	Aim to spread the population evenly across the Pareto Front	X
	D	Use the same selection operators as single-objective EAs.	

Problem 19

Supervised Learning	A	Training data is needed	X
	B	A class label for each input vector is required	X
	C	A validation data set can reduce the training set error	
	D	No information about how to correct error is provided	

Problem 20

Which data set(s) is/are important for obtaining generalization during training?	A	Test data	
	B	Training data	X
	C	Validation data	X
	D	All data sets are equally important	

Problem 21

McCulloch and Pitts Neurons	A	Are simple computer models of biological neurons	X
	B	Their activation function is a sigmoid	
	C	Learn by adapting connection weights	X
	D	The thresholding function is used in multi-layer perceptrons	

Problem 22

(Single-layer) Perceptrons	A	Can learn to approximate any function	
	B	Learn by modifying their weights to minimize the output error	X
	C	Can only learn linearly separable problems	X
	D	Usually apply a learning rate of 1	

Problem 23

Multilayer Perceptrons	A	Can not learn to solve the XOR-problem	
	B	Are a type of recurrent network	
	C	Are trained by adapting only the weights at the output layer	
	D	Are applicable in deep learning	X

Problem 24

Backpropagation	A	Is applicable for unsupervised learning	
	B	Minimizes the sum-of-squares error	X
	C	May apply a momentum to avoid local optima	X
	D	Is used to train recurrent neural networks	

Problem 25

Which is true about reinforcement learning?	A	No learning rate is needed	
	B	The objective is to minimize the reward function	
	C	Each state is the result of the previous actions	X
	D	The set of actions we take defines our policy	X

Problem 26

In reinforcement learning:	A	Learning is guided by the reward	X
	B	The reward tells us what we should have done	
	C	The greedy action selection strategy includes exploitation and exploration	
	D	The expected reward of a policy is the "value" of that policy	X

Problem 27

Which is true about the Q-learning algorithm?	A	Always assume the optimal action	X
	B	Follows an ε-greedy search behavior	X
	C	Results in a safer plan than the SARSA algorithm	
	D	Discounts the rewards based on the reward's value	

Problem 28

Deep Learning	A	Is more suitable in optimization than classification	
	B	Is usually implemented using a neural network architecture	X
	C	Is an ANN structure with one or two hidden layers	
	D	Several different kinds of layers are used	X

Problem 29

Support Vectors	A	Are data points which lie closest to the classification line	X
	B	We can ignore other data and just use support vectors for classification	X
	C	Are limited to one per class of data points	
	D	Slack variables are relevant for linearly separable problems	

Problem 30

Support Vector Machines	A	Use Kernels to make the input data linearly separable	X
	B	Try to minimize the marginal area	
	C	The best classifier goes through the middle of the marginal area	X
	D	Is not applicable for classification problems	

Problem 31

Ensemble Learning	A	Each classifier should be slightly better than the previous one	
	B	Uses bagging or boosting to combine the results of the classifiers	X
	C	Makes classification based on the combined results of all classifiers	X
	D	Provides significantly better results than each individual classifier	X

Problem 32

AdaBoost Algorithm	A	Updates the weights based on previous errors	X
	B	Continuous until the combined error is sufficiently low	X
	C	Initially, all weights are equal	X

	D	Works in most cases better than Bagging	

Problem 33

Principal Components Analysis	A	Does not directly help with data that is not linearly separable	X
	B	Is a linear transformation	X
	C	Principal component is the direction in the data with the smallest variance	
	D	Removes some data which might be noise	X

Problem 34

Which is true in unsupervised learning?	A	The data applied in training is labeled	
	B	Aims to spot similarities between data points	X
	C	The objective is to minimize an external error function	
	D	Can be used when competitive learning is not applicable	

Problem 35

k-mean clustering	A	Always overfits the data points	
	B	Number of clusters must be specified	X
	C	The initial cluster centers are randomly selected	X
	D	Is suitable for discovering clusters with non-convex shapes	

Problem 36

Self-Organizing Maps	A	Is not able to handle high dimensional data	
	B	It is always better to use a small network	
	C	Is a neural network with topological meaning	X
	D	Each neuron is only connected to an input	

Problem 37

Swarm Intelligence	A	Avoids local interactions between units	
	B	Follows a centralized control style	
	C	Swarms are expected to adapt to environmental fluctuations	X
	D	Mimics the collective behavior of insects	X

Problem 38

Ant Colony Optimization	A	Is not able to solve the Traveling Salesman problem	
	B	All initial directions are randomly selected	X
	C	Learning happens based on the concentration of pheromone in each path	X
	D	Considers the quality of the food source and the length of the path	X

Problem 39

39 Fuzzy Logic	A	Is an optimization technique	
	B	Uses Characteristic Function instead of Membership Function	
	C	Does not need much detailed knowledge of the system	X
	D	Can account for variability in the inputs	X

Problem 40

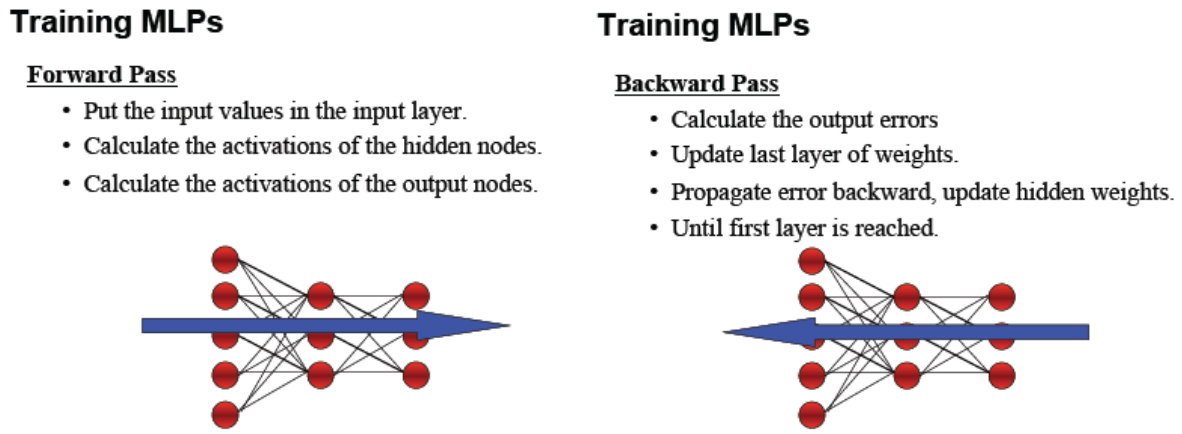
Fuzzy Logic	A	Is a decision making tool	X
	B	Number of rules depends on the number of output variables	
	C	Each input variable should have three membership functions	
	D	We can use the Center of Gravity method to Defuzzify the output	X

Problem 41 (14%)

a) (8%) Sketch the steps in the forward and backward phase of the multi-layer perceptron algorithm (backpropagation) for sequential training. Use words (and not equations) and by referring to input, hidden and output nodes, respectively.

1. an input vector is put into the input nodes
2. the inputs are fed forward through the network
• Each input is multiplied with one weight for each hidden node in the first layer, and the products are summed and a bias (weight value) is added and the result is sent through an activation function like the sigmoid for each hidden node.
• the outputs of these nodes and the second-layer weights are multiplied and processed in the same way as for the hidden layer to compute the output nodes.
3. the error is computed as the sum-of-squares difference between the network outputs and the targets
4. this error is fed backwards through the network in order to
• first update the second-layer weights
• and then afterwards, the first-layer weights

This process is repeated for each training vector. This is the sequential version of the backpropagation algorithm (other variants called batch and epoch training accumulate the error for a number of training vectors before the weights are updated).



b) (6%) An error term is used for updating the weights of the output layer:

$$\delta_o(\kappa) = (y_\kappa - t_\kappa)\, y_\kappa(1 - y_\kappa)$$

Explain the different parts (including what they represent) of the equations (you don´t need to use indices).

y-t is the output error and y(1-y) is the derivate of the (sigmoid) activation function.

How is the term above used for updating the weights in the hidden layer?

The delta term is multiplied with the corresponding weights in the output layer. This sum-of-products is then multiplied with the derivate of the hidden layer activation function to compute delta terms for the hidden layer nodes. The computed error values impact how much each weight in the hidden layer is updated.

$$\delta_h(\zeta) = a_\zeta(1 - a_\zeta) \sum_{k=1}^{N} w_\zeta \delta_o(k)$$

Problem 42 (6%)

a) We are going to make a mobile robot that can be used to monitor the mental and medical state of elderly living on their own at home. List and briefly explain at least three different key ethical related challenges that need to be addressed for such a robot.

The three most important challenges:
• Safety: There must be mechanisms (or opportunities for an operator) to control and limit a robot's autonomy and make sure it doesn´t result in physical damage.
• Security: There must be a password or other keys to avoid inappropriate and illegal use of a robot.
• Privacy policy: Software and hardware should be used to encrypt and password protect sensitive data that the robot needs to save. Further, try to limit the data collection and distribution beyond what is needed for mental and medical state monitoring.

Also relevant (but less important):
• Traceability: Similarly as aircraft, robots should have a "black box" to record and document their own behavior.
• Identifiability: Robots should have serial numbers and registration number similar cars.

b) Would the type of sensors applied or processing of the sensor data substantially impact any of the challenges in a) Justify your answer.

Avoid using video and sound sensors and do local processing in the robot to increase privacy. That is, rather than distributing raw sensor data, send only the computed state of the person from the robot/home to the caregiver.