(a) (1 punt) Design a neural network which could do the job (indicate number of inputs, outputs, hidden neurons ? assume one hidden layer, activation functions)

(b) (2 points) Suppose Joey has put together a set of 400 games, for which he has collected both input information and outcomes (targets). Would it be a good idea for Joey to use a deep architecture instead of the one you laid out at point (a)? Why? Provide arguments to support your answer.

(c) (1 punt) Suppose that Joey wants to use backpropagation to train the network. Would it be a good idea for him to start the training with a small learning rate, and then increase it over time? Why? Describe what the effect that this learning rate adaptation over time would have on the learning process.

(d) (1 punt) Joey is a bit worried about overfitting, which can happen often with neural networks. List three things that he could do to minimize (or at least control for) the risk of overfitting when training his network.

16. Despite all his efforts (and your advice), Joey can't manage to avoid overfitting. Hence, he decides to resort to an evolutionary strategy to train the weights of his network.

(a) (1 punt) Why would that be a good idea? Briefly describe, if any, the advantages of using evolutionary strategies over backpropagation to train neural networks

(b) (5 points) Assuming that Joey established a single hidden layer architecture for his network, with K hidden neurons:

1. describe the encoding of the problem into a chromosome. Use a graphical representation if it can help.
2. describe the initialization of the population
3. define the fitness function
4. describe the selection strategy
5. describe the genetic operators

(c) (2 points) Suppose that Joey would want to detemine a better architecture for the network (i.e., optimizing the number of neurons in the hidden layer). Could he use an evolutionary strategy to accomplish that task too? If so, how (describe briefly which steps you would take, also in terms of b.1 - b.5)? If not, why, and which technique could Joey use instead?

End exam.
Check if you answered all questions. You should have answered 16 questions.

Open questions

13. The following question is about using CI techniques in practice. Indicate which of the technique(s) (learned in this course) you would use to solve the challenge. To optimize usage of transportation trucks, a transportation company wants to investigate the best boxing dimensions for their trucks. The main requirement for the box is that it is a one-size fits all box that can be stacked and can contain multiple products. The products they transport range from 5x5cm to 30x30cm. They have three types of trucks: one with a 2x2x3 Meter (Height x Width x Length) cargo hold, one with a 2.5 x 2 x 3 cargo hold and a large truck with a 3 x 2 x 5 cargo hold. They ask you to solve this challenge. Motivate your answer by explicitly addressing the following:
    (a) (1 punt) What is the TYPE of problem(s) that you need to solve (optimalisation, clustering, etc..)?
    (b) (1 punt) Which method do you pick to solve each problem and why?
    (c) (1 punt) How would you use each method to solve the problem(s) in the challenge (you can be brief here, but your explanation must clearly indicate how you would approach solving the problem).

14. When I was a kid, I could not believe that both Saint Nicolas and Santa Claus would drop by and give me presents. So, I figured I could probably deduce who was the one giving me gifts based on observable evidence. Saint Nicolas puts presents in shoes, while Santa puts them in socks. So I decided to put this to the test. I put both a sock and a shoe in front of the chimney, and sang two songs (to lure both of them to my house). When I woke up, I saw a present in my sock but not in my shoe. I estimated the a-priori chance of Saint Nicolas being in the Netherlands at about 3/52 (he is here 3 weeks of the year), while the a-priori chance of Santa being here is only 3/365 (usually around Christmas). Further, I figured that both Saint Nicolas and Santa could in principle put presents in both the sock and the shoe. However, the odds of them doing so would be different. So, I estimated the chance of getting a present in my shoe when both Saint Nicolas and Santa dropped by to be 0.9, equal to the chance of getting something in my Sock when they both dropped by. The chance of getting something in my shoe when only Saint Nicolas is present is still high, about 0.8, equal to the chance of Santa putting something in my sock. The odds of getting something in my sock due to an action of Saint Nicolas are rare, about 0.2 (it could be by mistake though), and I estimated the same odds for getting something in my shoe from Santa. The odds of getting something in my shoe when both the good guys did not drop by are very small (who else would do that, right?), about 0.01 (and the same counts for the sock).
    (a) (2 points) Draw(!) a Bayesian network - using the correct causal relations - that can be used to determine the cause of a gift. Show the variables (hint: there are 4 in total). Give the probabilities IN TABLE FORM showing ALL the conditional and a-priori chances.
    (b) (4 points) Calculate the chance that Santa Claus dropped by, given that I know I got a present in my Sock and NOT in my Shoe. Show your calculation.
    (c) (2 points) Assuming the previous situation, what happens to the chance of Santa being the one responsible for the gift if the odds of Sint making an error and putting a present in my Sock would increase to 0.5 (so, P(Sock|Sint,-Santa)=0.5? Explain your answer. You do not have to recalculate the odds for santa (though you can of course). Your answer must be consistent with the given odds and must be complete using the evidence available.

15. Joey just graduated from his masters in computer science, but doesn't really feel like finding a job. Rather, he'd like to win a lot of money betting on football games. So, he decides to train a neural network to predict the outcome of a game (between two teams A and B) in terms of (1) winner team and (2) goal difference. The network bases its prediction on the following information:
    - A parameter that indicates the shape of each player (assume 11 players per team) within the range [1,10]
    - The number of goals in the season per player
    - The statistics of former matches between the same two teams within the past 10 years (e.g., Team A won 30% of the matches, Team B 45%, and 25% of the matches were tied)
    - Where the match takes place (Team A's stadium, Team B's stadium, or neutral place)

8. (2 points) In Particle Swarm Optimization, particles move throughout the solution space depending on their velocity v, expressed as:

$$v_{ij}(t+1) = \phi v_{ij}(t) + c_1 r_{1j} \left[y_{ij}(t) - x_{ij}(t)\right] + c_2 r_{2j} \left[\hat{y}_{ij}(t) - x_{ij}(t)\right]. \quad (1)$$

What does the term $y_{ij}(t)$ represent in the above equation?

   A. the highest fitness value achieved by the particle so far
   B. the particle position which achieved the highest fitness so far
   C. the highest fitness value achieved by the swarm so far
   D. the swarm position which achieved the highest fitness so far

9. (2 points) The next 4 multiple choice questions are about the tradeoff between exploration and exploitation. In Computational Intelligence exploration refers to "broadening the search", while exploitation refers to "narrowing down the search". Which of the following statements is true. When learning a solution (using e.g., Reinforcement Learning, Genetic Algorithms or Swarm Intelligence):

   A. Exploitation is important in the beginning, because you want to avoid having to cover the complete search space.
   B. Exploitation is to be avoided during the complete search process, because this increases the risk of ending up in a local optimum.
   C. Exploration is important in the beginning, because you want to avoid ending up in a local optimum.
   D. Exploration is to be avoided during the complete search process, because this increases the risk of never converging.

10. (2 points) The function of τ in the Boltzmann (or Softmax) Action Selection function in Reinforcement Learning $P(a_x) = e^{Q(a_x)/\tau} / Sum_i e^{Q(a_i)/\tau}$ is as follows (select one statement that is true):

   A. A high τ favors exploration, while a τ close to 0 favors exploitation.
   B. A high τ favors exploitation, while a τ close to 0 favors exploration.
   C. τ has nothing to do with the exploration versus exploitation tradeoff.
   D. When τ is set to a large number, action selection will strongly favor good actions over worse actions.

11. (2 points) Given an Evolutionary Strategy, which of the following statements is true:

   A. Mutation to the genes drives exploration, while elitism (winner take all offspring selection) drives exploitation.
   B. Mutation to the genes drives exploitation, just like elitism.
   C. A large population (to select from each round) favors exploitation, while cross-over favors exploration.
   D. Evolutionary Strategies only use exploitation (and not exploration) as they try to increase fitness of chromosomes at every round.

12. (2 points) Given Particle Swarm Optimization (PSO), which of the following statements is FALSE:

   A. High inertia (ϕ) helps exploration because particles can overshoot local minima and thereby find novel places in search space.
   B. Local best PSO introduces extra exploration compared to Global best PSO.
   C. A strong contribution of the cognitive component (high c_1) helps exploration more than a strong contribution of the social component (high c_2).
   D. A non uniformly distributed random vector for the starting points of the particles helps exploration.

Multiple choice

1. (2 points) The main difference between clustering and classification is:
   A. clustering cannot be done with neural networks.
   B. clustering is unsupervised while classification is supervised learning.
   C. classification is a binary decision problem, while clustering can have more than two outcomes.
   D. there is no difference, apart from the name.

2. (2 points) A feature vector of length n, with each feature having possible values [1-10] (Natural numbers):
   A. Has 10^n possible states.
   B. Has n possible states.
   C. Has $n * 10$ possible states.
   D. Has n^{10} possible states.

3. (2 points) Which of the following statements about a single perceptron is FALSE:
   A. Training a binary classification task is *guaranteed* to converge if the two classes are linearly separable.
   B. A single perceptron can not learn the X-OR function.
   C. A single perceptron has |x| number of weights (|x| for the number of inputs).
   D. A single perceptron can be trained with backpropagation.

4. (2 points) The ultimate goal in Reinforcement Learning is to learn:
   A. A reward function.
   B. A value propagation mechanism.
   C. An action selection policy that optimizes reward.
   D. A world model (state transition model).

5. (2 points) In neural networks, if two neurons i and j are connected, their connection is associated with a weight w_{ij}. Suppose you have an *auto-associative memory*, designed to retrieve K n-dimensional memories. Then the number of weights (i.e. of the connections between neurons) in the network will be:
   A. nK
   B. $n + k$
   C. K^2.
   D. $n * (n - 1)/2$

6. (2 points) Suppose you are dealing now with an *auto-encoder* with n-dimensional input and K hidden neurons. Then the number of weights (i.e. of the connections between neurons) in the network will be:
   A. nK
   B. $2(nk)$
   C. $n + K$
   D. n^2.

7. (2 points) Suppose you have now yet another network, a *competitive learning one*, with an n-dimensional input and K output neurons. Then the number of weights (i.e. of the connections between neurons) in the network will be:
   A. nK
   B. $2(nk)$
   C. $n + K^2$
   D. $n^2 + K$.

TECHNISCHE UNIVERSITEIT DELFT
Faculteit Elektrotechniek, Wiskunde en Informatica

TU Delft

Exam
TI2736-A – Computational Intelligence

2-nov-2015, 13:30–16:00

- The complete exam has 12 multiple choice (MC) questions with a total of 24 points and 4 open questions with a total of 24 points.
- Regarding MC questions:
  - Each question has only one correct answer.
- Regarding open questions:
  - Write correct English or Dutch (English preferred!) and make sure your handwriting is readable.
  - Motivate your answers.
  - Irrelevant information may lead to less points.
- You may NOT use your book or any notes during this exam.
- You MAY use a calculator. Your Phone must be IN your bag.
- Every paper must have your name and study number, and the first paper must contain how many papers have been handed over in total.
- The material for the exam is as indicated in the course overview document and dealt with during the lectures.
- One exam can not cover all material. Do not conclude anything about what will or will not be tested at exams based on older exams.
- Total pages: 4.

Good luck!