Question 1.

a) Explain the main characteristics of Artificial Neural Networks?

Sol-
The ANN are designed so that they possess many desirable characteristics. These include:
- distributed representation and computation;
- learning capacity;
- generalization ability;
- adaptability;
- fault tolerance;
- massive parallelism

b) Explain the main differences between the von Neumann computer and the biological neural system with respect to memory, processing organization, computing mode, and reliability?

Sol-

<table>
<thead>
<tr>
<th>Processor</th>
<th>Complex</th>
<th>Simple</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-speed</td>
<td>Low-speed</td>
<td></td>
</tr>
<tr>
<td>One (or a few)</td>
<td>A large number</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Memory</th>
<th>Separate</th>
<th>Integrated into processor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Localized</td>
<td>Distributed</td>
<td></td>
</tr>
<tr>
<td>Noncontent addressable</td>
<td>Content addressable</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Computing</th>
<th>Centralized</th>
<th>Distributed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential</td>
<td>Parallel</td>
<td></td>
</tr>
<tr>
<td>Stored Programs</td>
<td>Self-learning</td>
<td></td>
</tr>
</tbody>
</table>

| Reliability     | Vulnerable       | Robust           |


c) Explain the main computational differences between the von Neumann computer and multilayer perceptron networks?

Sol-

With respect to the organization of the computation process the multilayer perceptron networks are parallel, noise resistant, while the von Neumann computer is serial, noise sensitive and operates with stored programs.

d) Describe briefly the basic types of neural network learning algorithms?

Sol-

There are four basic types of connectionist learning algorithms depending on the learning rules:
- Error-correction learning / rule
- Boltzmann learning / rule
- Hebbian learning / rule
- Competitive learning / rule

Question 2.

a) Explain the notions: axon, dendrite, and synapse?

Sol-
The axon is a protuberance that delivers the neuron's output to connections with other neurons. The dendrites are protuberances that with plenty of surface area facilitate the connections with the axons of other neurons.
At the terminals of the dendrite branches are the synapses. A synapse is a functional unit between two neurons: an axon strand of one neuron and a dendrite branch of another.

b) What are the formulae for the Widrow-Hoff rule (incremental gradient descent) for training single-layer networks? Explain the meaning of every term in the formulae.
The formulae of the Widrow-Hoff rule (incremental gradient descent) are:

\[ w_i^{n+1} = w_i^n + \eta^* (y_e - o_e) x_{i|E} \]

where \( \eta^* (y_e - o_e) x_{i|E} \) is the weight update computed after each example \( x_{i|E}, 1 \leq i \leq N \)

\( y_e \) is the desired target outcome and \( o_e \) is the network output.

Consider a single-layer network (perceptron) with two inputs, having weights \( w_1 = 0.5 \) and \( w_2 = -0.25 \). Determine the output of the network when given an input vector \( (1,1) \) for each of the following activation functions:

i) linear activation function without threshold
ii) sigmoidal activation function
iii) thresholded activation function using threshold 0.25

The weighted sum computed at the neuron output is:

\[ o = 0.5 \times 1 + (-0.25) \times 1 = 0.25 \]

i) In case of a linear activation function without threshold: Output \( o = 0.25 > 0 \) therefore 1

ii) In case of a sigmoidal activation function: Output \( o = 1 / (1 + \exp(-0.25 \times 1)) = 0.56 \)

iii) In case of a thresholded activation function using threshold 0.25:
\( o = 0.25 + 0.25 = 0.5 \) therefore 1
\( o = 0.25 - 0.25 = 0 \) therefore 0

Question 3.

a) What is the formula for the straight line function learned by single-layer networks with two inputs and a third thresholding link with input always fixed at 1?

The straight line function learned by single-layer networks is:

\[ w_1 x_1 + w_2 x_2 + w_0 x_0 = 0 \]

where \( w_i \) are the weights, and \( w_0 \) is a constant applied always to \( x_0 = 1 \).

b) What are the equations for computing the errors according to the backpropagation algorithm:

i) in the output layer of the multilayer perceptron?
   Explain the meanings of all terms in the equation.

ii) in the hidden layer of the multilayer perceptron?
   Explain the meanings of all terms in the equation.

c) Consider a multilayer neural with a stair-step threshold function. Suppose that you add a constant to all weights and thresholds. Will the learning behaviour change?

Yes, the learning algorithm behaviour will change since the weighted sum in this case is not modified proportionally to the threshold value (they could be modified proportionally.
d) What is the weight update rule with momentum parameter used in the backpropagation algorithm?

i) Explain the meaning of all the terms in this weight update rule.

ii) Describe briefly the three effects of the momentum parameter on the performance of the backpropagation algorithm?

Sol-

The weight update rule with momentum parameter used in the backpropagation algorithm is:

\[
\Delta w(t) = -\frac{\partial E}{\partial w(t)} + \alpha \Delta w(t-1)
\]

where: \( \alpha \) is the momentum

\( \Delta w(t-1) \) is the previous weight update

\( -\frac{\partial E}{\partial w(t)} \) is the error derivative with respect to the weight

The momentum has the following effects parameter on the performance of the backpropagation algorithm:

- it smooths the weight changes and suppresses cross-stitching, that is cancels side-to-side oscillations across the error valley;
- when all weight changes are all in the same direction the momentum amplifies the learning rate causing a faster convergence;
- enables to escape from small local minima on the error surface.