Question 1.

a) Explain the main characteristics of Artificial Neural Networks?

Sol- The ANN are designed so that they posses 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-

processor	complex	simple
	high-speed	low-speed
	one (or a few)	a large number
memory	separate	integrated into processor
	localized	distributed
	noncontent addressable	content addressable
computing	centralized	distributed
	sequential	parallel
	strored-programs	self-learning
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 resistent, 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.

Sol- The formulae of the Widrow-Hoff rule (incremental gradient descent) are:

$w_i = w_i + \eta * (y_e - o_e) * x_{ie}$

where $\eta * (y_e - o_e) * x_{ie}$ is the weight update computed after each example x_{ie}, $1 \le i \le N$

y_e is the desired target outcome and o_e is the network output.

c) Consider a single-layer network (perceptron) with two inputs, having weights w1=0.5 and w2=-.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

Sol- The weighted sum computed at the neuron output is:
o = 0.5 * 1 + (-0.25) * 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 * 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?

Sol- The straight line function learned by single-layer networks is:
w1x1 + w2x2 + w0x0 = 0 where wi are the weights, and w0 is a constant applied always to x0=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.

Sol- The errors at the output layer according to the backpropagation algorithm are computed as

$\delta_k = o_k (1 - o_k) (y_k - o_k)$

where: o_k is the output of the k-th output node, and y_k is the given true output

The errors at the hidden layer according to the backpropagation algorithm are computed as

$\delta_j = o_j (1 - o_j) \Sigma_k \delta_k w_{jk}$

where: o_j is the output of the j-th hidden node, δ_k is the error from the k-th output node, and w_{jk} is the weight on the connection between the j-th hidden node and the k-th output n

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?

Sol- 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

if the weights and the threshold are multiplied together by the same constant).

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:

$$\Delta w(t) = -\partial Ee / \partial w(t) + \alpha \, \Delta w(t-1)$$

where: α is the momentum

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

$-\partial Ee / \partial w(t)$ 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
accross the error valey;
- 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.