Q1-1: Select the correct option.

A. A perceptron is guaranteed to perfectly learn a given linearly well-separable function within a finite number of training steps.
B. A single perceptron can compute the XOR function.

1. Both statements are true.
2. Both statements are false.
3. Statement A is true, Statement B is false.
4. Statement B is true, Statement A is false.

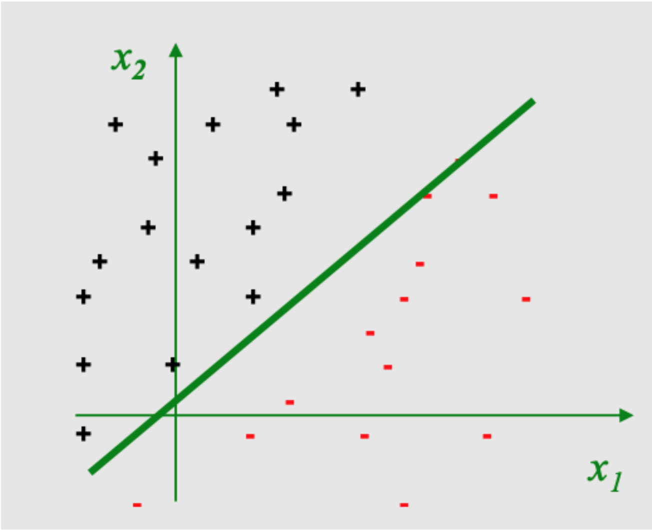
Q1-1: Select the correct option.

A. *A perceptron is guaranteed to perfectly learn a given linearly well-separable function within a finite number of training steps.*
B. *A single perceptron can compute the XOR function.*

1. Both statements are true.
2. Both statements are false.
3. Statement A is true, Statement B is false.
4. Statement B is true, Statement A is false.

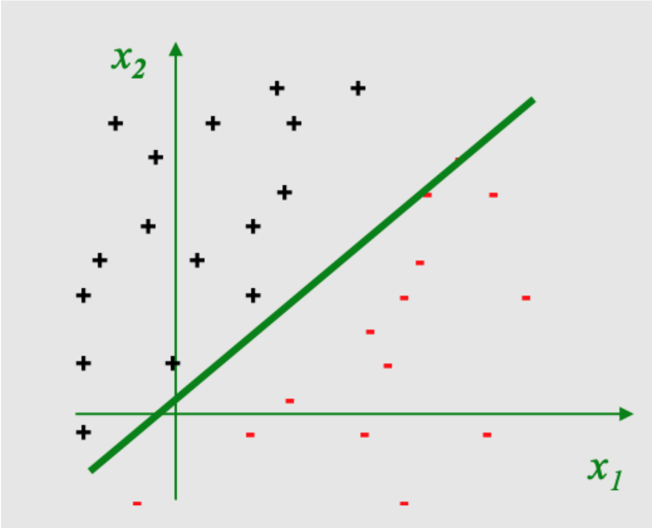
Q1-2: The decision boundary obtained by a perceptron for the given dataset is shown in green and is of following form: $w_0 + w_1x_1 + w_2x_2 = 0$. Which of the following set of values for $\{w_0, w_1, w_2\}$ can NOT depict the given boundary?

1. {-0.5, -1, 1}
2. {0.5, 1, -1}
3. {-0.5, -1, -1}
4. All of the above are valid candidates.



Q1-2: The decision boundary obtained by a perceptron for the given dataset is shown in green and is of following form: $w_0 + w_1x_1 + w_2x_2 = 0$. Which of the following set of values for $\{w_0, w_1, w_2\}$ can NOT depict the given boundary?

1. {-0.5, -1, 1}
2. {0.5, 1, -1}
3. {-0.5, -1, -1}
4. All of the above are valid candidates.



Equation of Line:
$x_2 = -(w_1/w_2)*x_1 - w_0/w_2$

From the figure, it can be inferred that slope > 0 and y-intercept > 0.
Slope: $-(w_1/w_2)$, y-intercept = $-w_0/w_2$

1. Slope = 1, y-intercept = 0.5 -- possible
2. Slope = 1, y-intercept = 0.5 -- possible
3. Slope = -1, y-intercept = -0.5 -- NOT possible

Q2-1: Select the correct option.

A. *The more hidden-layer units a Neural Network has, the better it can predict desired outputs for new inputs that it was not trained with.*
B. *A 3-layers Neural Network with 5 neurons in the input and hidden representations and 1 neuron in the output has a total of 55 connections.*

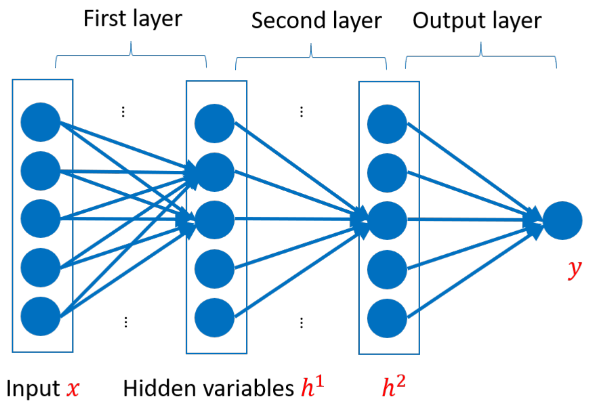
1. Both statements are true.
2. Both statements are false.
3. Statement A is true, Statement B is false.
4. Statement B is true, Statement A is false.

Q2-1: Select the correct option.

A. *The more hidden-layer units a Neural Network has, the better it can predict desired outputs for new inputs that it was not trained with.*

B. *A 3-layers Neural Network with 5 neurons in the input and hidden representations and 1 neuron in the output has a total of 55 connections.*

1. Both statements are true.
2. Both statements are false.
3. Statement A is true, Statement B is false.
4. Statement B is true, Statement A is false.



Q2-2: Select the correct option.

A. *The range of tanh activation function* $\tanh(z) = 2\sigma(2z) - 1$ *is (-1, 1).*

B. *Nominal features are usually represented using a thermometer encoding.*

1. True, True
2. True, False
3. False, True
4. False, False

Q2-2: Select the correct option.

A. *The range of tanh activation function* $\tanh(z) = 2\sigma(2z) - 1$ *is (-1, 1).*

B. *Nominal features are usually represented using a thermometer encoding.*

1. True, True
2. True, False
3. False, True
4. False, False

Nominal features are usually represented using a 1-of-k encoding.

Q3-1: Select the correct option.

A. *The backpropagation learning algorithm is based on the gradient-descent method.*

B. *In backpropagation learning, we usually start with a small learning parameter η and slowly increase it during the learning process.*

1. Both statements are true.
2. Both statements are false.
3. Statement A is true, Statement B is false.
4. Statement B is true, Statement A is false.

Q3-1: Select the correct option.

A. The backpropagation learning algorithm is based on the gradient-descent method.

B. In backpropagation learning, we usually start with a small learning parameter η and slowly increase it during the learning process.

1. Both statements are true.
2. Both statements are false.
3. Statement A is true, Statement B is false.
4. Statement B is true, Statement A is false.

Q3-2: Consider $w \in R$, the objective function to be minimized is the regularized loss $L(w) + \lambda w^2$. With $w_t = 1$, $dL(w_t) = 1$, step size $\eta = 0.1$, $\lambda = 2$, perform one step of gradient descent by computing the value of w_{t+1}. Hint: $w_{t+1} = w_t - \eta \nabla(L(w) + \lambda w^2)$

1. 0.5
2. 1
3. 1.5
4. 0.1

Q3-2: Consider $w \in R$, the objective function to be minimized is the regularized loss $L(w) + \lambda w^2$. With $w_t = 1$, $dL(w_t) = 1$, step size $\eta = 0.1$, $\lambda = 2$, perform one step of gradient descent by computing the value of w_{t+1}. Hint: $w_{t+1} = w_t - \eta \nabla(L(w) + \lambda w^2)$

1. 0.5
2. 1
3. 1.5
4. 0.1

$w_{t+1} = w_t - \eta d(L+\lambda w^2) = w_t - \eta dL(w_t) - \eta\lambda * 2w_t$
$= 1 - 0.1 - 0.1 \times 2 \times 2 = 1 - 0.1 - 0.4 = 0.5$