- **Question 1**

Consider 3-layers NN defined as sign (W'''(W''(W'x + b') b'') b''') where b are biases and W are weights and sign are the sign function. Which of the following is FALSE?
a. The 3-layer NN cannot solve the XOR problem.
b. Adding one more linear hidden layer will not increase the expressiveness of the model.
c. The hypothesis space of this model is the same as Perceptron.
d. The network cannot be a universal approximation.
e. none of above.

- **Question 2**

Let's consider a family of parametric probability distributions represented by a model $p_{model}(x;\theta)$ and the probability of the data of a task $\hat{p}_{data}$.Which of the following if TRUE?
a. Maximum likelihood assume uniform prior distribution.
b. Minimize KL between the two probabilities means maximizing the cross-entropy between the two probabilities.
c. Bayesian approach make predictions using a single point estimate of θ
d. Maximum likelihood male prediction using a full probability over θ
e. none of above

- **Question 3**

Consider a CNN layer with 2 filters of size 3x3, a stride of 1 and input images of size 4x4 and padding=" same" on a single channel (i.e. black and white). How many multiplications (between two numbers) are performed to compute the output (feature map) of such a layer? (do not consider bias terms or activation functions) Please answer with the exact number of the parameters (no formulas).

- **Question 4**

How can we appropriately select an appropriate number of hidden layers and the corresponding activation function in a DNN? Why is it crucial to accurately evaluate the model's performance and how can this be achieved?

- **Question 5**

What is a Graph Convolutional NN? Explain how it is formulated in graph spectral domain and in the node domain.

- **Question 6**

$$h^{(1)} = W^{(1)}x + b^{(1)}$$

$$x = \begin{bmatrix} 1 \\ 0 \\ 0.5 \end{bmatrix}, W^{(1)} = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 0 & -1 \end{bmatrix},$$

$$a^{(1)} = \sigma\left(h^{(1)}\right)$$

$$b^{(1)} = \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}, \quad b^{(2)} = 0,$$

$$h^{(2)} = \left(\omega^{(2)}\right)^{T} a^{(1)} + b^{(2)}$$

$$w^{(2)} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad t = 1$$

$$y = \sigma\left(h^{(2)}\right)$$

$$J = \frac{1}{2}(t-y)^2$$

$$Compute\ \frac{\partial J}{\partial W^{(1)}}\ and\ \frac{\partial J}{\partial b^{(1)}}\ where\ \sigma(0) = 0.5$$