Question 1. (1 Mark) Which of these architectures would have the best chance of learning long range dependencies?

a) Simple Recurrent Network
b) Feedforward network with sliding window
c) Long Short Term Memory
d) Elman Network

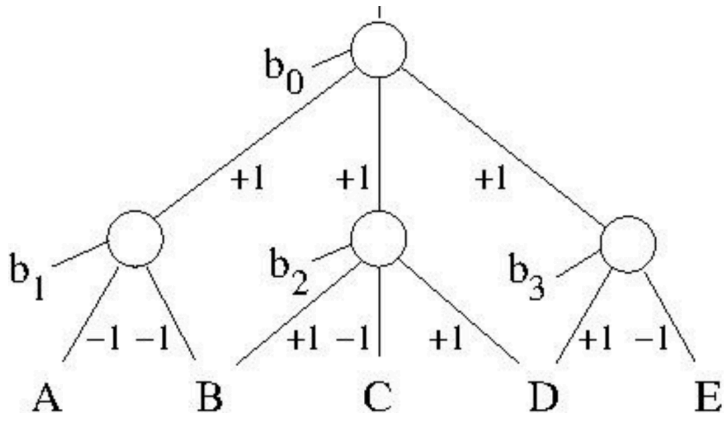
Question 2. (3 marks) Consider a Perceptron whose output is given by $h(w_0 + w_1x_1 + w_2x_2)$, where x_1, x_2 are inputs and $h()$ is the Heaviside (step) function.

Assume this Perceptron is being trained on the data in the following table, and that the current values of the weights w_0, w_1 and w_2 are -0.5, -1, and -2.

Training Example	x_1	x_2	Class
(a)	2	1	+1
(b)	-2	2	+1
(c)	-1	-1	-1

Suppose the Perceptron Learning Rule is applied to the current weights, using a learning rate of η=1.0, where only item (a) is learned. The new values for w_0, w_1 and w_2 at the end of this training step will be:

Question 3. (2 Marks) Consider the following multi-layer perceptron, using the threshold activation function, and assume that TRUE is represented by 1; FALSE by 0.



For what values of the biases would this network compute the logical function
(¬A ∨ ¬B) ∧ (B ∨ ¬C ∨ D) ∧ (D ∨ ¬E)?

Question 4. (2 Marks) Only 30% of the population have been vaccinated against a certain disease. Among those who are vaccinated, only 1% of them have the disease. But, among those who are not vaccinated, 3% of them have the disease.

If a random person is found to have the disease, what is the probability that they have been vaccinated?

Question 5. (3 Marks) Consider these two probability distributions on the same space Ω = {A, B, C, D, E}

p = ⟨ 1⁄2 , 1⁄4 , 1⁄8 , 1⁄16 , 1⁄16 ⟩
q = ⟨ 1⁄4 , 1⁄8 , 1⁄16 , 1⁄2 , 1⁄16 ⟩

Compute the Entropy H(p) and the KL-Divergence $D_{KL}(p||q)$ to at least 2 decimal places.

Question 6. (3 Marks) Consider a neural network trained using softmax for a classification task with three classes 1, 2, 3. Suppose a particular input is presented, producing outputs:

$z_1 = 2.5$, $z_2 = 1.5$, $z_3 = 0$.

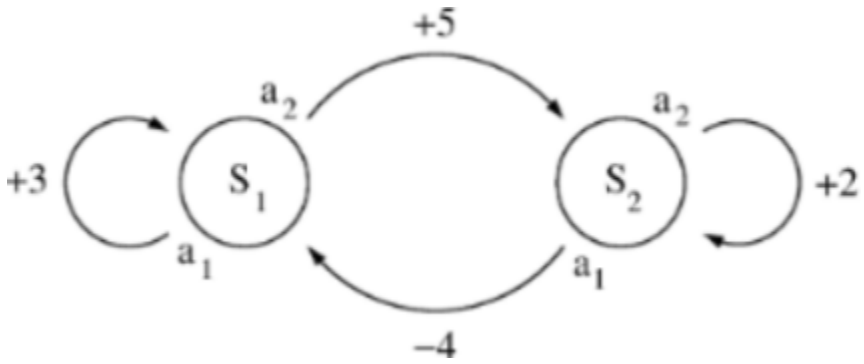
Assuming the correct class for this input is Class 2, and that Prob(2) is the softmax probability of the network choosing Class 2, and that $\log_e$ is the natural logarithm (ln), compute the following (correct to at least two decimal places):

$d(\log_e Prob(2))/dz_1 = -0.69$
$d(\log_e Prob(2))/dz_2 = 0.75$
$d(\log_e Prob(2))/dz_3 = -0.06$

Question 7. (3 Marks) Consider a convolutional neural network which takes as input a 65 x 77 color image (i.e. with three channels R, G, B). The first convolutional layer has 18 filters that are 5-by-5, with stride 3 and no zero-padding. Compute the number of:

weights per filter in this layer (including bias):
neurons in this layer:
connections into the neurons in this layer:
independent parameters in this layer:

Question 8. (8 Marks) Consider an environment with two states S={S_1, S_2} and two actions A={a_1, a_2}, where the deterministic transitions δ and reward R for each state and action are as follows:



With a discount factor of $\gamma = 0.7$, the optimal policy π for this environment is given by:

$\pi(S_1) = a_1$, $\pi(S_2) = a_2$

Using the above policy, and assuming $\gamma = 0.7$, compute these values (correct to 2 decimal places).

$Q^{\pi}(S_1, a_1)$ =
$Q^{\pi}(S_1, a_2)$ =
$Q^{\pi}(S_2, a_1)$ =
$Q^{\pi}(S_2, a_2)$ =

If γ is allowed to vary between 0 and 1, for which range of values of γ is this policy optimal?

Minimum value of γ =

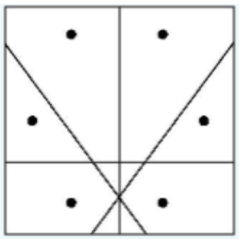
Maximum value of γ =

Question 9. (3 Marks) Consider a fully connected feedforward neural network with 6 inputs, 2 hidden units and 4 outputs, using tanh activation at the hidden units and sigmoid at the outputs. Suppose this network is trained on the following data, and that training is successful.

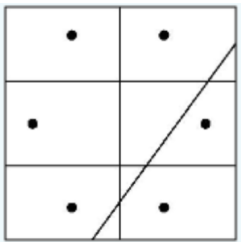
Item	Inputs	Outputs
	123456	1234
1	100000	0001
2	010000	0011
3	001000	0100
4	000100	1010
5	000010	1011
6	000001	1110

Which of these diagrams correctly shows a point in hidden unit space corresponding to each input, and for each output, a line dividing the hidden unit space into regions for which the value of that output is greater/less than one half?

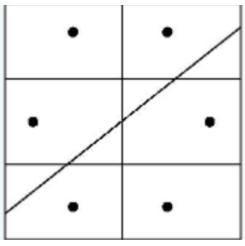
a)



b)



c)



d)