Name: _____

Problem	Points	Score
1(a)	10	
1(b)	10	
1(c)	10	
2	20	
3(a)	5	
3(b)	5	
3(c)	5	
3(d)	10	
4(a)	15	
4(b)	10	
Total	100	

Notes:

(1) The exam is closed books and notes. You are allowed on double-sides 8.5x11 in. page of notes. If you don't have an equation that you think you need, please ask me during the exam.

(2) Please clearly indicate your answer to the problem.

(3) Note that ungrammatical sentences, incoherent statements, or general illegible scratches will get zero credit.

(4) If I can't read or follow your solution, it is wrong, and no partial credit will be awarded.

Problem No. 1: Consider the following binary classification problem. The training set consists of patterns A and B in all possible translations, with wraparound. The dark cells are labeled as class 1 and the light cells are labeled as class 2.

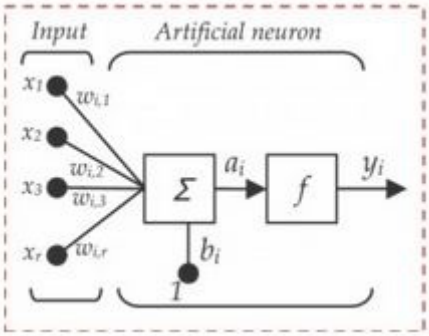
(a) (10) Can a linear classifier solve this problem, such as QDA, solve this problem? Explain.

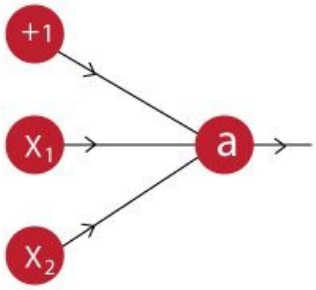
(b) (10) Consider a neural network that consists of a 1D convolution layer with a linear activation function, followed by a linear layer with a logistic output. Can such an architecture perfectly classify all of the training examples? Explain.

(c) (10) What type of network architecture could solve this problem? Explain why being as specific as possible.

Problem No. 2: (20) Given the following table, and the neural network shown, design the weights so that the network can reproduce the mapping.

X1	X2	X1 AND X2
0	0	0
0	1	0
1	0	0
1	1	1





$$a = f\left(\sum_{i=0}^{N} w_i x_i\right)$$

$$f(x) = \begin{cases} 0, & for\ x < 0 \\ 1, & for\ x \geq 0 \end{cases}$$

Problem 3: Explain the role of the following algorithms and how they might apply to learning issues such as overfitting, regularization and generalization.

(a) (5) Splitting the data in /train, /dev and /eval.

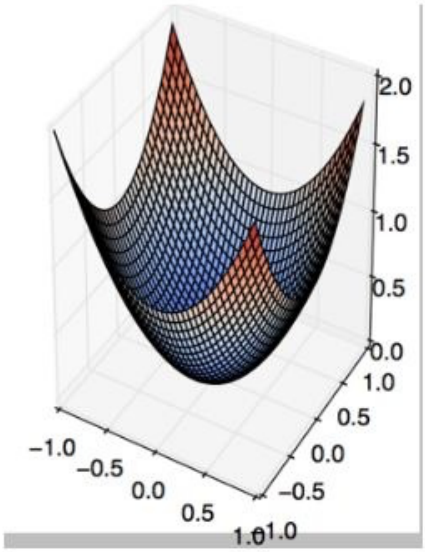
(b) (5) Introducing max pooling into a CNN.

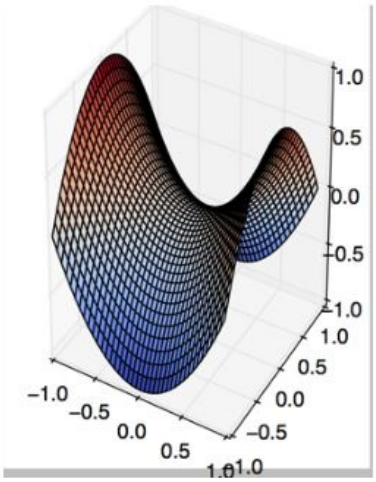
(c) (5) Introducing dropouts into a deep network.

(d) (10) You are designing a deep learning system to detect driver fatigue in cars. It is crucial that that your model detects fatigue, to prevent any accidents. Consider a simple related problem – stop sign detection. You run an experiment and generate a confusion matrix (2x2). Explain how each component of this matrix reflects the challenges of this problem. Which element do you think is he most important?

Problem No. 4: Discuss the following properties of networks.

(a) (15) First-order gradient descent will obviously work on the first surface to the right. Will it work on the second surface? Explain. If you think it would not work on the second surface, how would you modify it to work?





(b) (10) Suppose we want to build a machine that replicates the decision surface in the first graphic using the inputs from the second graphic. What logic would that machine need to implement?