Homework 04

Submission Notices:

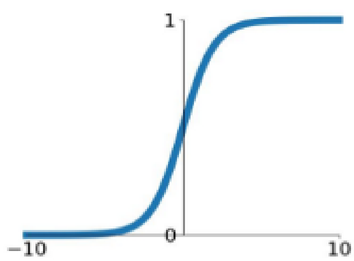
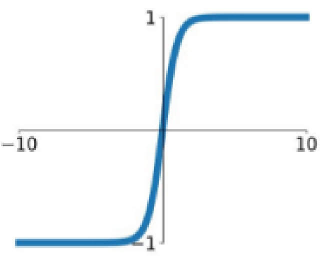
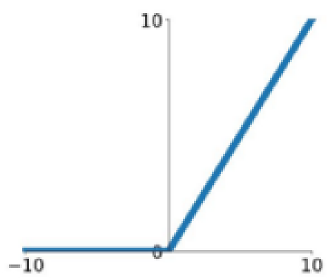
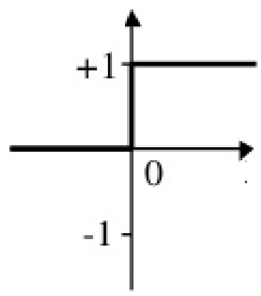
- *Conduct your homework by filling answers into the placeholders in this file (in Microsoft Word format). Questions are shown in black color, instructions/hints are shown in italics and blue color, and your content should use any color that is different from those.*
- *After completing your homework, prepare the file for submission by exporting the Word file (filled with answers) to a PDF file, whose filename follows the following format,*

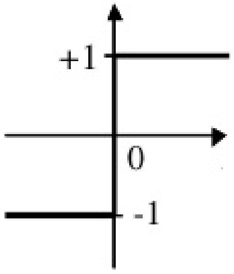
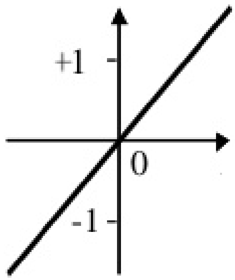
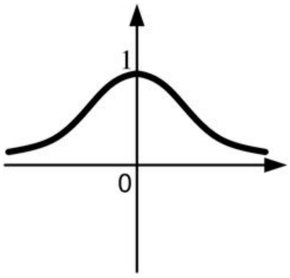
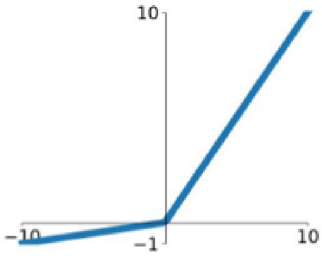
  <StudentID-1>_<StudentID-2>_HW01.pdf (Student IDs are sorted in ascending order)

  E.g., ***2112001_2112002_HW02.pdf***

  and then submit the file to Moodle directly WITHOUT any kinds of compression (.zip, .rar, .tar, etc.).
- *Note that you will get zero credit for any careless mistake, including, but not limited to, the following things.*
  1. *Wrong file/filename format, e.g., not a pdf file, use "-" instead of "_" for separators, etc.*
  2. *Disorder format of problems and answers*
  3. *Conducted not in English*
  4. *Cheating, i.e., copying other students' works or letting other students copy your work.*

Problem 1. (2pts) Identify each of the following activation functions.

#	2-D representation	#	2-D representation
a		b	
c		d	

e		f	
g		h	

Please fill your answer in the table below

#	a	b	c	d
Function name	Sigmoid	Tanh	ReLU (Rectified Linear Unit)	Binary Step
#	e	f	g	h
Function name	Sign	Linear Activation	Gaussian	Leaky ReLU

Problem 2. (1pt) Present two objective metrics that can be used to evaluate the attributes for a node on the decision tree. For each metric, you need to present the formula, identify its domain (i.e., range of values), and explain for every term in the formula.

Please fill your answer in the table below

Metric name	Formula	Explanation
Information Gain	$IG(a) = H(S) - \sum_{i=1}^{m} p_i H(S \mid a = a_i)$	H(S): Entropy of the entire dataset S. pi: Proportion of the total number of instances that belong to the ith subset S_ai. H(S|a = ai): Entropy of the ith subset S_ai.

Gini Impurity	1 - ∑(p_i)^2	p_i: Proportion of the total number of instances that belong to the ith class.

Problem 3. (2pts) You are given the following tables, which represent the outcomes of some functions. The functions take two values x and y and output the outcomes of the operations. Please identify **at least two models** for each of the functions that are perfectly represent the functions for some choice of parameters. Justify your answer. Note: there are no constraints on the architecture (e.g, the number of neurons, activation function, or the best splitting criterion), and the depth of decision tree is 0-index.

a) (1pt) $f(x,\ y)\ =\ x \oplus y$

x	y	$x \oplus y$
0	0	**0**
0	1	**1**
1	0	**1**
1	1	**0**

☐ A neural network with no hidden layer

☐ A neural network with a single hidden layer

☐ A decision tree of depth one

☐ A decision tree of depth two

Explanation: ....

1. A neural network with no hidden layer

   A single-layer perceptron with two input neurons and one output neuron. The perceptron takes in the two input values x and y, and the output is calculated by taking the exclusive OR of the inputs. The weights and bias of the perceptron can be set as follows:

   Weight of the first input neuron is set to 1.

   Weight of the second input neuron is set to 1.

   The bias is set to -1/2.

The output of the perceptron is then passed through a step function, which outputs 1 if the input is positive and 0 otherwise. This model can perfectly represent the XOR function, as it can output the correct output for any input combination of x and y.

2. A decision tree of depth one

   We can split the input space based on the value of x, and then output the XOR of the values of y for each split. For example, if x = 0, we output y ⊕ 0, and if x = 1, we output y ⊕ 1.

b) (1pt) $f(x, y) = \neg(x \lor y)$

x	y	$\neg(x \lor y)$
0	0	**1**
0	1	**0**
1	0	**0**
1	1	**0**

☐ A neural network with no hidden layer

☐ A neural network with a single hidden layer

☐ A decision tree of depth one

☐ A decision tree of depth two

Explanation: ……………………….

1. For a decision tree of depth 1, we have one root node that tests for either x or y being true, and two leaf nodes representing the output values for each case. Specifically, if x or y is true, the output will be 0, and if x and y are both false, the output will be 1.
2. For a decision tree of depth 2. we have one root node that tests for either x or y being true, and two child nodes representing the two possible outcomes. Each child node would then have another node that tests for the other input being false or true, respectively, and the corresponding output values at the leaf nodes. Specifically, if x or y is true, the first child node will output 0, and the second child node will test for the other input, with the corresponding output. If x and y are both false, the first child node will test for either input being false, and the second child node will output 1.

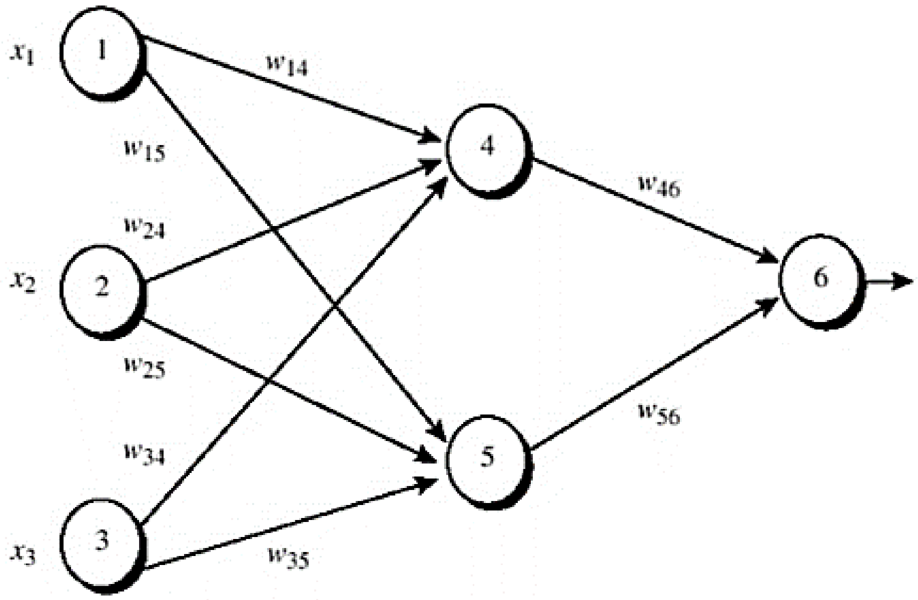
Problem 4. (2pts) Consider the following training dataset, in which **Transportation** is the target attribute. Show calculations to choose an attribute for the **root node** of the ID3 decision tree

Gender	Car Ownership	Travel Cost	Income Level	Transportation
Male	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Female	1	Cheap	Medium	Train
Female	0	Cheap	Low	Bus
Male	1	Cheap	Medium	Bus
Male	0	Standard	Medium	Train
Female	1	Standard	Medium	Train
Female	1	Expensive	High	Car
Male	2	Expensive	Medium	Car
Female	2	Expensive	High	Car

Please fill your answer in the table below

		Counts			Metric values		
	Attribute values	Bus	Car	Train	H	AE	IG
Whole		4	3	3	1.571		
Gender (0.5pt)	Female	1	2	2	1.522	1.447	0.125
	Male	3	1	1	1.371		
Car Ownership (0.5pt)	0	2	0	1	0.918	1.0364	0,535
	1	2	1	2	1.522		
	2	0	2	0	0		
Travel Cost (0.5pt)	Cheap	4	0	1	0.722	0.361	1.21
	Expensive	0	3	0	0		
	Standard	0	0	2	0		
Income Level (0.5pt)	Low	2	0	0	0	0.875	0.696
	Medium	2	1	3	1.459		
	High	0	2	0	0		

Problem 5. (3pts) Consider the following neuron network, which includes 3 input neurons, 2 hidden neurons and 1 output neurons.



Initial input, weight and bias values are

x_1	x_2	x_3	w_{14}	w_{15}	w_{24}	w_{25}	w_{34}	w_{35}	w_{46}	w_{56}	θ_4	θ_5	θ_6
1	0	1	0.2	–0.3	0.4	0.1	–0.5	0.2	–0.3	–0.2	–0.4	0.2	0.1

The expected output value is 1. The learning rate is 0.9.
Knowing that the actual output at some neuron j is calculated as follows.

$$y_j(p) = sigmoid\left[\sum_{i=1}^{n} x_i(p) \times w_{ij}(p) + \theta_j\right]$$

where n is the number of inputs of neuron j, w_{ij} is the corresponding link from a neuron i in the previous layer to neuron j, and θ_j is the bias at neuron j.

Present all calculations required to perform the backpropagation once (i.e., one forward pass and one backward pass) on the given neural network in the following cases

a) Ignore all biases *(precision to 3 decimal places).*

(0.25pt) Ignore all biases – Forward

Neuron	4	5	6
Output	0.426	0.475	0.445

(1pt) Ignore all biases – Backward

Weight	w_{46}	w_{56}	w_{14}	w_{15}	w_{24}	w_{25}	w_{34}	w_{35}
Value	-0.247	-0.141	0.191	-0.222	0.4	0.1	-0.509	0.278

b) Consider all biases such that each bias is treated as a neuron and thus it will be also updated *(precision to 3 decimal places).*

(0.25pt) Consider all biases – Forward

Neuron	4	5	6
Output	0.332	0.535	0.474

(1.5pt) Consider all biases – Backward

Weight	w_{46}	w_{56}	w_{14}	w_{15}	w_{24}	w_{25}	w_{34}	w_{35}	θ_4	θ_5	θ_6
Value	-0.261	-0.138	0.192	-0.306	0.400	0.100	-0.508	0.194	-0.408	0.194	0.218