Math Homework 2

CS540

May 30, 2019

1 Instruction

Please submit your answers on Canvas → Assignments → M2. Late submission will not be accepted. Please add a file named "comments.txt", and in the first line of the file, grade yourself: $1, 1.5, 2$ (for the entire homework, not for individual questions). In your submission, please do not write your name if you do not want other students to see it (in the case it is posted as a sample solution).

Grade	Meaning
1	You attempted something but mostly incorrect.
1.5	You attempted something but there are mistakes.
2	You have the correct answers + permission to post as a sample solution.

2 Questions

2.1 Question 1

Similar to 2011 Oct Midterm Q10

Given the following sample for functions AND, OR, NOT, NAND. Find the perceptron among the four choices that represent each function.

x_1	x_2	y for AND	y for OR	y for NOT	y for NAND
0	0	0	0	1	1
0	1	0	1	1	1
1	0	0	1	0	1
1	1	1	1	0	0

- $\hat{y} = \mathbb{1}_{\{1x_1+1x_2-1.5\geqslant 0\}}$
- $\hat{y} = \mathbb{1}_{\{1x_1+1x_2-0.5\geqslant 0\}}$
- $\hat{y} = \mathbb{1}_{\{-1x_1+0.5\geqslant 0\}}$
- $\hat{y} = \mathbb{1}_{\{-1x_1-1x_2+1.5\geqslant 0\}}$

2.2 Question 2

2010 Fall Final Q17

Show that a neural network with the following weights represent XOR.

$$w_{11}^{(1)} = 1, w_{12}^{(1)} = -10, b_1^{(1)} = -0.5$$

$$w_{21}^{(1)} = -1, w_{22}^{(1)} = 10, b_2^{(1)} = -5$$
$$w_{11}^{(2)} = 1, w_{21}^{(2)} = 1, b_1^{(2)} = -0.5$$

The perceptrons are LTUs. The notation $w_{ij}^{(l)}$ represents the weight in layer l from unit i in the previous layer to unit j in the next layer.
Hint: compute the outputs and show that they are identical to the one for XOR.

x_1	x_2	y from neural network	XOR
0	0	?	0
0	1	?	1
1	0	?	1
1	1	?	0

2.3 Question 3

Given an L layer neural network with $m^{(l)}$ units in layer l and $m^{(L)} = 1$ unit in layer L and $m^{(0)} = m$ input units, prove the following gradient formula. All units have sigmoid (logistic) activation functions, and the cost is the sum of squared errors.

$$\frac{\partial C}{\partial w_{j'j}^{(l)}} = \sum_{i=1}^{n} \delta_{ij}^{(l)} a_{ij'}^{(l-1)}, l = 1, 2, ..., L$$

where,

$$\delta_{i1}^{(L)} = (y_i - a_i)\, a_i\, (1 - a_i)$$
$$\delta_{ij}^{(l)} = \sum_{j'=1}^{m^{(l+1)}} \delta_{j'}^{(l+1)} w_{jj'}^{(l+1)} a_{ij}^{(l)} \left(1 - a_{ij}^{(l)}\right), l = 1, 2, ..., L - 1$$

Hint: when proving the formula for layer l, you can assume the formula holds for all weights in layer $l + 1$. This is called proof by induction.

2.4 Question 4

March 2018 Midterm Q9
Consider the majority classifier that predict $\hat{y}$ = mode of the training data labels (this means the classifier predicts either 0 or 1 whichever appears more times in the training set labels).

1. What is the 2-fold cross validation accuracy (percentage of correct classification) on the following training set?
2. What is the 5-fold cross validation accuracy for this training set?
3. What is the leave one out cross validation accuracy for this training set?

x	1	2	3	4	5	6	7	8	9	10
y	1	1	0	1	1	0	0	1	0	0