MLAI Week 6 Exercise: Logistic regression & PyTorch for deep learning

Note: An indicative mark is in front of each question. The total mark is 12. You may mark your own work when we release the solutions.

$\boxed{3}$ 1. Figure 1 shows the COVID test results at a centre:
1) convert it to a probability table following a similar example in Lecture 6
2) calculate the observed odds of COVID positive for the age group of 20–29.

Age	COVID	Age	COVID	Age	COVID
9	1	41	0	54	1
10	1	42	1	55	0
15	0	46	0	58	1
17	1	47	1	60	1
23	1	48	1	60	0
25	1	49	1	62	1
28	0	49	0	65	0
30	0	50	1	67	1
33	0	51	1	71	1
33	1	51	0	77	1
38	0	52	1	81	1

Figure 1: Age and COVID test results: 0 = negative, 1 = positive.

Solution:

Conversion to probability table [2 marks].

		COVID Positive	
Age Group	# in Group	#	%
0-9	1	1	100
10-19	3	2	67
20-29	3	2	67
30-39	4	1	25
40-49	7	4	57
50-59	7	5	71
60-69	5	3	60
70-79	2	2	100
80-89	1	1	100

Compute the odds [1 mark].

Using the probability computed in the table above for the 20-29 age group of $\frac{2}{3}$, we can compute the odds as follows:

$$
\begin{aligned}
\text{Odds} &= \frac{\pi}{1-\pi} \\
&= \frac{\frac{2}{3}}{1-\frac{2}{3}} \\
&= 2
\end{aligned}
\tag{1}
$$

$\boxed{2}$ 2. Derive π from $\log \frac{\pi}{1-\pi} = \mathbf{w}^\top \mathbf{x}$ (slide 22), i.e. derive the logistic function from the logit function.

Solution:

starting at the expression of the logit function,

$$
log\left(\frac{\pi}{1-\pi}\right) = \mathbf{w^T x} \tag{2}
$$

taking the exponent of both sides

$$
\frac{\pi}{1-\pi} = e^{\mathbf{w^T x}} \tag{3}
$$

also,

$$
\begin{aligned}
\frac{1-\pi}{\pi} &= e^{-\mathbf{w^T x}} \\
\frac{1}{\pi} - 1 &= e^{-\mathbf{w^T x}} \\
\frac{1}{\pi} &= 1 + e^{-\mathbf{w^T x}} \\
\pi &= \frac{1}{1+e^{-\mathbf{w^T x}}}
\end{aligned}
\tag{4}
$$

$\boxed{2}$ 3. The last equation on slide 23 writes the log likelihood in terms of π_i. Rewrite the equation in terms of the weight vector $\mathbf{w}$ and input vector $\mathbf{x}$.

Solution:

$$
\log P(y \mid X) = \sum_{i=1}^{n} \log P(y_i \mid x_i) \tag{5}
$$

$$
= \sum_{i=1}^{n} y_i \log\left(\frac{1}{1+e^{-\mathbf{w^T x_i}}}\right) + \sum_{i=1}^{n}(1-y_i)\log\left(1 - \frac{1}{1+e^{-\mathbf{w^T x_i}}}\right) \tag{6}
$$

$$
= \sum_{i=1}^{n} y_i \log\left(\frac{1}{1+e^{-\mathbf{w^T x_i}}}\right) + \sum_{i=1}^{n}(1-y_i)\log\left(\frac{1+e^{-\mathbf{w^T x_i}}-1}{1+e^{-\mathbf{w^T x_i}}}\right) \tag{7}
$$

$$\log P(y \mid X) = \sum_{i=1}^{n} y_i \log\left(\frac{1}{1+e^{-\mathbf{w}^{\mathbf{T}}\mathbf{x_i}}}\right) + \sum_{i=1}^{n}(1-y_i)\log\left(\frac{1}{e^{\mathbf{w}^{\mathbf{T}}\mathbf{x}}+1}\right) \tag{8}$$

2 4. In a binary (two-class) logistic regression model, the weight vector $\mathbf{w}$ = [4, -2, 5, -3, 11, 9]. We apply it to some object that we'd like to classify; the vectorized feature representation of this object is $\mathbf{x}$ = [6, 8, 2, 7, -3, 5]. What is the probability, according to the model, that this instance belongs to the positive class?

Solution:

We can compute this probability using the following expression,

$$P(y = 1|x) = \frac{1}{1+e^{-\mathbf{w}^{\mathbf{T}}\mathbf{x}}}. \tag{9}$$

This first requires computation of $\mathbf{w}^{\mathbf{T}}\mathbf{x}$, which in this case is,

$$\begin{aligned}\mathbf{w}^{\mathbf{T}}\mathbf{x} &= (4\times 6) + (-2\times 8) + (5\times 2) + (-3\times 7) + (11\times -3) + (9\times 5) \\ &= 9.\end{aligned} \tag{10}$$

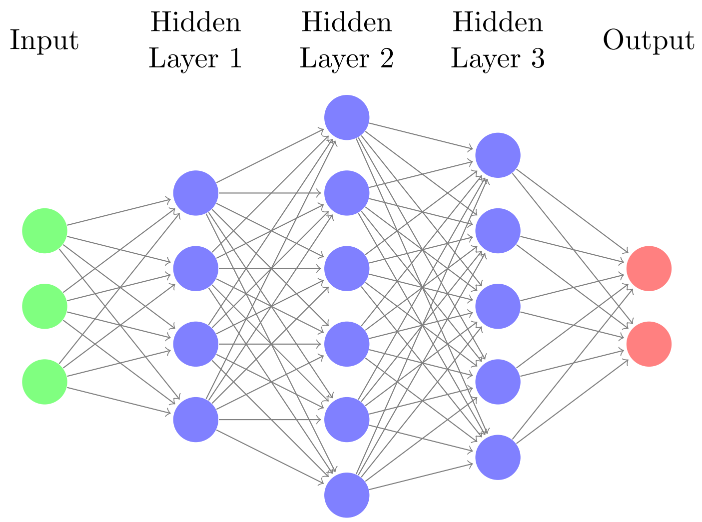
We can then substitute this value into (9) to obtain the final answer (given to 5 significant figures),

$$\begin{aligned}P(y = 1|x) &= \frac{1}{1+e^{-9}} \\ &= 0.99988.\end{aligned} \tag{11}$$

3 5. Consider the fully connected neural network (multilayer perceptron) on slide 33. If we insert two new hidden layers between the old Hidden Layer (4 neurons) and the Output Layer (2 neurons), i.e., New Layer 1 (6 neurons) after old Hidden layer, New Layer 2 (5 neurons) after New layer 1, and Output Layer after New Layer 2, with full connections between all adjacent layers and no other connections. The same activation function sigma (sigmoid) is used in the new hidden layers. How many learnable parameters in total are there for this three-hidden-layer neural network?

Solution:

The neural network described in the question has the following structure:



Firstly we must count all of the weights which connect the layers of our model,

$$\begin{aligned}\text{Number of weights} &= (3 \times 4) + (4 \times 6) + (6 \times 5) + (5 \times 2) \\ &= 76.\end{aligned} \tag{12}$$

Next, we count up all of the bias parameters,

$$\begin{aligned}\text{Number of biases} &= 4 + 6 + 5 + 2 \\ &= 17.\end{aligned} \tag{13}$$

The sum of these two values is the total number of model parameters, therefore the answer is $76 + 17 = 93$.