Week 6 Exercise: Logistic regression & PyTorch for deep learning

Note: An indicative mark is in front of each question. The total mark is 10. You may mark your own work when we release the solutions.

[1] 1. From Table 2 on slide 15 of Lecture 6, what are the observed odds of diseased for the age group of 50–59?

> **Solution:**
>
> Number of positive outcomes (diseased): 4
> Number of negative outcomes (not diseased): 7-4=3
> Odds of diseased = Number of positive outcomes / Number of negative outcomes= 4/3 = 1.33

[2] 2. Derive π from $\log \frac{\pi}{1-\pi} = \mathbf{w}^\top \mathbf{x}$ (slide 21), i.e. derive the logistic function from the logit function.

> **Solution:**
>
> starting at the expression of the logit function,
>
> $$log\left(\frac{\pi}{1-\pi}\right) = \mathbf{w^T x} \tag{1}$$
>
> taking the exponent of both sides
>
> $$\frac{\pi}{1-\pi} = e^{\mathbf{w^T x}} \tag{2}$$
>
> also,
>
> $$\begin{aligned} \frac{1-\pi}{\pi} &= e^{-\mathbf{w^T x}} \\ \frac{1}{\pi} - 1 &= e^{-\mathbf{w^T x}} \\ \frac{1}{\pi} &= 1 + e^{-\mathbf{w^T x}} \\ \pi &= \frac{1}{1+e^{-\mathbf{w^T x}}} \end{aligned} \tag{3}$$

$\boxed{2}$ 3. The last equation on slide 22 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:

$$\begin{aligned} logP(y \mid X) &= \sum_{i=1}^{n} logP(y_i \mid x_i) && (4) \\ &= \sum_{i=1}^{n} log\left(\frac{1}{1+e^{-\mathbf{w^T x_i}}}\right) + \sum_{i=1}^{n}(1-y_i)log\left(1-\frac{1}{1+e^{-\mathbf{w^T x_i}}}\right) && (5) \\ &= \sum_{i=1}^{n} log\left(\frac{1}{1+e^{-\mathbf{w^T x_i}}}\right) + \sum_{i=1}^{n}(1-y_i)log\left(\frac{1+e^{-\mathbf{w^T x_i}}-1}{1+e^{-\mathbf{w^T x_i}}}\right) && (6) \end{aligned}$$

$$logP(y \mid X) = \sum_{i=1}^{n} log\left(\frac{1}{1+e^{-\mathbf{w^T x_i}}}\right) + \sum_{i=1}^{n}(1-y_i)log\left(\frac{1}{e^{\mathbf{w^T x}}+1}\right) \quad (7)$$

$\boxed{2}$ 4. In a binary (two-class) logistic regression model, the weight vector $\mathbf{w}$ = [2, 8, 3, -5, -1]. We apply it to some object that we'd like to classify; the vectorized feature representation of this object is $\mathbf{x}$ = [-3, 2, 6, 8, -9]. What is the probability, according to the model, that this instance belongs to the positive class?

Solution:

Simply, multiply the $\mathbf{w}$ vector with the object features in $\mathbf{x}$, this gives,

$$\begin{aligned} P(object = +\text{veclass}) &= \frac{1}{1+e^{-\mathbf{w^T x}}} && (8) \\ w^T x &= (2)(-3)+(8)(2)+(3)(6)+(-5)(8)+(-1)(-9) && (9) \\ &= -6+16+18-40+9 = -3 && (10) \\ P(object = +\text{veclass}) &= \frac{1}{1+e^{--3}} = 0.0474 = 4.7\% && (11) \end{aligned}$$

$\boxed{3}$ 5. Consider the fully connected neural network (multilayer perceptron) on slide 32. If we insert another new hidden layer with 9 neurons between the old Hidden layer (4 neurons) and the Output layer (2 neurons), with full connections between the old hidden layer and new hidden layer, and full connections between new hidden layer and output layer (no direct connections between the old hidden layer and the output layer). The same activation function sigma (sigmoid) is used in the new hidden layer. How many learnable parameters in total are there for this two-hidden-layer neural network?

Solution:

3x4 + 4x9 + 9x2 = 66 weights
4+9+2= 15 biases
total= 81