Assignment #2: COMP4434 Big Data Analytics

Due Date: 23:59pm, Monday, 8 Mar. 2021

Question 1 **[10 marks]**

(a). [5 point] Consider using linear regression for binary classification on the label {0, 1}. Here, we use a linear model

$$h_\theta(x) = \theta_1 x + \theta_0$$

and squared error loss $L = \frac{1}{2}(h_\theta(x) - y)^2$. The threshold of the prediction is set as 0.5, which means the prediction result is 1 if $h_\theta(x) \geq 0.5$ and 0 if $h_\theta(x) < 0.5$. However, this loss has the problem that it penalizes confident correct predictions, i.e., $h_\theta(x)$ is larger than 1 or less than 0. Some students try to fix this problem by using an absolute error loss $L = |h_\theta(x) - y|$. The question is: Will it fix the problem? Please answer the question and explain it. Furthermore, some other students try designing another loss function as follows

$$L = \begin{cases} \max(0, h_\theta(x)), & y = 0 \\ \cdots, & y = 1 \end{cases}.$$

Although it is not complete yet, if it is correct in principle, please complete it and explain how it can fix the problem. Otherwise, please explain the reason.

(b). [5 point] Consider the logistic regression model $h_\theta(x) = g(\theta^T x)$, trained using the binary cross entropy loss function, where $g(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function. Some students try modifying the original sigmoid function into the following one

$$g(z) = \frac{e^{-z}}{1+e^{-z}}.$$

The model would still be trained using the binary cross entropy loss. How would the model prediction rule, as well as the learnt model parameters θ, differ from conventional logistic regression? Please show your answer and explanation.

Question 2 **[20 marks]**

Consider using logistic regression for classification problems. Four 3-dimensional data points $(x_1, x_2, x_3)^i$ and the corresponding labels y^{i} are given as follows.

Data point	x_1	x_2	x_3	y
D1	-0.120	0.300	-0.010	1
D2	0.200	-0.030	-0.350	-1
D3	-0.370	0.250	0.070	-1
D4	-0.100	0.140	-0.520	1

The learning rate η is set as 0.2 and the initial parameter $\theta[0]$ is set as [-0.09, 0, -0.19, -0.21]. Please answer the following questions.

a) [5 point] Calculate the initial predicted label for each data point.

b) [10 point] Calculate the parameter in the first and second iterations, i.e., $\theta[1], \theta[2]$, by using gradient descent algorithm.

c) [5 point] Implement the gradient descent algorithm to update the parameters θ using python language. Please show the change trend diagram of loss function $J(\theta)$ in 50000 rounds and upload the source code file.

ps. For a) and b), the detailed calculation process is required and the intermediate and final results should be rounded to 3 decimal places.