Computer Vision

Assignment 3

Tongji University
School of Software Engineering

ID:	1754060
Name:	Zhe Zhang
Adviser:	Prof. Lin Zhang
Time:	Mon. 2-4 [1-17]
Email:	dbzdbz@tongji.edu.cn

1 Problem 1(Math)

In our lecture, we mentioned that for logistic regression, the cost function is,

$$J(\theta) = -\sum_{i=1}^{m} y_i \log\left(h_\theta\left(\boldsymbol{x}_i\right)\right) + \left(1 - y_i\right) \log\left(1 - h_\theta\left(\boldsymbol{x}_i\right)\right) \tag{1}$$

Please verify that the gradient of this cost function is

$$\nabla_\theta J(\boldsymbol{\theta}) = \sum_{i=1}^{m} \boldsymbol{x}_i \left(h_\theta\left(\boldsymbol{x}_i\right) - y_i\right) \tag{2}$$

Solution: According to the defination of gradient,

$$\nabla_\theta J(\theta) = \begin{bmatrix} \frac{\partial J(\theta)}{\partial \theta_1} \\ \frac{\partial J(\theta)}{\partial \theta_2} \\ \ldots \\ \frac{\partial J(\theta)}{\partial \theta_m} \end{bmatrix} \tag{3}$$

where x is the input vector, y is the ground-truth, θ is the parameter vector

For one item $\frac{\partial J(\theta)}{\partial \theta_j}$, the calculation process is showing below:

$$\begin{aligned} \frac{\partial J(\theta)}{\partial \theta_j} &= -\sum_{i=1}^{m} [y_i \frac{1}{h_\theta(x_i)} \frac{\partial h_\theta(x_i)}{\partial \theta_j} + (1 - y_i) \frac{-\frac{\partial h_\theta(x_i)}{\partial \theta_j}}{1 - h_\theta(x_i}] \\ &= -\sum_{i=1}^{m} [\frac{\partial h_{theta}(x_i)}{\partial \theta_j} (\frac{y_i}{h_\theta(x_i} + \frac{y_i - 1}{1 - h_\theta(x_i})] \\ &= \frac{h_\theta(x_i)}{1 - h_\theta(x_i)} e^{-\theta^T x} x_{ij} (y_i - h_\theta(x_i)) \\ &= x_{ij}(y_i - h_\theta(x_i)) \end{aligned} \tag{4}$$

During the calculation, I omitted part of the calculation, add them below:

$$\begin{aligned} h_\theta(x_i) &= \frac{1}{1 + exp(-\theta^T x_i)} \\ \frac{\partial h_\theta(x_i)}{\partial \theta_J} &= -\frac{e^{-\theta^T x}(-x_j)}{[1 + exp(-\theta^T x)]^2} \\ &= (h_\theta(x_i))^2 e^{-\theta^T x} x_{ij} \\ \frac{h_\theta(x_i)}{1 - h_\theta(x_i)} e^{-\theta^T x} &= \frac{\frac{1}{1+exp(-\theta^T x_i)}}{1 - \frac{1}{1+exp(-\theta^T x_i)}} exp(-\theta^T x_i) \\ &= \frac{\frac{1}{1+exp(-\theta^T x_i)}}{\frac{exp(-\theta^T x_i)}{1+exp(-\theta^T x_i)}} exp(-\theta^T x_i) \\ &= 1 \end{aligned}$$

From what have be mentioned above, $\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{m} (h_\theta(x_i) - y_i)x_{ij}$. Hence, the gradient of this cost function is $\nabla_\theta J(\boldsymbol{\theta}) = \sum_{i=1}^{m} \boldsymbol{x}_i \left(h_\theta\left(\boldsymbol{x}_i\right) - y_i\right)$.