THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY

CSIT6000G/MSBD5012: Machine Learning

Homework 1

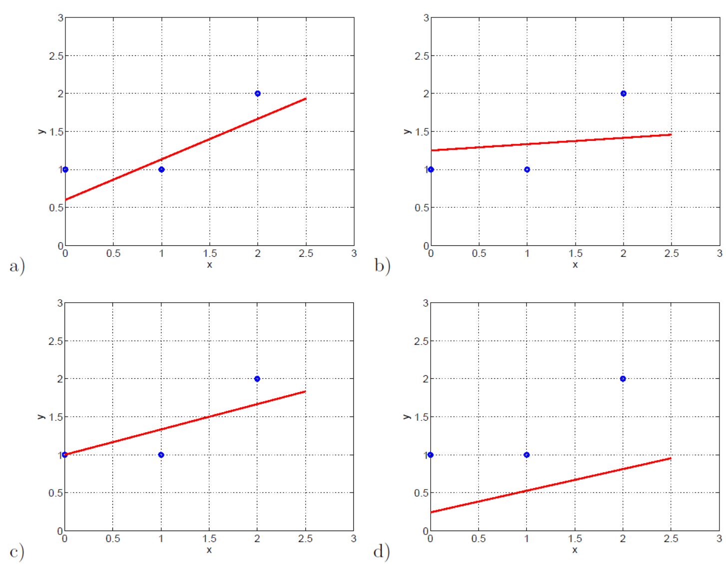
Assigned: 13/09/2019 (CSIT), 14/09/2019 (MSBD),
Due Date: 04/10/2019(CSIT), 05/10/2019 (MDBD),

To submit your work, hand it to the instructor on the due date.

Question 1 Consider carrying out linear regression on the following dataset. Manually compute the ordinary least squares solution.

x_1	0	0	1	1	1
x_2	1	1	1	0	0
y	0	1	2	3	4

Question 2 The following figures show linear regression results on a dataset of only three data points (marked blue).



The results were obtained using following regularization schemes:

1. $\frac{1}{3}\sum_{i=1}^{3}(y_i - w_0 - w_1 x_i)^2 + \lambda w_1^2$ where $\lambda = 1$.
2. $\frac{1}{3}\sum_{i=1}^{3}(y_i - w_0 - w_1 x_i)^2 + \lambda w_1^2$ where $\lambda = 10$.
3. $\frac{1}{3}\sum_{i=1}^{3}(y_i - w_0 - w_1 x_i)^2 + \lambda(w_0^2 + w_1^2)$ where $\lambda = 1$.
4. $\frac{1}{3}\sum_{i=1}^{3}(y_i - w_0 - w_1 x_i)^2 + \lambda(w_0^2 + w_1^2)$ where $\lambda = 10$.

Match the regularization schemes with the regress results. Briefly explain your answers.

Question 3 Consider applying logistic regression to the following dataset:

x_1	0	0	1	1
x_2	0	1	0	1
y	0	0	0	1

The target is to learn a model of the form $p(y = 1|\mathbf{x}, \mathbf{w}) = \sigma(w_0 + w_1 x_1 + w_2 x_2)$.

Suppose $w_0 = -2$, $w_1 = 1$ and $w_2 = 1$ initially and $\alpha = 0.1$. Manually run the batch gradient descent algorithm for one iteration. Give the weights and training error after the iteration.

Question 4 Consider applying logistic regression to the following dataset:

x_1	0	0	1	1
x_2	0	1	0	1
y	1	0	0	1

1. If we use raw feature x_1 and x_2, the model is

$$p(y = 1|\mathbf{x}, \mathbf{w}) = \sigma(w_0 + w_1 x_1 + w_2 x_2).$$

   What is the minimum achievable training error in this case? Give weights that achieve the minimum error.

2. Next consider using an additional feature $x_1 x_2$ in addition to the raw feature x_1 and x_2. The model now is

$$p(y = 1|\mathbf{x}, \mathbf{w}) = \sigma(w_0 + w_1 x1 + w_2 x_2 + w_3 x_1 x_2).$$

   What is the minimum achievable training error in this case? Give weights that achieve the minimum error.

Question 5 Consider the gradient vector in logistic regression $\nabla_{\mathbf{w}} NNL(\mathbf{w}) = (\frac{\partial NNL(\mathbf{w})}{\partial w_0}, \frac{\partial NNL(\mathbf{w})}{\partial w_1}, \ldots, \frac{\partial NNL(\mathbf{w})}{\partial w_D})$ where

$$\frac{\partial NNL(\mathbf{w})}{\partial w_i} = -\sum_{i=1}^{N} [y_i - \sigma(z_i)] x_{i,j}.$$

Suppose the feature x_1 is binary and, in the training set, it takes value 1 only in a small number of training examples with class label 1 (i.e., $y = 1$), and it takes value 0 in all training examples with class label 0 (i.e., $y = 0$). What will happen to the weight w_1 if we update it repeatedly using the following rule:

$$w_1 \leftarrow w_1 + \alpha \sum_{i=1}^{N} [y_i - \sigma(\mathbf{w}^\top \mathbf{x}_i)] x_{i,1}$$

What if we use the following update rule instead:

$$w_1 \leftarrow w_1 + \alpha[-\lambda w_1 + \sum_{i=1}^{N} [y_i - \sigma(\mathbf{w}^\top \mathbf{x}_i)] x_{i,1}],$$

where λ is the regularization constant?