King Mongkut's University of Technology Thonburi
Faculty of Engineering, Department of Computer Engineering
CPE 663 Special Topic: Deep Learning, 1/2019

Assignment 1: Gradient-based learning with regularization constraints

Due: 19 September 2019. Please submit your report in PDF to MyLE

Introduction

One of the main problems in machine learning is to select suitable features for the model to learn. Embedded learning is one of feature selection approaches that aim to select suitable features and at the same time fit the model with data. In this assignment, you will work on deriving and implementing the stochastic gradient descent optimization on the logistic regression with ridge regularization. Logistic regression with ridge regularization can be expressed in the following formula,

$$f\left(\mathbf{x},\mathbf{w}\right) = w_0 + \sum_{i=1}^{n} w_i x_i$$

$$H_{\mathbf{w}}\left(\mathbf{x},\mathbf{w}\right) = \frac{1}{1+e^{-f(\mathbf{x},\mathbf{w})}}$$

where **x** is the vector of input feature and **w** is the vector of the regression coefficient. Error function for ridge regression can be expressed in the form of

$$E(\mathbf{w}) = L(\mathbf{w}) + \lambda \left|\left|\mathbf{w}\right|\right|_2$$

$$= L(\mathbf{w}) + \lambda \sum_{i=1}^{n} w_i^2$$

$L(\mathbf{w})$ is the loss/error function of the model and λ is a regularization parameter. Ridge regularization will penalize the use of weight and the outcome of the optimization will be a fitted model with minimal weight.

For demonstration, we will use a very simple iris dataset. The report needs to address the following problems/tasks.

Iris dataset: https://www.kaggle.com/arshid/iris-flower-dataset

Tasks

1. Prepare the data in one-against-the-rest strategy. This can be done by converting the "Species" column into 3 binary columns.
2. Formulate the error function of the logistic regression with ridge regularization criterion.
3. Derive the gradient of the error function by deriving the partial derivative of the error function in Task 2.
4. Implement the gradient descent using all of the dataset in each iteration.
5. Implement the stochastic gradient descent using the subset of dataset in each iteration.
6. Test to see the effect of λ.
7. Test to see the effect of sampling proportion in Task 5.