ECE421S – Introduction to Machine Learning

Assignment 1

Linear and Logistic Regression

**New Deadline:** **February 6, 2019 @ BA3014, 4:00-5:00 PM EST**

Code Submission: ece421ta2019@gmail.com **February 6, 2019 @ 5:00 PM EST**

General Notes:

* Attach this cover page to your hard copy submission
* For assignment related questions, please contact Matthew Wong (matthewck.wong@mail.utoronto.ca)
* For general questions regarding Python or Tensorflow, please contact Tianrui Xiao (tianrui.xiao@mail.utoronto.ca) or see him in person in his office hours, Tuesdays, 4:00-6:00 PM in BA-3128 (Robotics Lab)

Please circle section to which you would like the assignment returned

Tutorial Sections

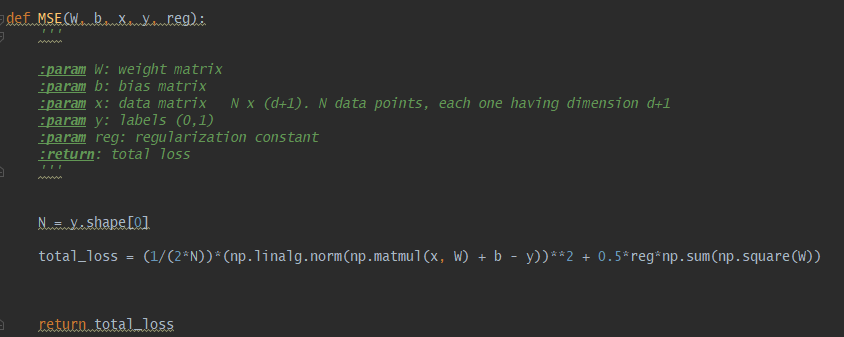
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 001 |  |  | 002 | 003 |  | 004 |
| 005 |  |  | 006 | 007 |  | Graduate |

|  |  |
| --- | --- |
| Group Members | |
| Names | StudentID |
| **Andy Linzi Zhou** | **1002296730** |
| **Ryan Do** | **1001254117** |

# Linear Regression

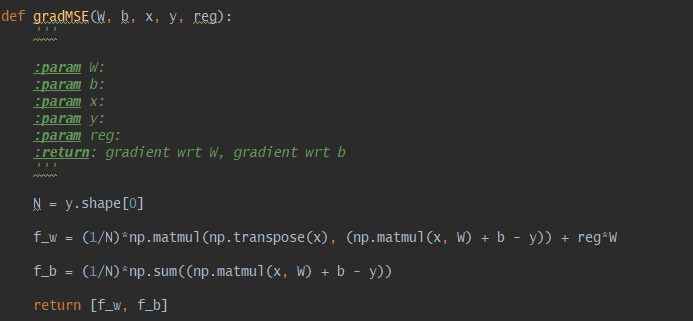
## **Loss Function and Gradient**

***MSE Loss Equation, Returns total loss***



***Gradient with respect to weights***

***Gradient with respect to bias scalar***



## **Tuning the learning rate**

|  |  |  |  |
| --- | --- | --- | --- |
| **Learning Rate α** | **Loss Curves** | **Training Time** | **Final Classification Accuracy** |
| **0.005** |  | 00:00:36 | Training Accuracy: 0.68371  Validation Accuracy: 0.67000  Test Accuracy: 0.64138 |
| **0.001** |  | 00:00:32 | Training Accuracy: 0.67429  Validation Accuracy: 0.65000  Test Accuracy: 0.66207 |
| **0.0001** |  | 00:00:33 | Training Accuracy: 0.69086  Validation Accuracy: 0.65000  Test Accuracy: 0.68966 |

The effect of α on the training time isn’t very pronounced. It takes roughly the same time to train since the number of epochs were the same for all three scenarios. The change of α only impacts the rate of convergence towards an optimal solution, not the time it takes to compute the gradients for descent.

The effect of α on the final classification accuracy can be observed. It seems that decreasing α leads to a greater final classification accuracy on the training set. On the other hand, a larger α leads to faster progress but lower accuracy.

## **1.4 Generalization**

|  |  |  |  |
| --- | --- | --- | --- |
| **λ** | **Loss curves** | **Accuracy Curves** | **Final Classification Accuracy** |
| **0.001** |  |  | Training Accuracy: 0.68371    Validation Accuracy: 0.67000  Test Accuracy: 0.64138 |
| **0.1** |  |  | Training Accuracy: 0.66743  Validation Accuracy: 0.66000  Test Accuracy: 0.62069 |
| **0.5** |  |  | Training Accuracy: 0.64800  Validation Accuracy: 0.67000  Test Accuracy: 0.61379 |

As regularization increases, the performance of the model on the training and testing set declines. The purpose of regularization is to induce generalization of the model to the validation set. Tuning λ to higher values allows the model to generalize and have good performance on the validation set despite worse performance on the training and testing sets. Tuning λ allows the model to have good generalization abilities even if training performance is not optimal.

## **1.5 Comparing Batch GD with Normal Equation**

Analytical solution

|  |  |  |
| --- | --- | --- |
|  | **Batch GD** | **Normal Equation** |
| **Training MSE loss** | 0.01366 | 0.011582 |
| **Accuracy** | 0.68371 | 0.733143 |
| **Computation Time** | 00:00:36 | 00:00:01 |

The normal equation is superior to Batch GD in all three metrics.

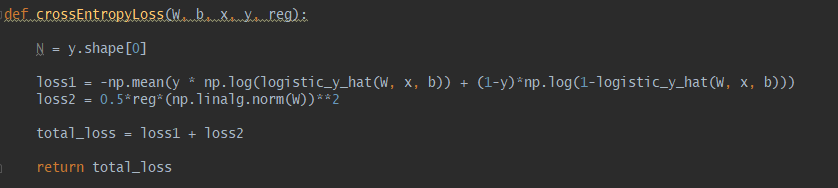
# **Part 2: Logistic Regression**

## **2.1 Loss Function and Gradient**

***sigmoid function***

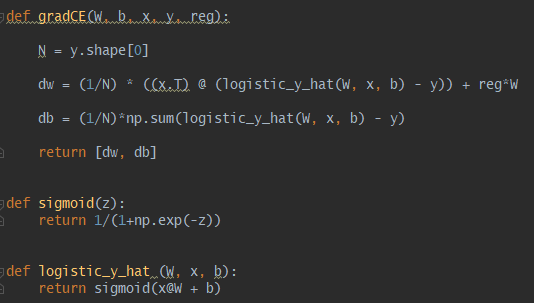
***logistic y\_hat***

***Cross Entropy loss***

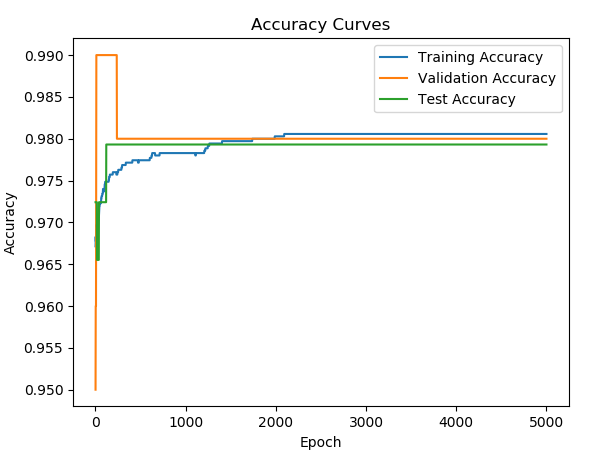
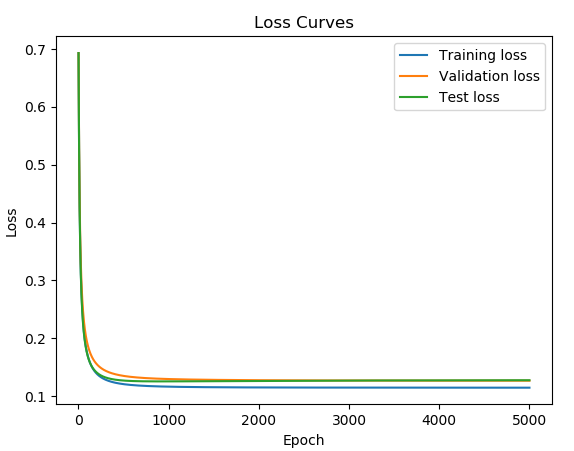


***Gradient with respect to weights***

***Gradient with respect to bias scalar***



## **2.2 Learning**



## **2.3 Comparison to Linear Regression**

|  |  |
| --- | --- |
| **Cross Entropy** | **MSE** |
|  |  |

The cross-entropy loss converges faster than MSE loss. This is expected since we are doing binary classification and the dataset is more suited towards logistic regression rather than linear regression.

# **Batch Gradient Descent VS SGD & Adam**

## **3.1.2 Implementing SGD Algorithm**

|  |  |  |
| --- | --- | --- |
|  | **Loss Curves** | **Accuracy Curves** |
| **MSE Loss** |  |  |
| **Binary Cross Entropy Loss** |  |  |

## **3.1.3 Batch Size Investigation**

|  |  |  |
| --- | --- | --- |
| **Minimizing MSE Loss** | | |
| **Batch Size** | **Loss Curves** | **Accuracy Curves** |
| **100** |  |  |
| **700** |  |  |
| **1750** |  |  |

|  |  |  |
| --- | --- | --- |
| **Minimizing Binary Cross Entropy Loss** | | |
| **Batch Size** | **Loss Curves** | **Accuracy Curves** |
| **100** |  |  |
| **700** |  |  |
| **1750** |  |  |

## **3.1.4 Hyperparameter Investigation**

|  |  |  |
| --- | --- | --- |
| **Minimizing MSE Loss** | | |
| **Hyperparameter** | **Accuracy Curves** | **Final Accuracies** |
| **β1 = 0.95** |  | Training Accuracy: 0.71857  Validation Accuracy: 0.75000  Test Accuracy: 0.69655 |
| **β1 = 0.99** |  | Training Accuracy: 0.71286  Validation Accuracy: 0.75000  Test Accuracy: 0.68966 |
| **β2 = 0.99** |  | Training Accuracy: 0.67257  Validation Accuracy: 0.72000  Test Accuracy: 0.64828 |
| **β2 = 0.9999** |  | Training Accuracy: 0.71829  Validation Accuracy: 0.69000  Test Accuracy: 0.68276 |
| **ϵ = 1e-9** |  | Training Accuracy: 0.73514  Validation Accuracy: 0.77000  Test Accuracy: 0.71724 |
| **ϵ = 1e-4** |  | Training Accuracy: 0.73514  Validation Accuracy: 0.77000  Test Accuracy: 0.71724 |

|  |  |  |
| --- | --- | --- |
| **Minimizing Binary Cross Entropy Loss** | | |
| **Hyperparameter** | **Accuracy Curves** | **Final Accuracies** |
| **β1 = 0.95** |  | Training Accuracy: 0.98857  Validation Accuracy:  0.99000  Test Accuracy: 0.97931 |
| **β1 = 0.99** |  | Training Accuracy: 0.98571  Validation Accuracy:  0.98000  Test Accuracy: 0.97931 |
| **β2 = 0.99** |  | Training Accuracy: 0.98971  Validation Accuracy:  0.97000  Test Accuracy: 0.98621 |
| **β2 = 0.9999** |  | Training Accuracy: 0.98771  Validation Accuracy:  0.98000  Test Accuracy: 0.97931 |
| **ϵ = 1e-9** |  | Training Accuracy: 0.98857  Validation Accuracy:  0.98000  Test Accuracy: 0.97931 |
| **ϵ = 1e-4** |  | Training Accuracy: 0.98857  Validation Accuracy:  0.98000  Test Accuracy: 0.97931 |

## **3.1.5 Cross Entropy Loss Investigation**

## **3.1.6 Comparison against Batch GD**