Boonrit Boonmarueng ID.61070701611

Assignment I

Gradient-based learning with regularization constraints

1. The logistic regression model is binary classification that means it can handle only 2 classes. However, the output class in iris data set is a category with 3 classes including “Iris-setosa”, “Iris-versicolor” and “Iris-virginica”. To solve this problem, I convert the output column into 3 binary columns and use concept of “one-against-the-rest” strategy to create the model that can classify “1” or “0” for each binary column.

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

	Iris-setosa	Iris-versicolor	Iris-virginica
0	1	0	0
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0

2. Formulate the error function of the logistic regression with ridge regularization criterion.

$$J(\theta) = -\frac{1}{n}\sum_{i=1}^{n}[\, y_i \, log \, H_\theta + (1 - y_i)log(\, 1 - H_\theta)\,] + \lambda \sum_{j=1}^{m} w_j^2$$

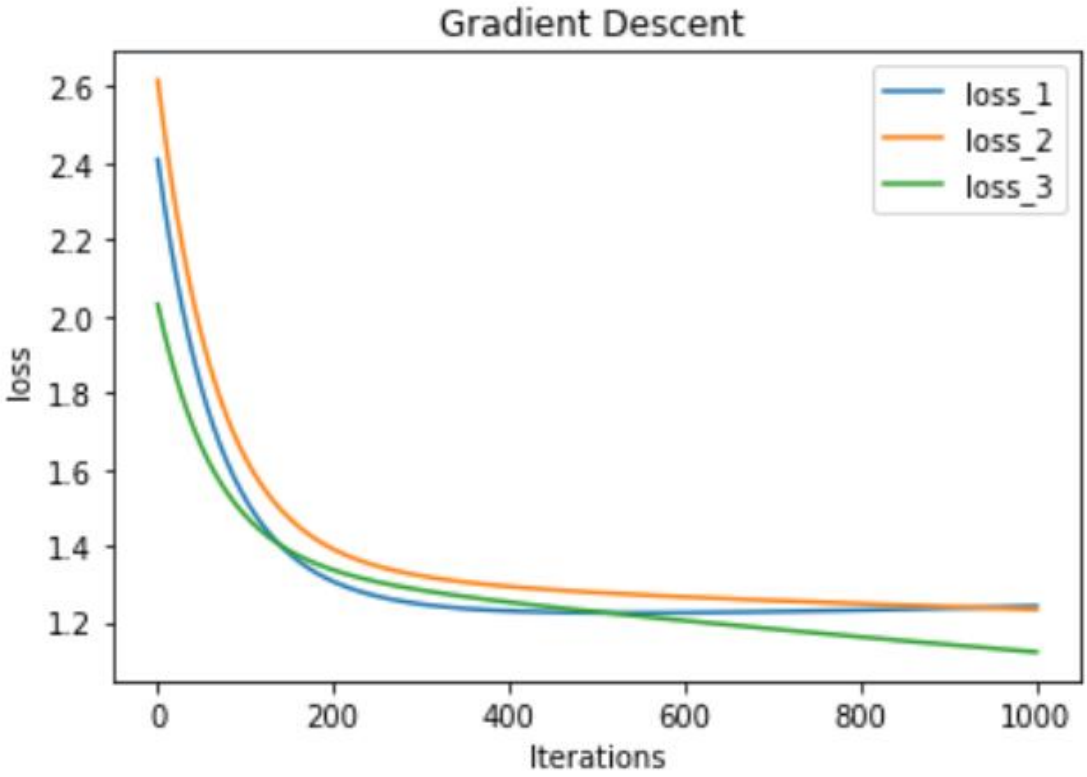
3. Derive the gradient of the error function by deriving the partial derivative of the error function in Task 2.

$$Gradient \; for \; W1 - W4 = \frac{1}{n}\sum_{i=1}^{n}[(\, H_\theta(x_i) - y_i)x_i] + 2\lambda \sum_{j=1}^{m} w_j$$

$$Gradient \; for \; W0 = \frac{1}{n}\sum_{i=1}^{n}[(\, H_\theta(x_i) - y_i)]$$

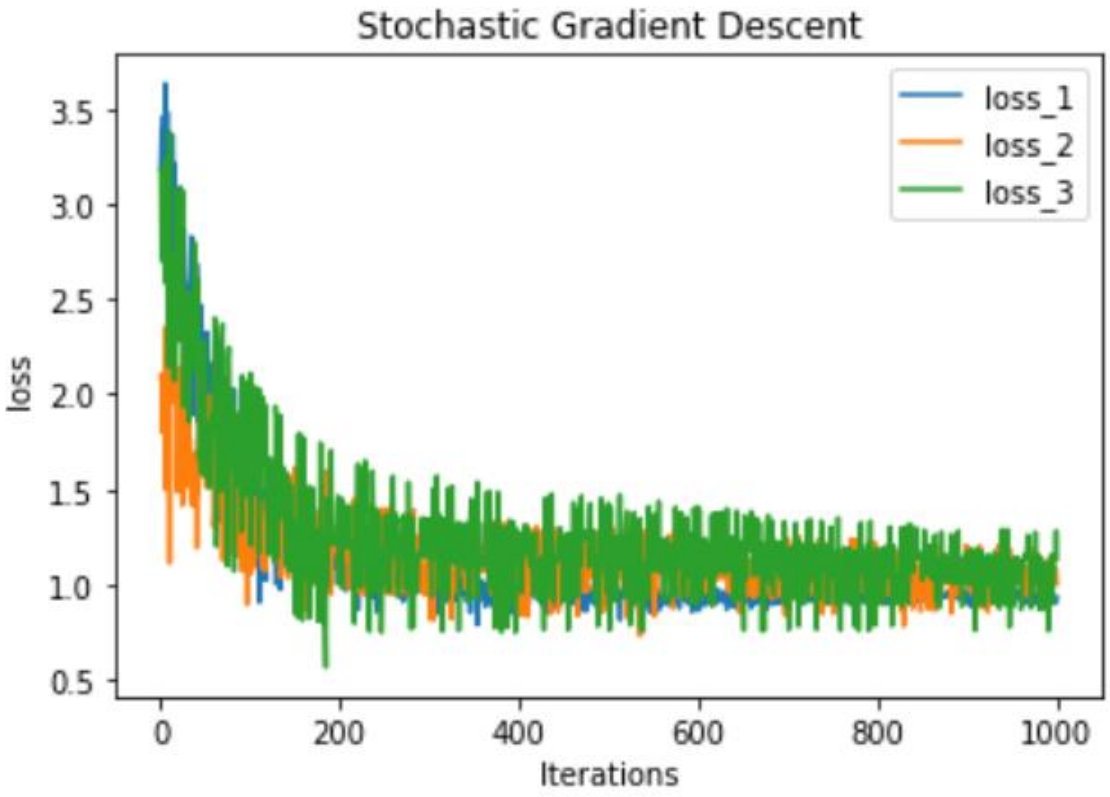
4. Implement the gradient descent using all the dataset in each iteration.

lamda = 1, learning rate = 0.001 and epoch = 1,000



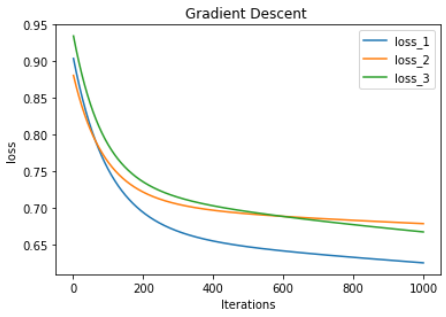
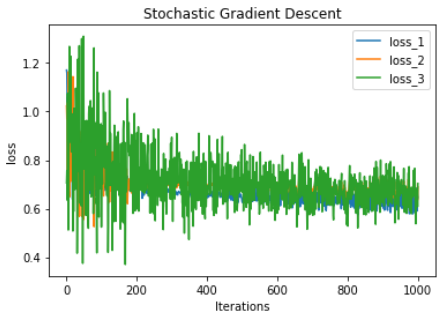
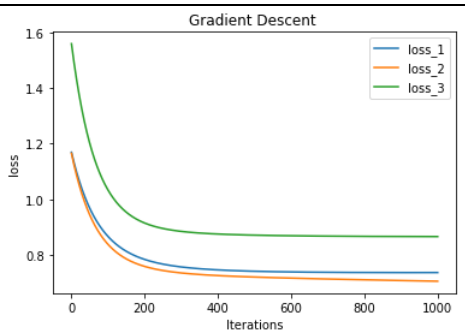
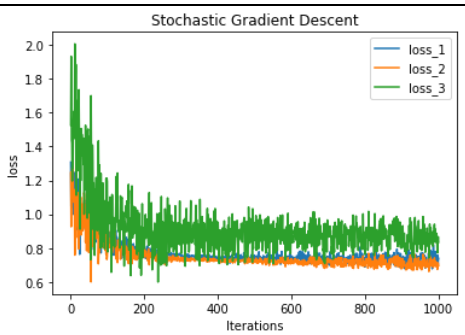
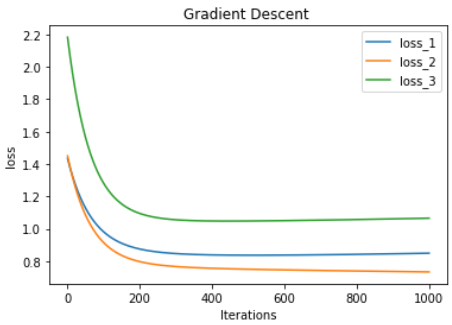
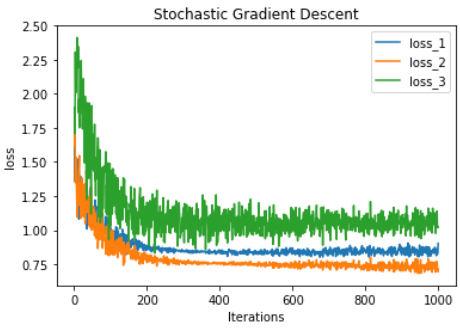
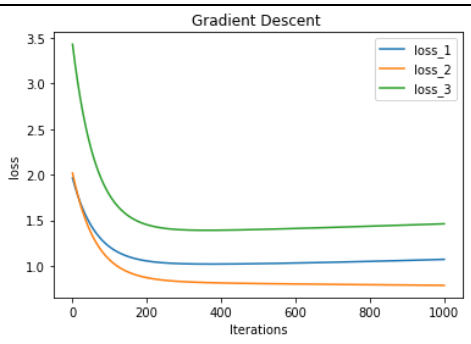
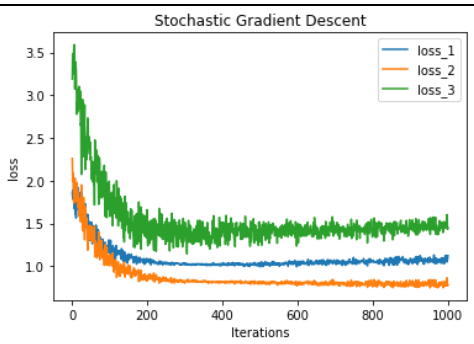
5. Implement the stochastic gradient descent using the subset of dataset in each iteration.

   lamda = 1, learning rate = 0.001, epoch = 1,000 and batch size = 5



6. Change λ value to see its effect.

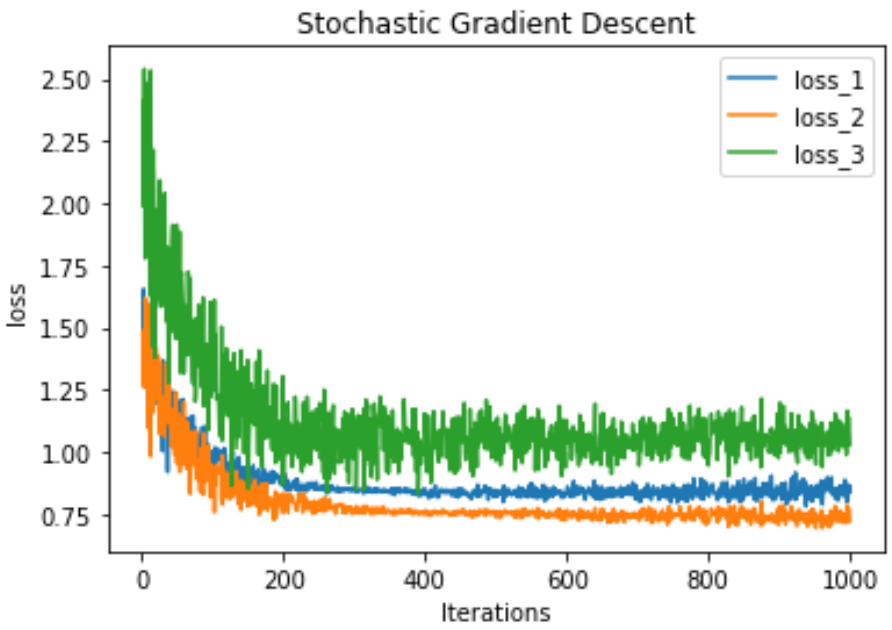
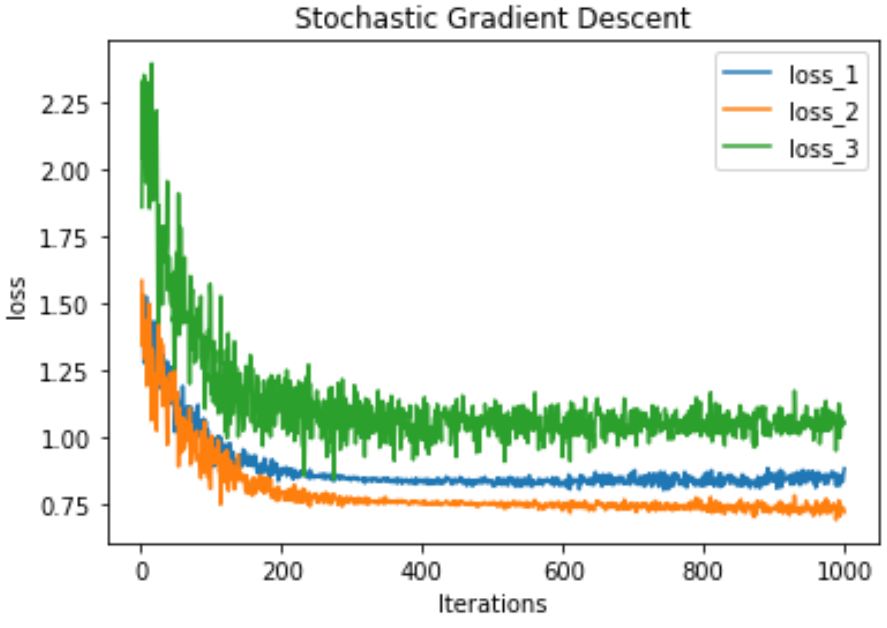
   Fixed value: learning rate = 0.001, epoch = 1,000 and batch size = 5

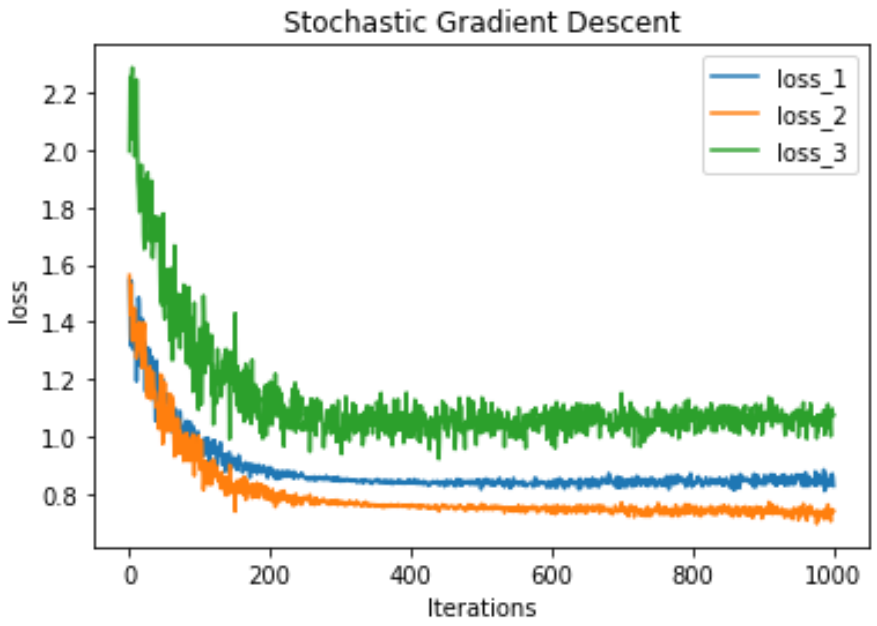
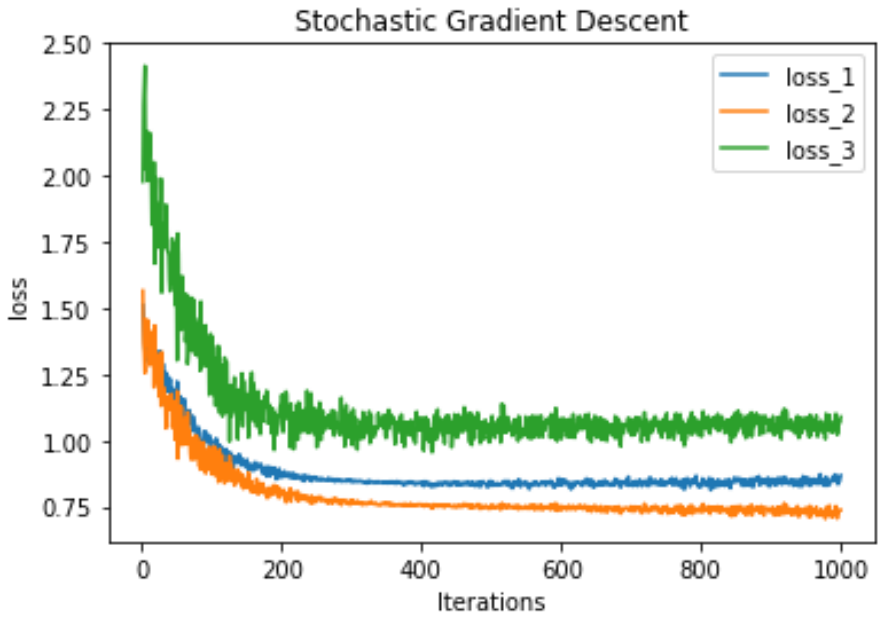
	Gradient Descent	Stochastic Gradient Descent
λ = 0.0		
λ = 0.5		
λ = 1		
λ = 2		

From the results above, increasing the value of λ will result in the loss values decrease and reache the optimum point faster.

7. Change the batch size to see its effect.

   Fixed value: learning rate = 0.001, λ = 1 and epoch = 1,000

	Stochastic Gradient Descent
Batch size = 5	
Batch size = 10	

Batch size = 15	
Batch size = 20	

As the results above, when increasing value of batch size from small to large will result in decreased variation of loss value.