CSCD 496-040 Prog 5

Blake Chalpin

22 February 2022

Sub-project 5

Dataset & Pre-Processing

For the following experiments, the `ionosphere` dataset is used. To normalize our dataset, the "negative 1 to positive 1" normalization method is used. This is all that constitutes the pre-processing phase.

Methodology

For the Stochastic Gradient Descent learning approach, we use an automated script to train our model for many different combinations of hyperparameters.

For Stochastic Gradient Descent, our tunable hyperparameters are

- The degree of Z-space
- λ (regularization)
- η (learning rate)
- The number of iterations
- Mini batch size

As mentioned above, we will use automated scripts to train our model for every unique combination of our hyperparameters. In the below section we will examine the results for our test conducted with hyperparameters:

- Z-transform Degree: $\{1, 2, 3\}$
- λ: $\{0.01, 0.001, 0\}$
- η: $\{0.1, 0.01, 0.001\}$
- Iterations: 10,000
- Mini batch size: $\{50, 100, 200\}$

Results

For the Stochastic Gradient Descent approach, we query our recorded output for each training run to find the optimal hyperparameters. This is the run with the least validation misclassifications.

For our **Stochastic Gradient Descent** training runs, we found that the run that yielded the least validation error is:

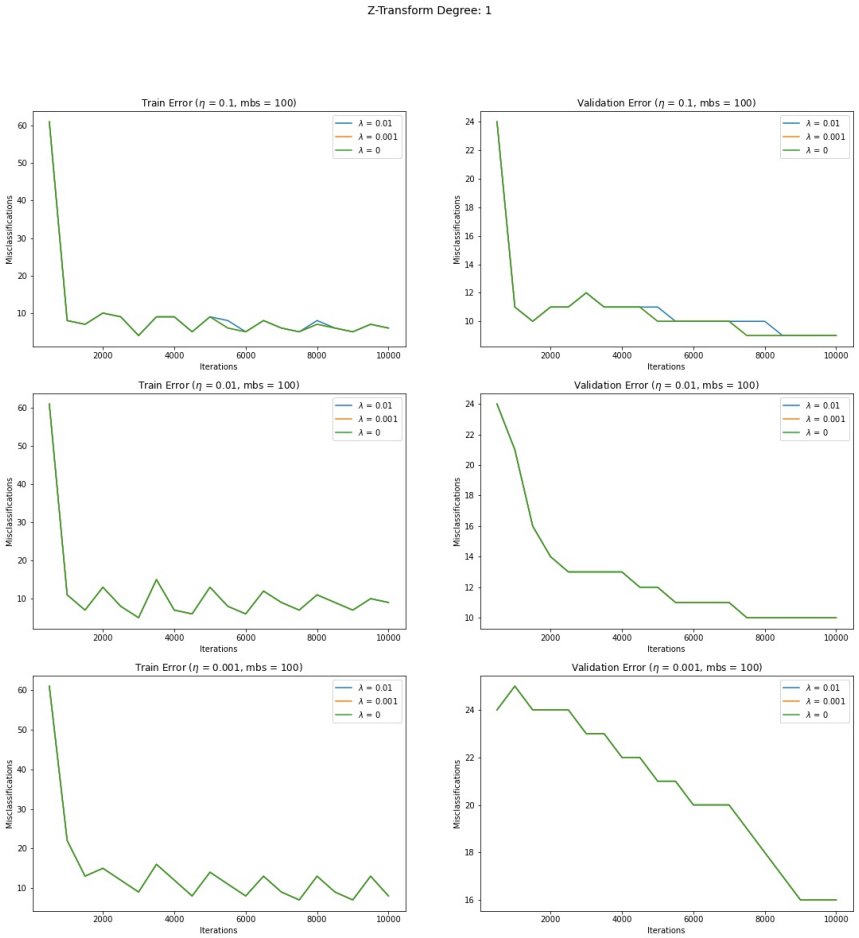
- Validation Misclassifications: 2/71
- Z-transform Degree: 2
- λ: 0.01
- η: 0.1
- Mini batch size: 50

Visualization

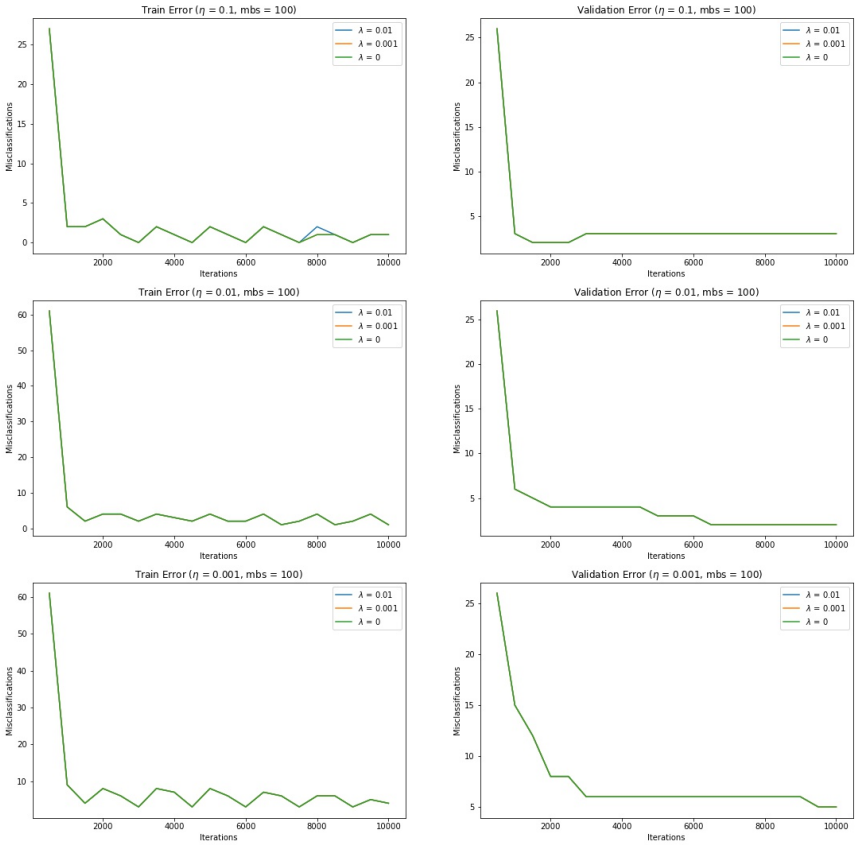
Gradient-Descent Method

Note: For each Z-transform degree, we will generate a new figure. Within this figure it will consist of two columns of sub-figures. The left column will examine training misclassifications, and the right validation misclassifications.

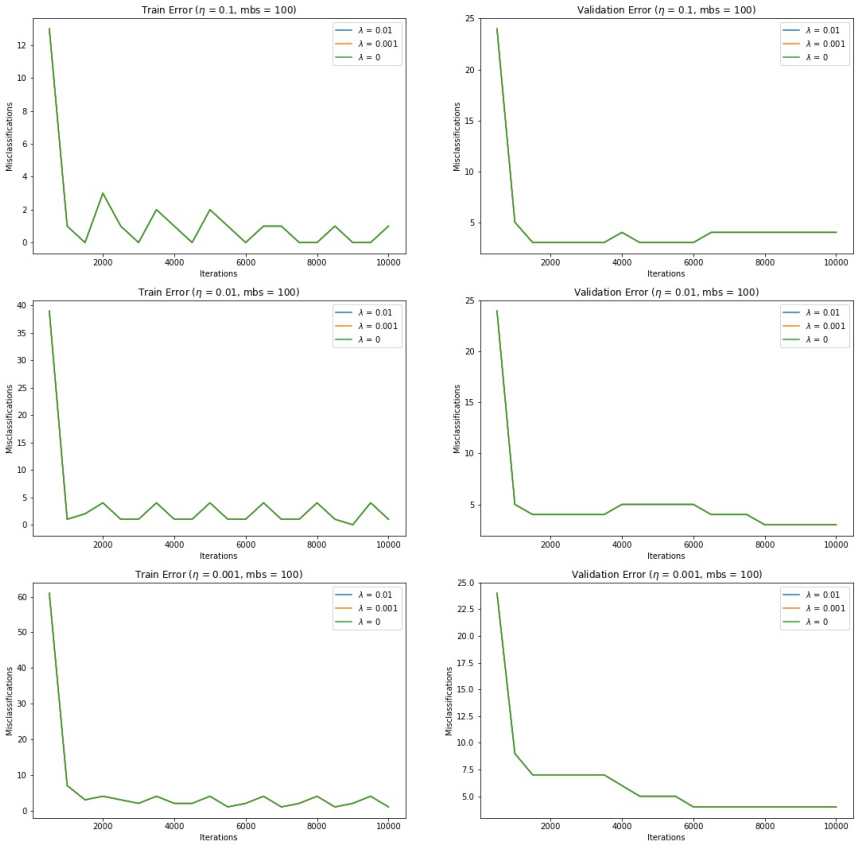
In the first set of figures, we examine the relationship between regularization (λ) and misclassifications. Note that for each pair of subplots, we fix the learning rate (η) and Mini batch size (mbs).


Z-Transform Degree: 1
Train Error (η = 0.1, mbs = 100)
Validation Error (η = 0.1, mbs = 100)
Train Error (η = 0.01, mbs = 100)
Validation Error (η = 0.01, mbs = 100)
Train Error (η = 0.001, mbs = 100)
Validation Error (η = 0.001, mbs = 100)
λ = 0.01
λ = 0.001
λ = 0
Misclassifications
Iterations

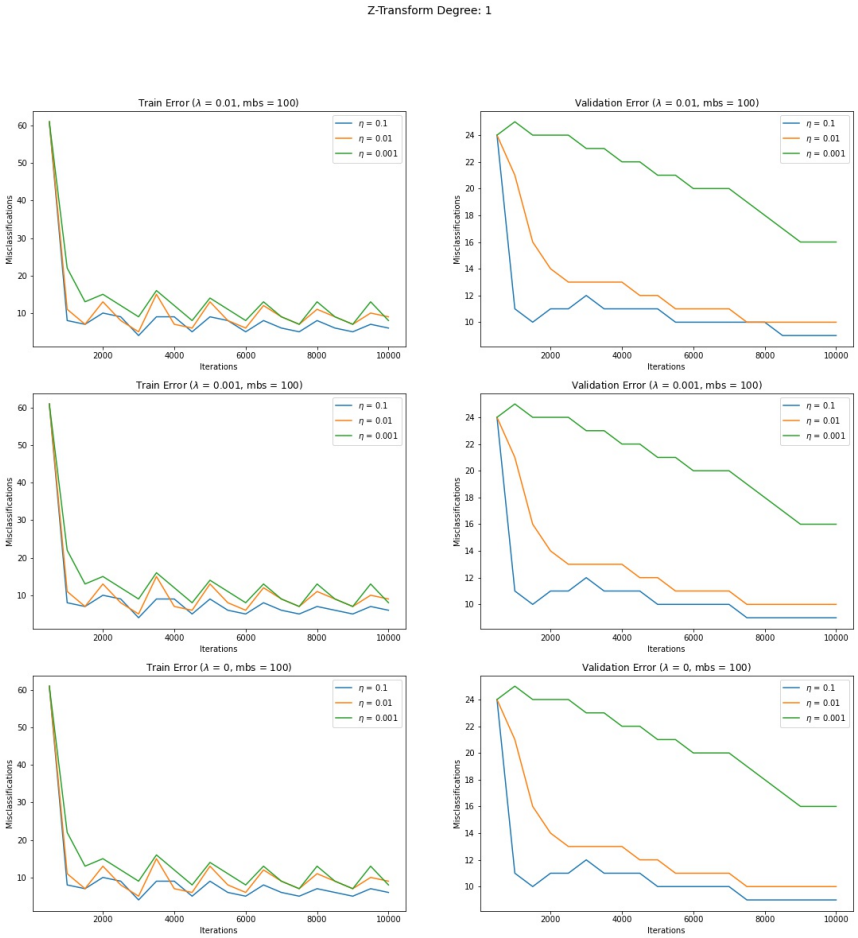
Z-Transform Degree: 2

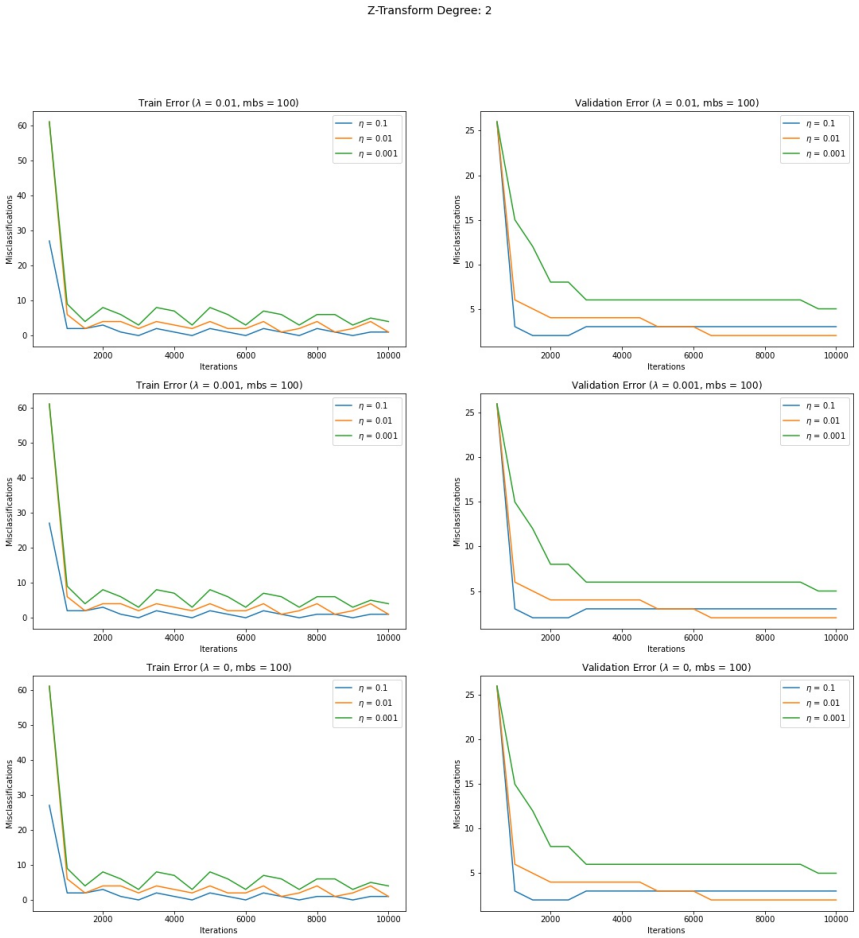


Z-Transform Degree: 3

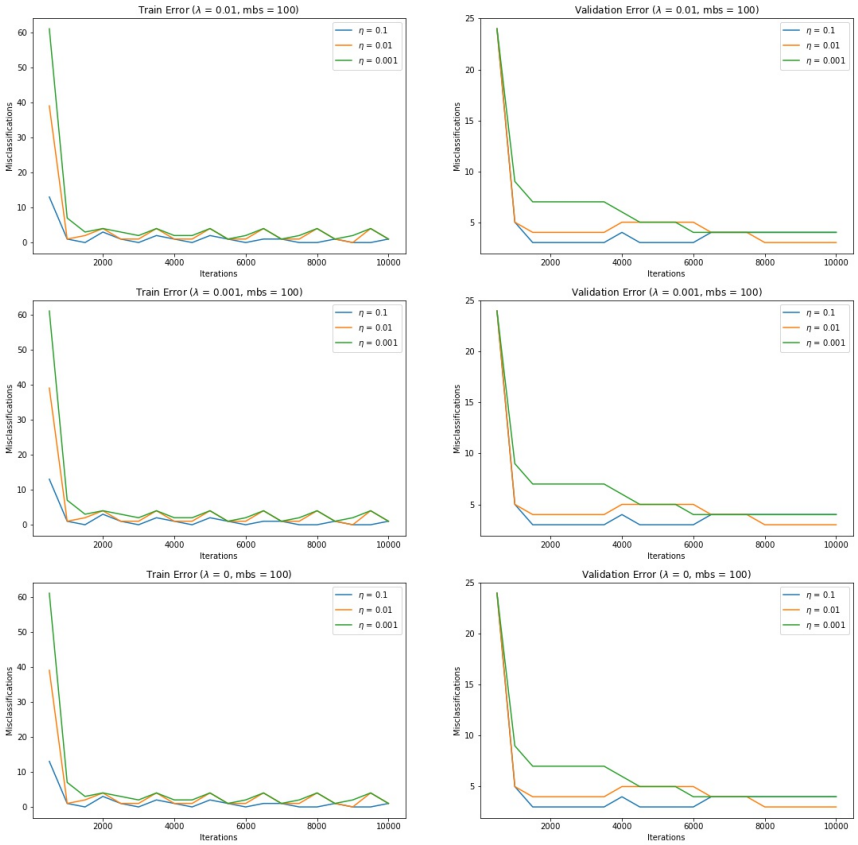


In the next set of figures, we examine the relationship between learning rate (η) and misclassifications. Note that for each pair of subplots, we fix the regularization (λ) and Mini batch size (mbs).

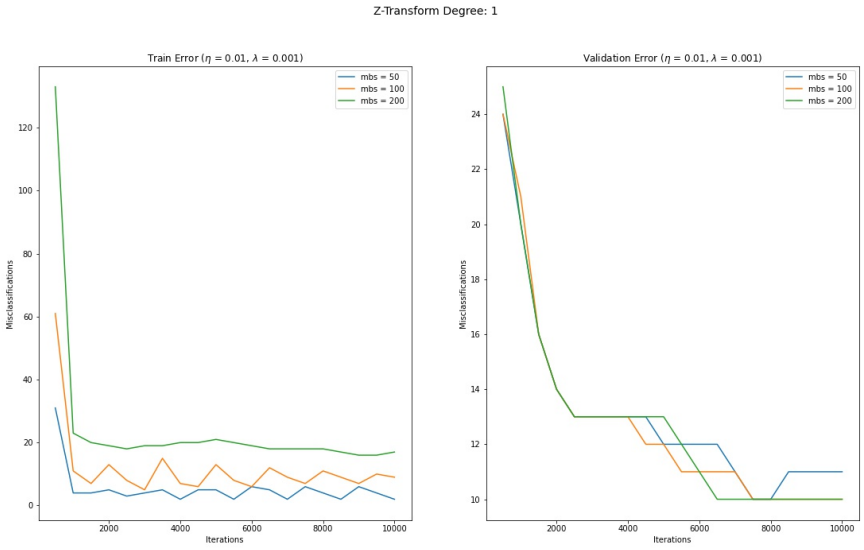


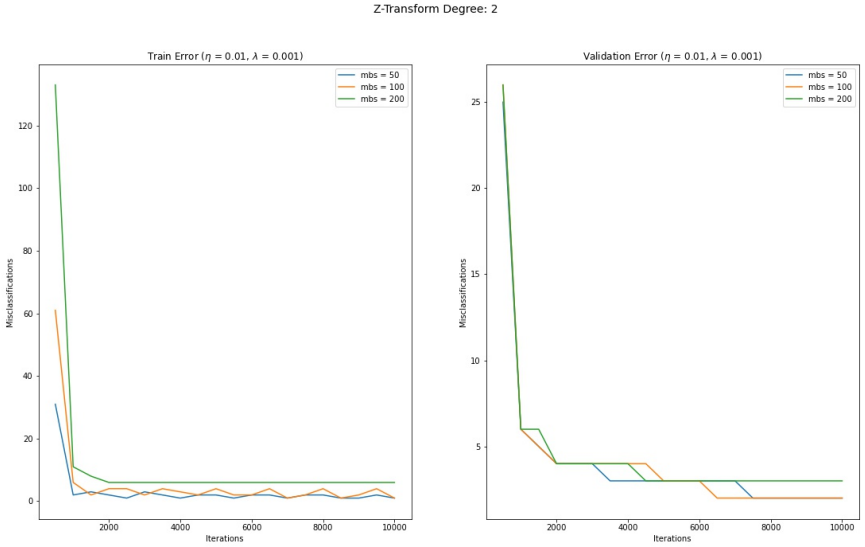

Z-Transform Degree: 2
Train Error (λ = 0.01, mbs = 100)
Validation Error (λ = 0.01, mbs = 100)
Train Error (λ = 0.001, mbs = 100)
Validation Error (λ = 0.001, mbs = 100)
Train Error (λ = 0, mbs = 100)
Validation Error (λ = 0, mbs = 100)
η = 0.1
η = 0.01
η = 0.001
Misclassifications
Iterations

Z-Transform Degree: 3



In the last set of figures, we examine the relationship between mini batch size (mbs) and misclassifications. Note that for each pair of subplots, we fix the regularization (λ) and learning rate (η).




Z-Transform Degree: 2
Train Error (η = 0.01, λ = 0.001)
Validation Error (η = 0.01, λ = 0.001)
mbs = 50
mbs = 100
mbs = 200
Misclassifications
Iterations

Z-Transform Degree: 3