Isai Martinez (im354)
Joshua Kruse (jek248)

Source Code: https://github.com/JoshEKruse/Gradient_Descent/

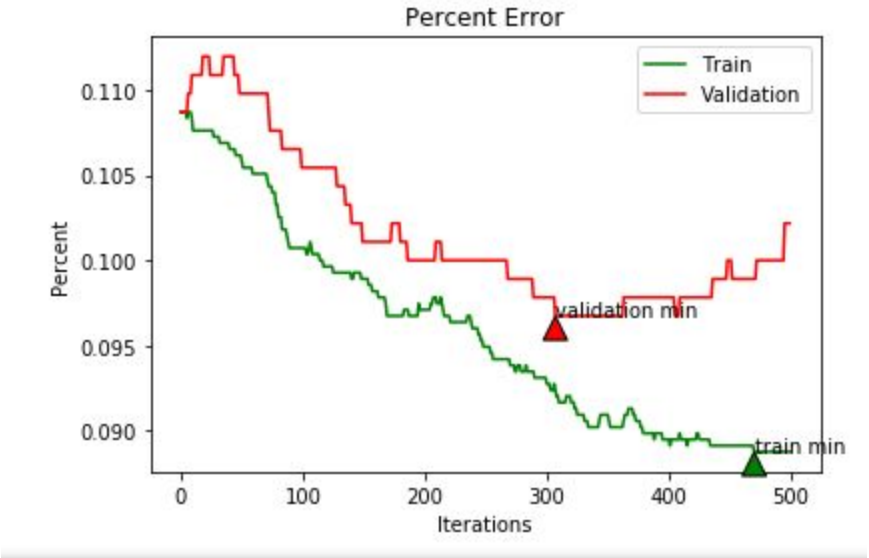
Group Coding Project 1

Spam.data

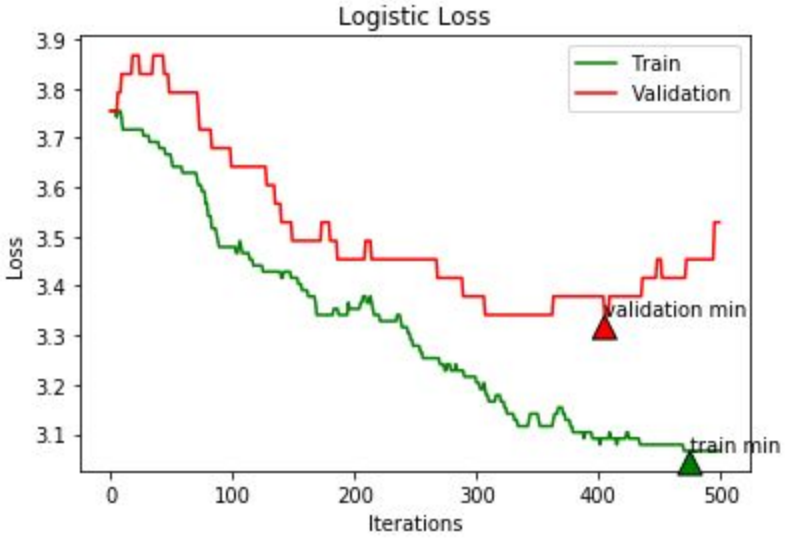
```
                y
        set     0    1
      train  1694 1066
       test   538  383
 validation   556  364
Number of iterations to minimize the validation error: 307
```

Spam data split into three sets train, test, and validation with each having 60%, 20%, and 20% respectively. The validation error is minimized at iteration 307. The most common class in this dataset is the 0 or “NOT SPAM” class, this class will be used to calculate the baseline.

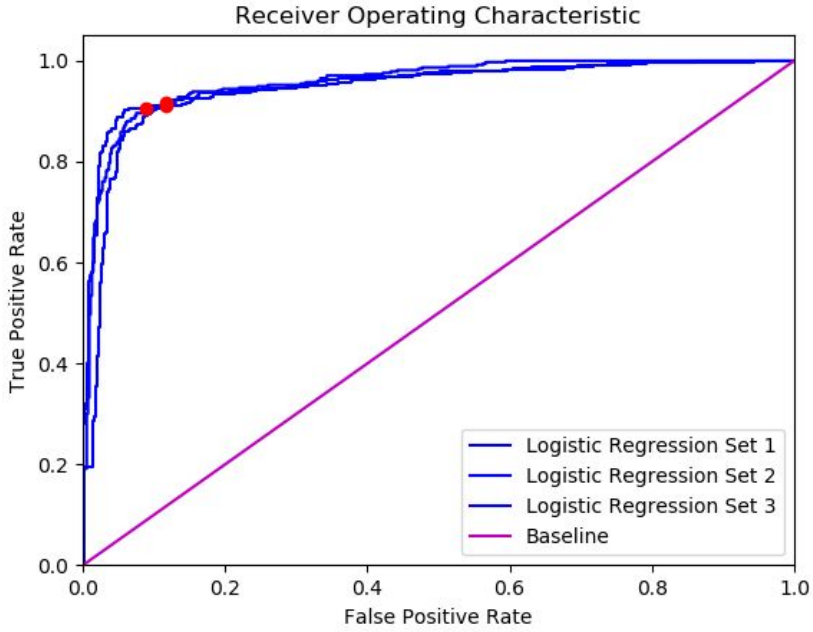
Step Size = 0.05
Max Iterations = 500



Graph displays the percent error for both the Train and Validation set, and the minimum value is shown with a green pointer for the Train and a red arrow for the Validation set.



Graph displays the logistic loss for both the Train and Validation set, and the minimum value is shown with a green pointer for the Train and a red arrow for the Validation set.



Three random samples were taken of the data to calculate three different train, test, and validation sets. The three different ROC curves these test sets produced are plotted above.

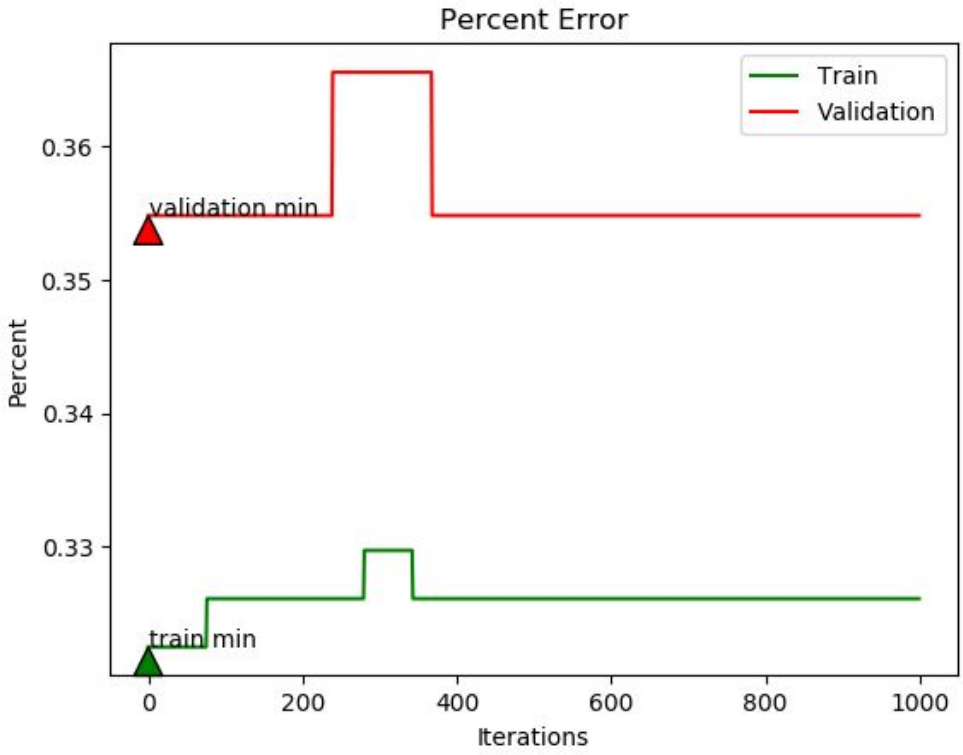
SAheart.data

	y	
set	0	1
train	187	89
test	61	32
validation	54	39

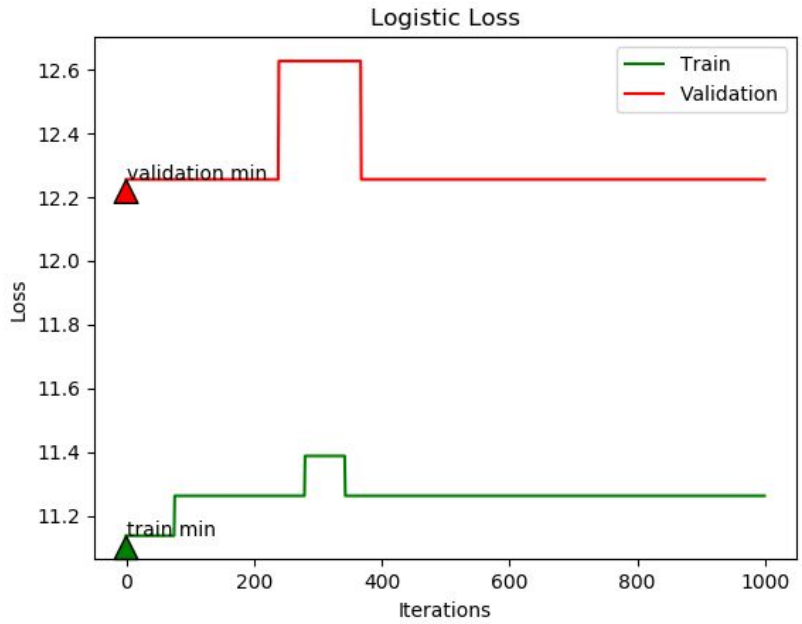
Number of iterations to minimize the validation error: 0

SAheart data split into three sets train, test, and validation with each having 60%, 20%, and 20% respectively. The validation error is minimized at iteration 0. The most common class in this dataset is the 0 or “NO DISEASE” class, this class will be used to calculate the baseline.

Step Size = 0.01
Max Iterations = 1000

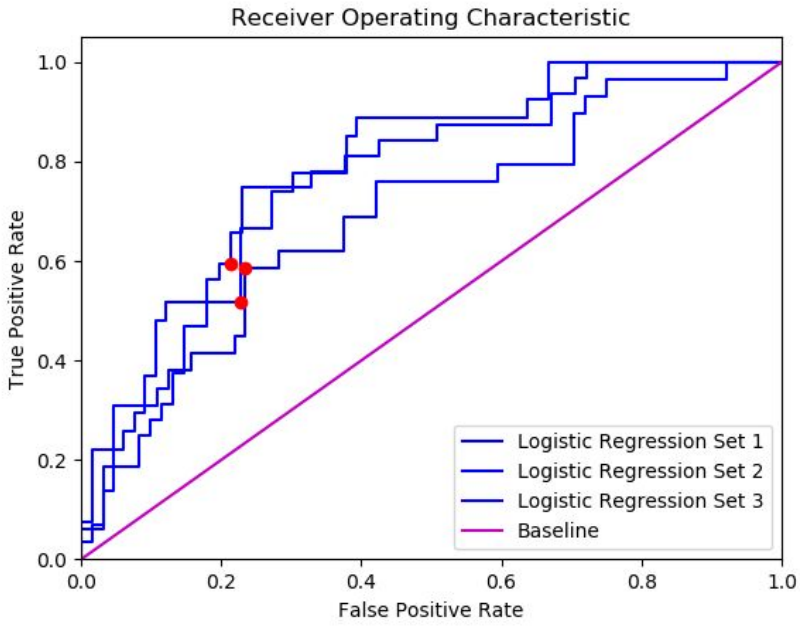


Graph displays the percent error for both the Train and Validation set, and the minimum value is shown with a green pointer for the Train and a red arrow for the Validation set.



Graph displays the logistic loss for both the Train and Validation set, and the minimum value is shown with a green pointer for the Train and a red arrow for the Validation set.

	log reg	baseline
train	32%	32%
validation	35%	41%
test	31%	34%



Three random samples were taken of the data to calculate three different train, test, and validation sets. The three different ROC curves these test sets produced are plotted above.

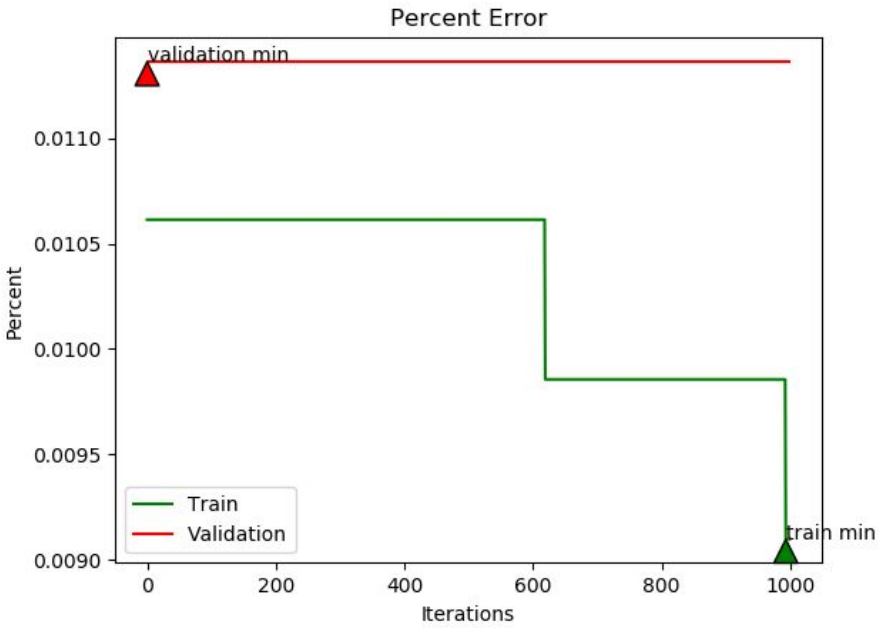
Zip.train

	y	
set	0	1
train	708	611
test	251	189
validation	235	205

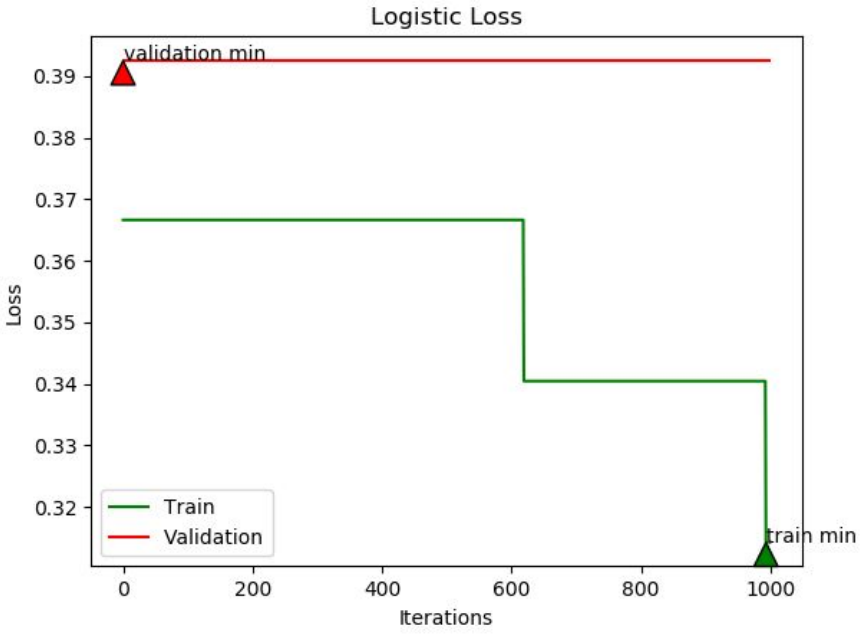
Number of iterations to minimize the validation error: 0

Zip.train data split into three sets train, test, and validation with each having 60%, 20%, and 20% respectively. The validation error is minimized at iteration 0. The most common class in this dataset is the 0 or “DIGIT ZERO” class, this class will be used to calculate the baseline.

Step Size = 0.1
Max Iterations = 1000

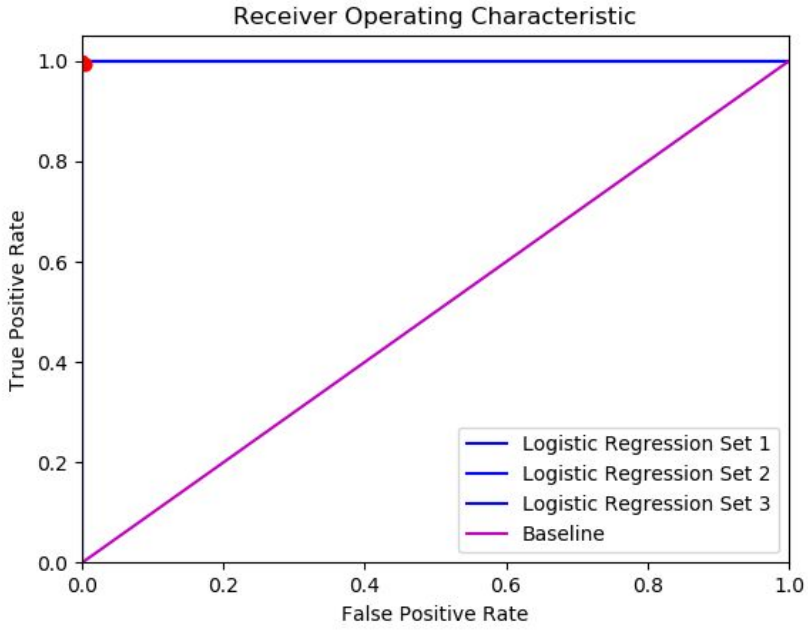


Graph displays the percent error for both the Train and Validation set, and the minimum value is shown with a green pointer for the Train and a red arrow for the Validation set.



Graph displays the logistic loss for both the Train and Validation set, and the minimum value is shown with a green pointer for the Train and a red arrow for the Validation set.

	log reg	baseline
train	1%	46%
validation	1%	46%
test	0%	42%



Three random samples were taken of the data to calculate three different train, test, and validation sets. The three different ROC curves these test sets produced are plotted above.