Machine Learning for Landsat Multi-Spectral Scanner image data

Landsat Multi-Spectral Scanner

- multi-spectral values of pixels in 3x3 neighborhoods
- classification for the central pixel in each neighborhood
- 7 classes describing different types of soil visible in the picture
- dataset does not contain records with class number 6

Algorithms

- multi-class logistic regression
- neural network
- support vector machines

No cross-validation

- in the dataset description there is an information that cross-validation should not be used with this dataset
- models were only trained and tested, without cross-validation

Multi-Class Logistic Regression

- one-vs-all approach
- performance evaluated using accuracy values, computed for model trained and tested with different values of regularization parameter lambda

Results

Accuracy training (lambda = 0.0) = 84.9830890643%

Accuracy test (lambda = 0.0) = 82.2%

Accuracy training (lambda = 0.5) = 84.2164599775%

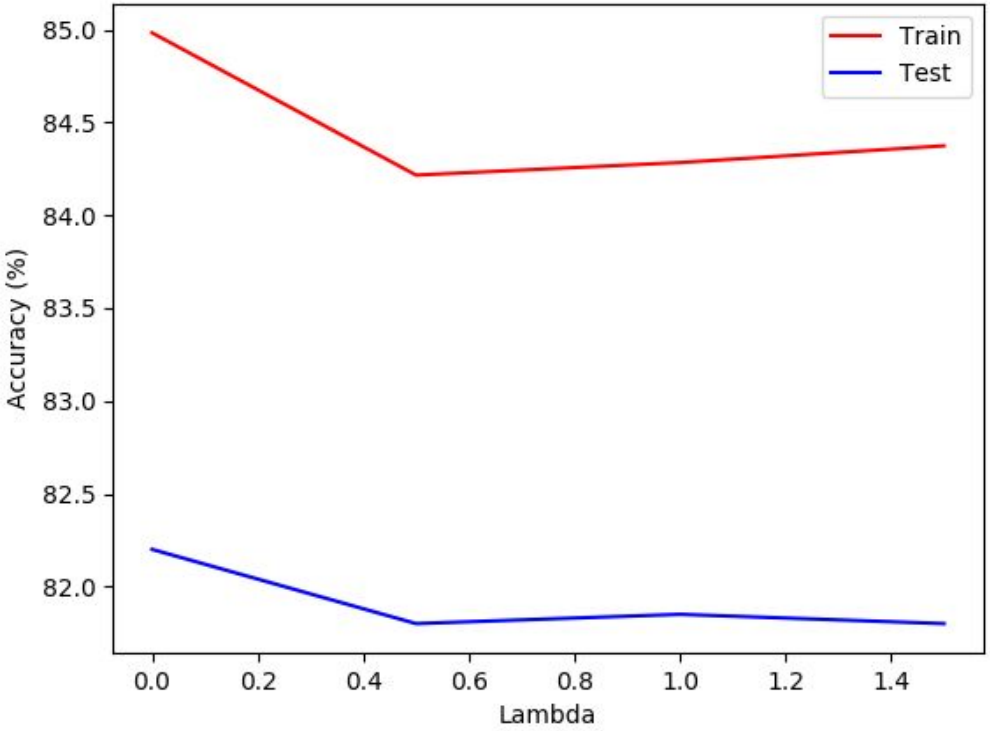
Accuracy test (lambda = 0.5) = 81.8%

Accuracy training (lambda = 1.0) = 84.2841037204%

Accuracy test (lambda = 1.0) = 81.85%

Accuracy training (lambda = 1.5) = 84.3742953777%

Accuracy test (lambda = 1.5) = 81.8%


Train
Test
85.0
84.5
84.0
83.5
83.0
82.5
82.0
Accuracy (%)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Lambda

Neural Network

- three-layer neural network
- random initialization of weights
- one-hot encoding for labels
- logistic regression cost function
- sigmoid activation function
- backpropagation implemented using gradient descent
- performance evaluated using accuracy values, computed for model trained and tested with different values of regularization parameter lambda

Results

Accuracy train for lambda = 0.0 = 79.00789177%

Accuracy test for lambda = 0.0 = 76.2%

Accuracy train for lambda = 0.5 = 89.3122886133%

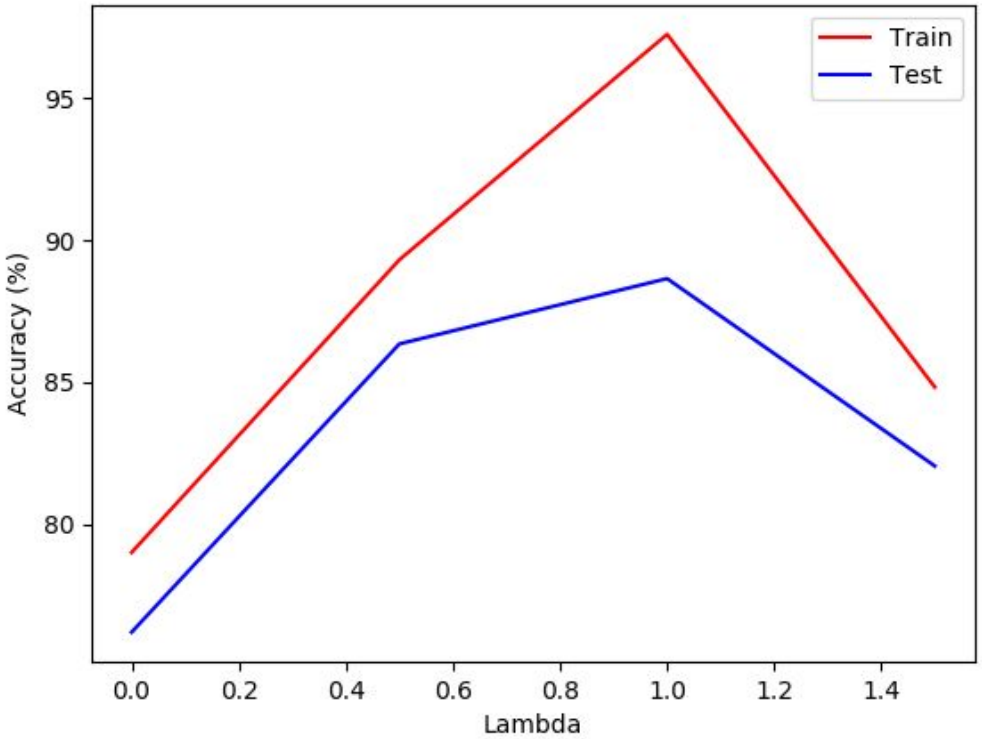
Accuracy test for lambda = 0.5 = 86.35%

Accuracy train for lambda = 1.0 = 97.2491544532%

Accuracy test for lambda = 1.0 = 88.65%

Accuracy train for lambda = 1.5 = 84.825253664%

Accuracy test for lambda = 1.5 = 82.05%


Train
Test
95
90
85
80
Accuracy (%)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Lambda

Support Vector Machines

- support vector machines implemented in scikit-learn module
- performance evaluated using accuracy values, computed for model trained and tested both with linear and RBF kernel and different values of C and sigma

Results for linear kernel

Train accuracy for SVM with linear kernel for C = 0.01 = 89.9887260428%

Test accuracy for SVM with linear kernel for C = 0.01 = 85.75%

Train accuracy for SVM with linear kernel for C = 0.1 = 90.1691093574%

Test accuracy for SVM with linear kernel for C = 0.1 = 85.2%

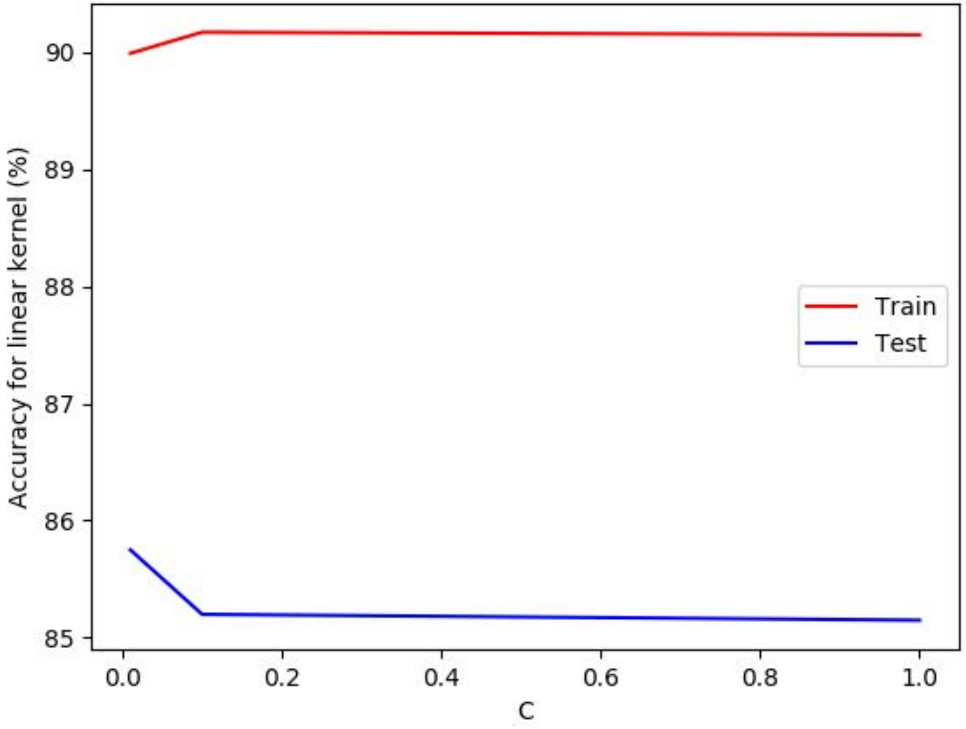
Train accuracy for SVM with linear kernel for C = 1 = 90.1465614431%

Test accuracy for SVM with linear kernel for C = 1 = 85.15%

Results for RBF kernel

Train accuracy for SVM with RBF kernel for C = 1 and sigma = 100 = 89.718151071%

Test accuracy for SVM with RBF kernel for C = 1 and sigma = 100 = 88.15%


Accuracy for linear kernel (%)
90
89
88
87
86
85
0.0
0.2
0.4
0.6
0.8
1.0
C
Train
Test

Summary

The obtained results show that for this problem the best choice would be a neural network with regularization parameter lambda set to 1, achieving 88.65% accuracy for test set