Logistic Regression

Huong Huynh & Maria Nazari

Math 285: Classification of Handwritten Digits

San José State
UNIVERSITY

Introduction

In machine learning, classification is the process of predicting the categories of a group of new observations using a training set to build a predictive model. Classification falls under supervised learning (labels provided in training set). A commonly used supervised classification method includes logistic regression. Normally in regression models use a continuous dependent variable, logistic regression is a model where the dependent variable is categorical. This project uses a combination of logistic regression and dimensionality reduction tools (Principal Component Analysis and 2DLDA) to classify the MNIST Hand Digit data set, compromising of a training set of 60,000 written hand digit samples and testing set of 10,000.

Method

Logistic Regregression

Binary Class Logistic Regression:

- Y variable is binary (only takes the values of 0 or 1)
- Need a function such that
  - *$f(E(Y_i)) = \theta_o + \theta_1 x_1 + \cdots + \theta_d x_d$*
- Need to use special function f( ) called the logistic function (or logistic transform):
  - $f(p) = \log\frac{p}{1-p}$ *, p used as the argument to LR because the function takes values between 0 and 1.*
- When Y is a binary variable. E(Y) = p, where p is the probability that Y takes the value 1.
- The Logistic Regression model is written as
  - $\log(\frac{p}{1-p}) = \theta_0 + \theta_1 x_1 + \cdots + \theta_d x_d$
- Notes:
  - *p: probability of "success" (i.e. Y = 1)*
  - *p/(1-p): odds of "winning"*
  - *log(p/(1-p)) is logit (a link function)*

Multiclass Logistic Regression:

There are two ways to extend it for multiclass classification:

- Union of binary models
  - *One-versus-one: construct a LR model for every pair of classes*
  - *One-versus-rest: construct a LR model for each class against the rest of training set*
- Softmax Regression (fixed versus rest) also known as multinomial logistic regression.
  - *The method fixes one class and fits c-1 binary logistic models for each of the remaining class against the fixed class, and The prediction for a new observation will be the class with the largest relative probability.*

$$\log\frac{P(Y=2\,|\,x)}{P(Y=1\,|\,x)} = \vec{\theta}_2\cdot x$$
$$\log\frac{P(Y=3\,|\,x)}{P(Y=1\,|\,x)} = \vec{\theta}\cdot x$$
$$\cdots$$
$$\log\frac{P(Y=c\,|\,x)}{P(Y=1\,|\,x)} = \vec{\theta}_c\cdot x$$

Feature Selection:

With high dimensional data, LR commonly overfits the data. Two methods can be used to resolve the problem:

- Reducing the dimensionality of the data using dimensionality reduction methods (such as PCA or 2DLDAA)
- Adding a regularization term to the objection function:

Result using Binary Logistic Regression

.Apply the binary logistic regression classifier to 45 pairs of digits:

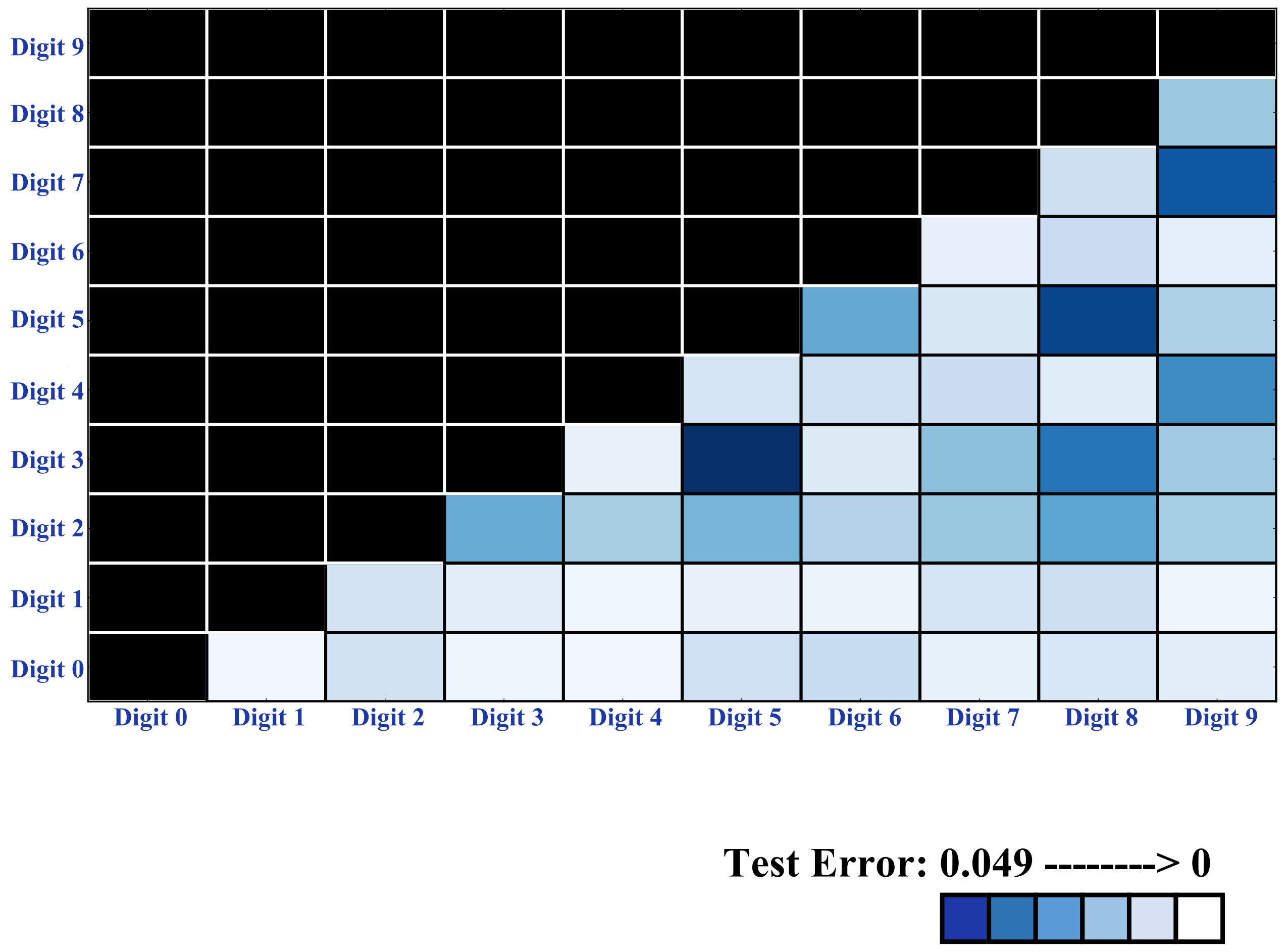

Figure 1: PCA 50 & Binary Logistic Regression Applied to All Digit Pairs

Figure 1 shows the binary regression model test errors for all possible pairs off the handwritten digits (Xi,Xj) where i and j are not equal, and X ranges from 0 to 9. The model for pair (3,5) resulted in the highest test error of 0.049. Pairs (5,8) and (7,9) resulted in high test errors. High test errors suggests that it is harder to predict the correct label between the pairs. Also, we can see that 0 versus other digits and 1 versus other digits produced low test errors. Pair (1,4) had lowest test error of 0.0019.

Apply binary logistic regression to all handwritten digits using one versus all method

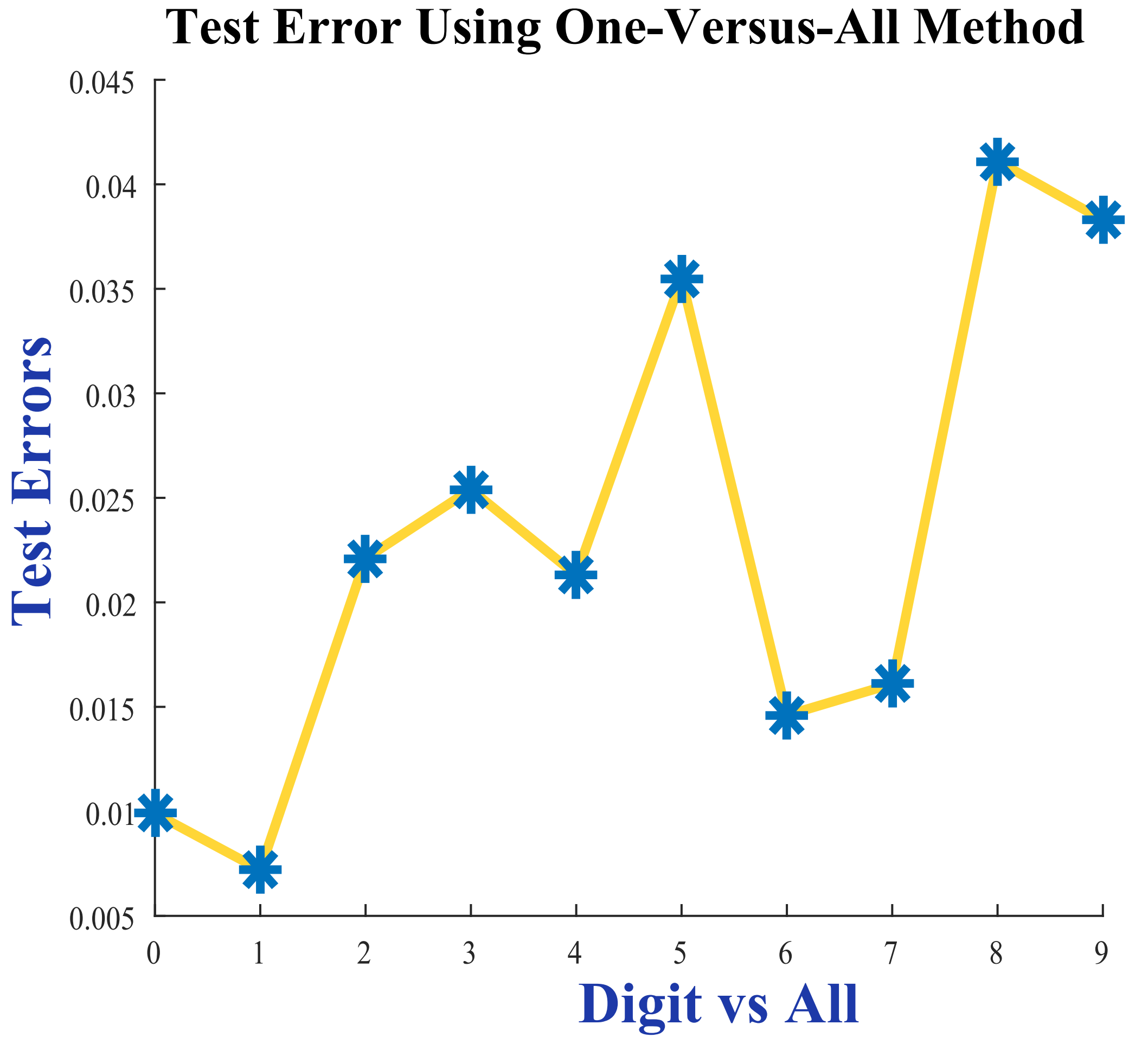

Figure 2: PCA 50 & One-Versus-All Method Applied to All Digits

Figure 2 shows the one versus all method for each digit. Digit 8 resulted in the largest test error of 0.0411 followed by digit 9 with an error of 0.0383. Digit 1 had the lowest test error when using one versus all binary logistic regression.

Result When Applying Different Logistic Regression Methods with PCA & 2DLDA

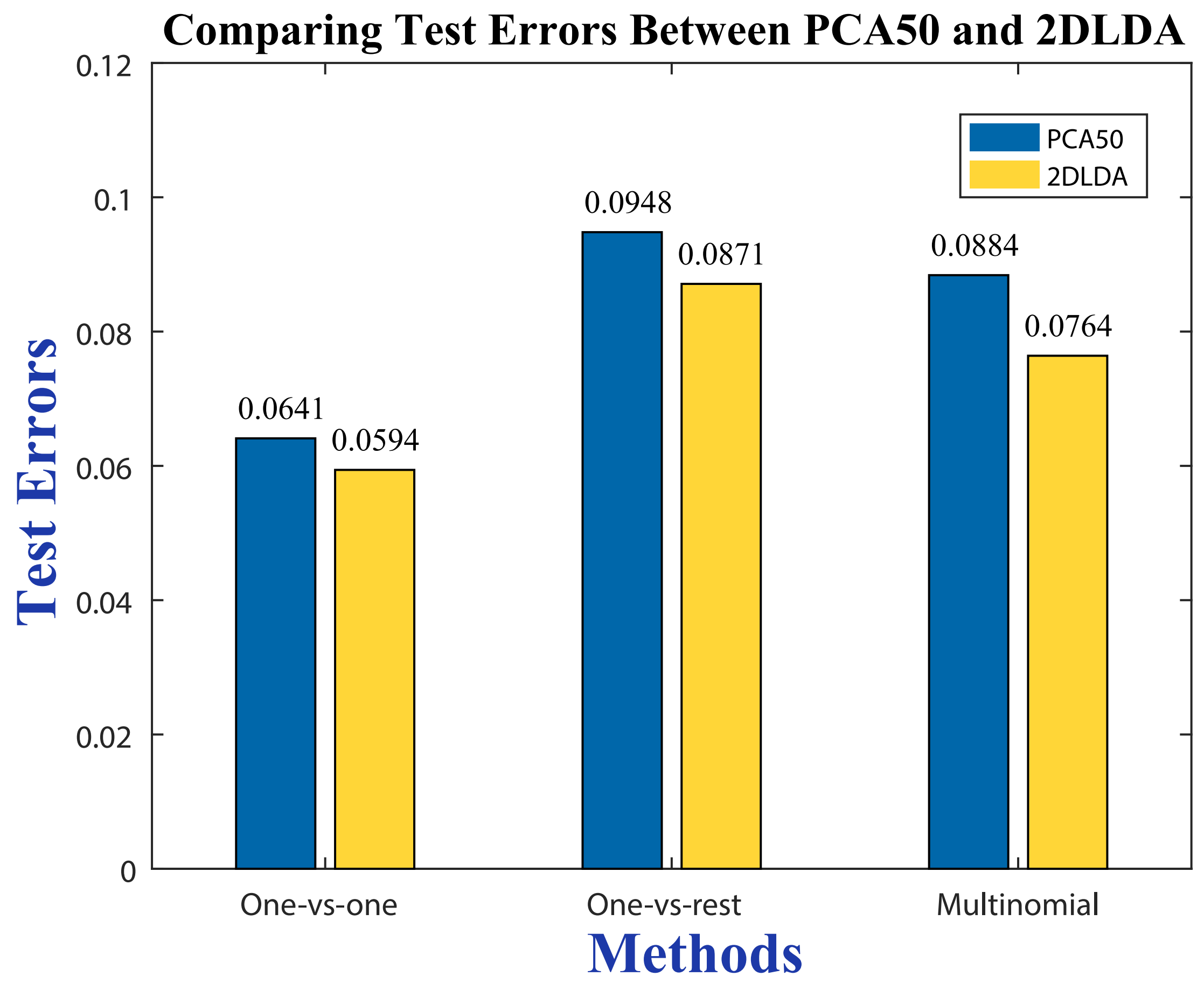

Figure 3: Binary Logistic Regression Methods Combined with PCA 50 and 2DLDA

Figure 3 shows the test errors when using logistic regression methods combined with PCA 50 and 2DLDA (reduce the data dimension to 11 by 11).
For PCA 50: The one-versus-rest method had the largest test error of 0.0948 and the one-versus-one method produced the smallest test error of 0.0641. The test error from the multinomial logistic regression method is 0.0884. For 2DLDA: The test error of one-vs-one binary logistic regression also give the smallest test error (0.0594) among the 4 methods. The method one-versus-rest binary method contributed the highest test error of 0.0871. Overall, 2DLDA produced lower test errors.

Result When Applying Logistic Regression with l1 & l2 Regularization

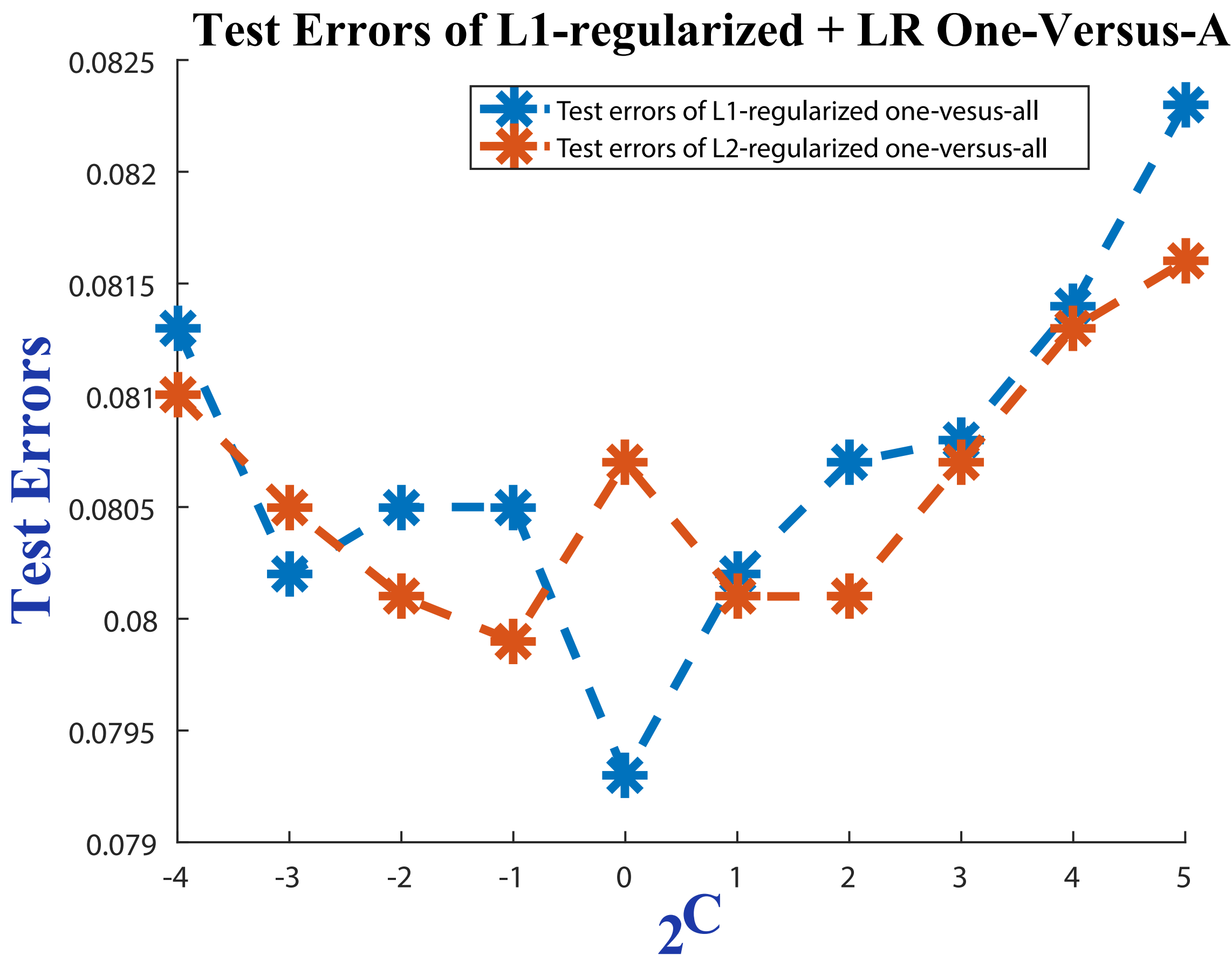

Figure 4: Binary Logistic Regression Combined with l1 & l2 Regularization

Figure 4: Uses the following objective function

$$\min_{\vec{\theta}=(\theta_0,\theta_1)} - \sum_{i=1}^{n} y_i \log p(x_i;\vec{\theta}) + (1 - y_i)\log(1 - p(x_i;\vec{\theta})) + C\,\|\vec{\theta}\|_p^p$$

For l1 regularization, p is set to 1 and l2 will have a p value of 2. Figure 4 contains different values of the regularization parameter (2^C).

l1 regularization has its lowest test error when C=1, and l2's lowest test error occurs at C=0.5.

As the value of C increases above 1, the test errors for both l1 and l2 continously increase.

Summary of Results

Comparing the Results of the Different Methods Applied to the MNIST Handwritten Dataset

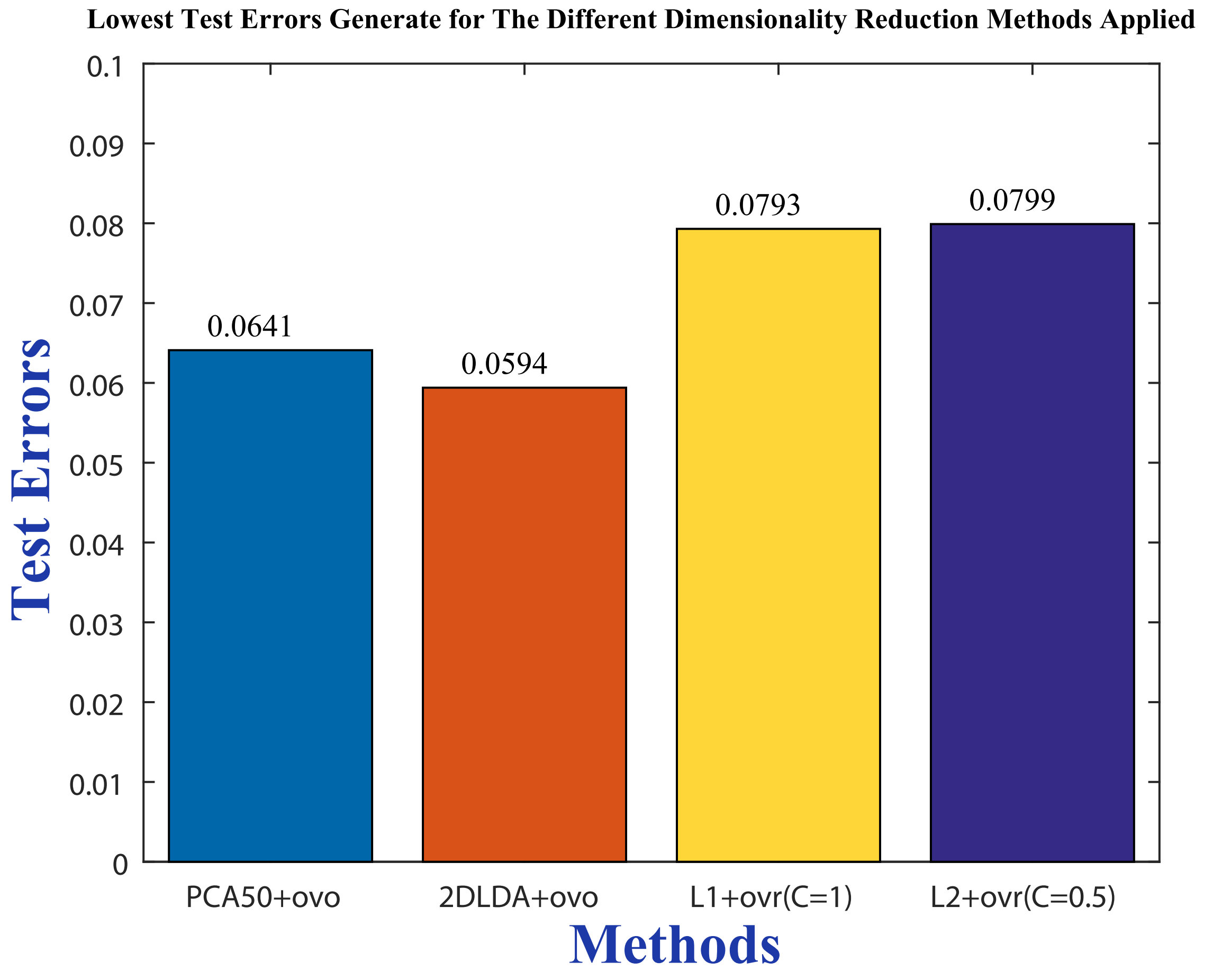

Figure 5: Comparing the Best Test Error for the Dimensionality Reduction Methods Applied

Figure 5 contains the best test errors achieved by the different dimensionality reduction methods analyzed in this project.

The best test errors for l1 and l2 regularizations using the one-versus-all method are relatively close in value (0.0793 vs 0.0799).

Using the one-versus-one method, PCA produced a test error of 0.0641 and 2DLDA resulted in a lower test error of 0.0594. Both PCA and 2DLDA outperformed the l1 and l2 regularization methods.

For the MNIST handwritten digit dataset, 2DLDA produced the lowest test error rates and had one of the lowest computational times to generate its test error.

Conclusion

This project implemented different variations of the logistic regression method of classification on the MNIST handwritten digit dataset. We found that the binary logistic regression method had trouble distinguishing between certain pairs of digits. This could be because of the similar ways the digits are written by the participants.

Different methods of logistic regression were applied on the dataset. To reduce the effects of overfitting, dimensionality tools such as PCA, 2DLDA, and 1 & l2 regularization were used to address the problem. 2DLDA generally outperformed PCA 50.

Overall, the logistic regression method performed very well in producing low test errors when classifying the test dataset. The one-versus-one model consistently outperformed the rest of the models examined in this project.

PCA and 2DLDA combined with the one-versus-one method outperformed l1 and l2 test errors. When using the one-versus-all method, l1 and l2 regularizations generated the small test errors. The best test error found is attributed to 2DLDA combine with the one-versus-one method.

Reference

[1] Guangliang Chen, "Logistic Regression", Lecture 6 – Math 285.

[2] Jeff Howbert, "Introduction to Machine Learning – Logistic Regression", Lecture – Winter 2012.