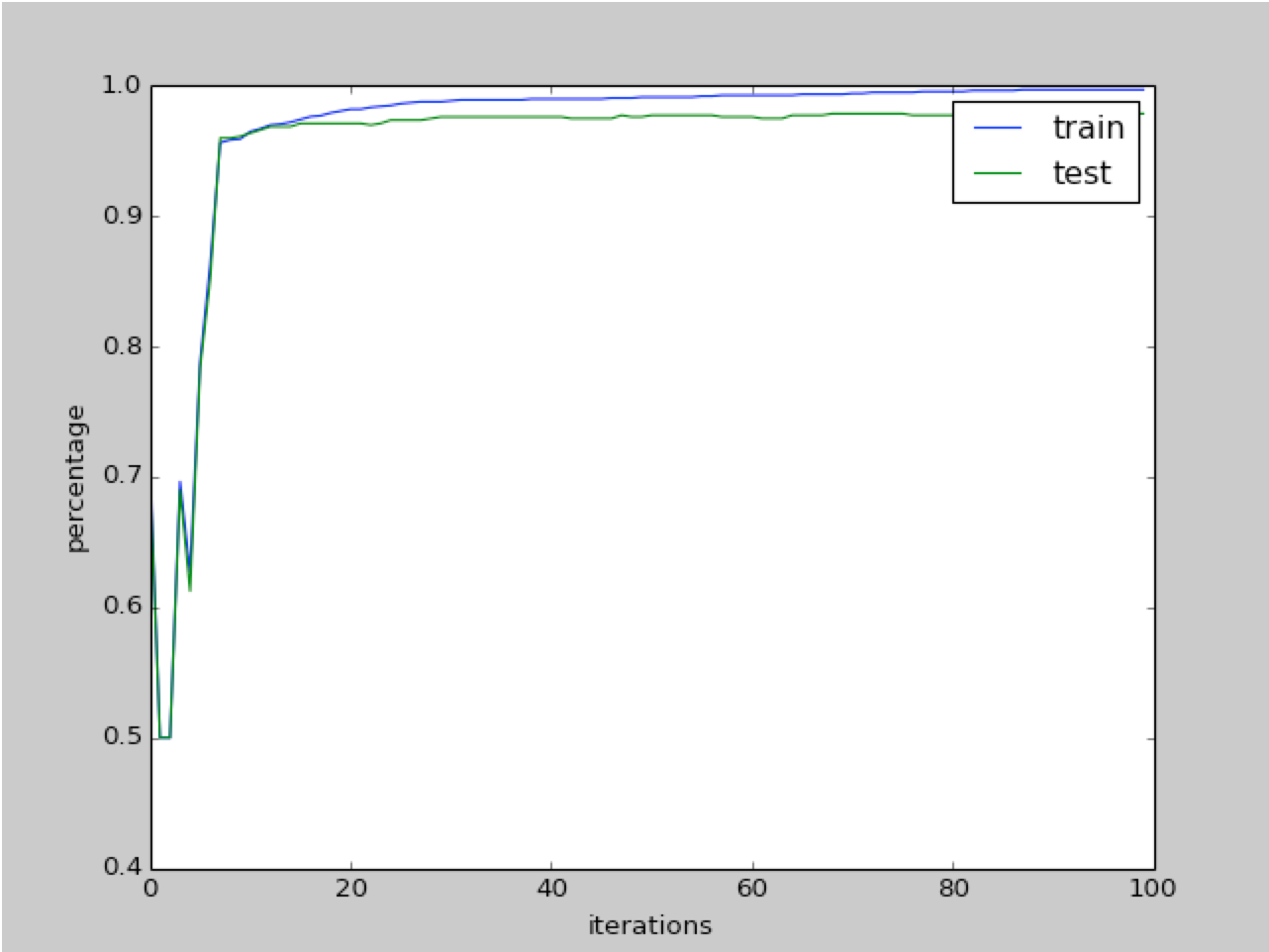
Alex Nguyen

Michael Lee

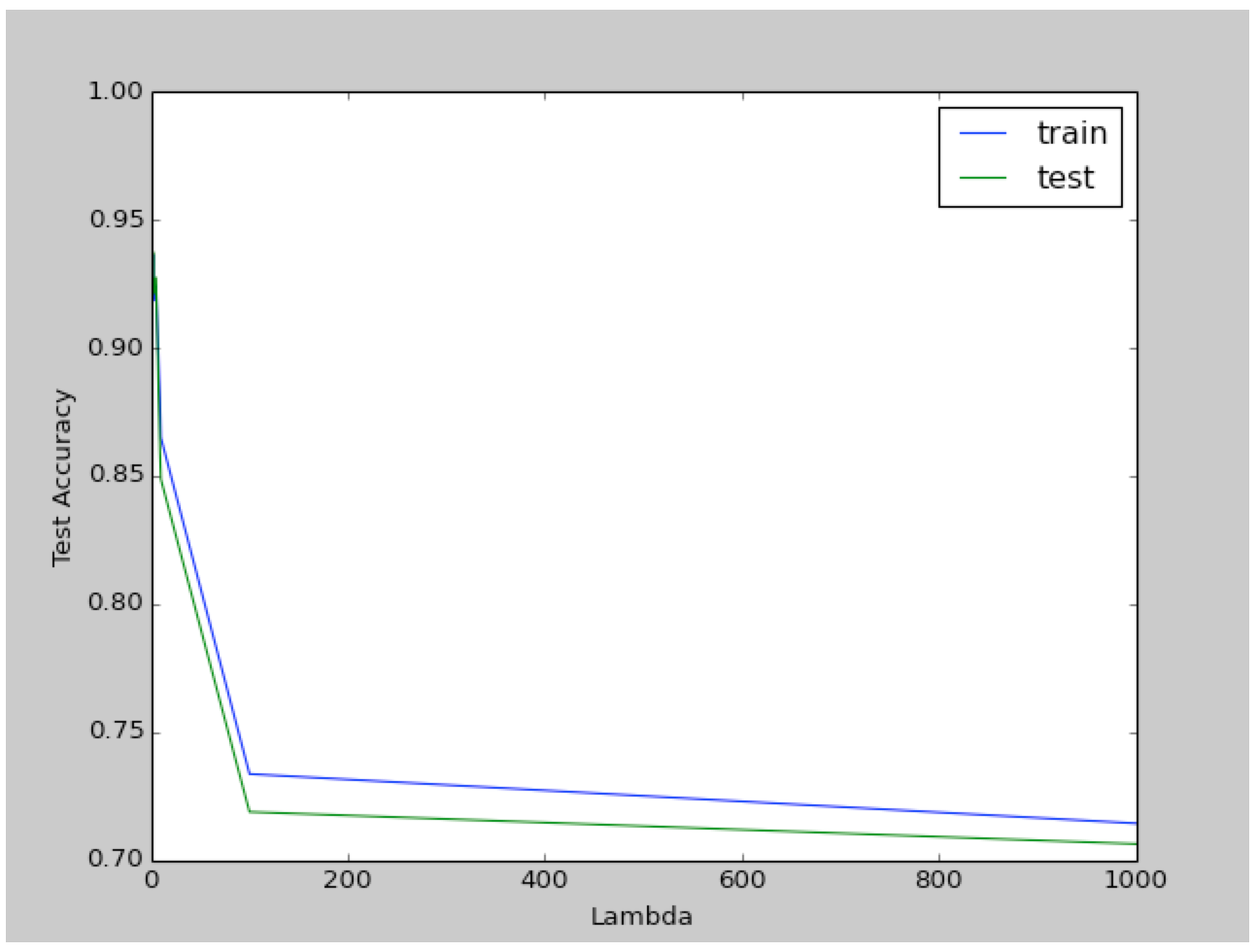
CS434

Assignment 2 - Logistic regression with L2 regularization

1. Batch gradient descent algorithm
   1. We found a small learning rate allowed us to not overflow when calculating W. Larger learning rates did not seem to converge in a reasonable amount of time or overflowed when we tried our processes.
   2. learningRate = 0.0000001
   3. Correct Percent = 0.97875



1. We noticed a where as each iteration of the gradient descent was performed our accuracy rose. This makes sense since our W is training itself to be more and more accurate to the training data. This may cause over-fitting but 100 iterations seemed to be a good fit for training and testing data.
2. Since the gradient descent algorithm uses the derivative of the loss function, we need to derive the regularization term. It becomes Lambda multiplied by W.
   1. Changes the Gradient Descent Algorithm:
      1. For each iteration:
         1. Calculation the regularization term before calculating d using lambda and the current value of W.
         2. Once d is calculated, W is now W + nd+ regularization term



1. We see that as Lambda gets larger and larger, the accuracy of our data becomes smaller and smaller. This is because the significance of lambda is so large it outshines our actual prediction. After a lambda of about 100, our “prediction” with regularization is always 1. The training data performs slightly better but has the same fallout rate like the test data.