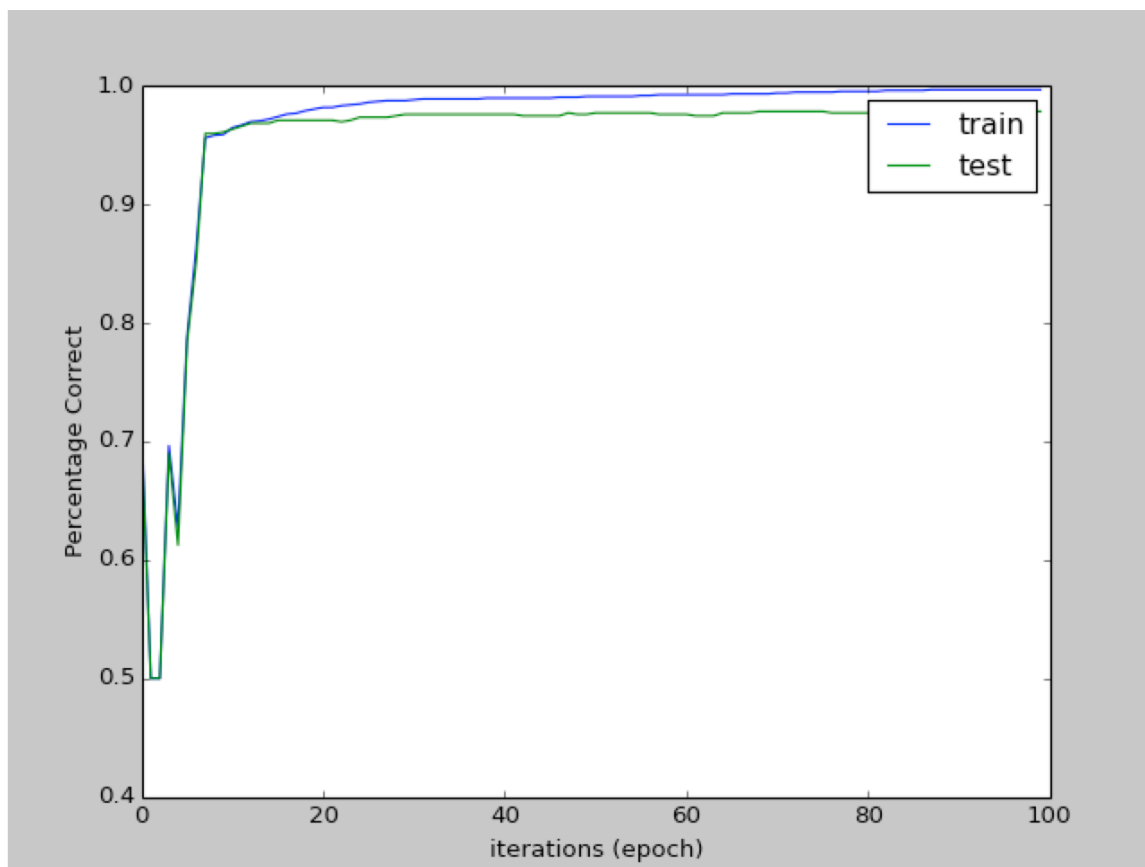


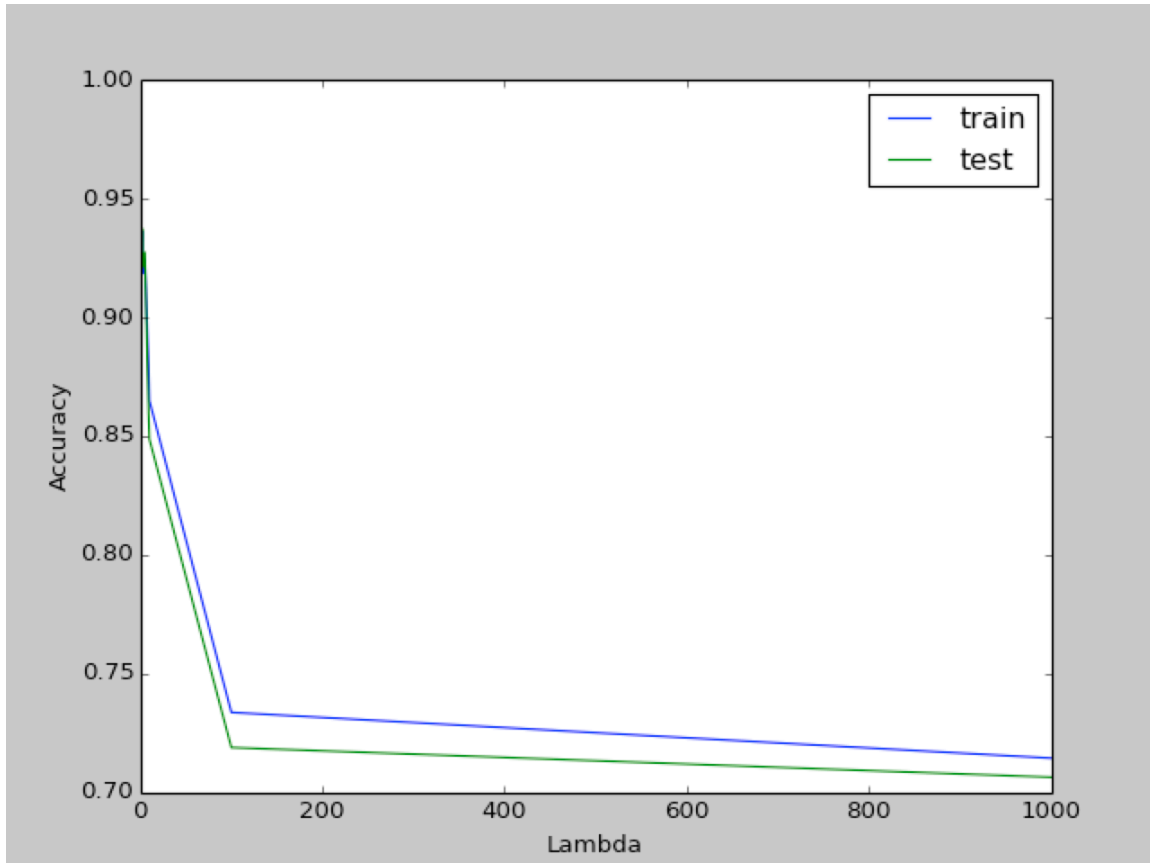
Assignment 2 - Logistic regression with L_2 regularization

1. Batch gradient descent algorithm
 - a. 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.
 - b. Learning Rate = 0.0000001
 - c. Correct Percent = 0.97875



2. We noticed 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 40 iterations seemed to be a good fit for training and testing data.
3. Since the gradient descent algorithm uses the derivative of the loss function, we need to derive the regularization term. The regularization term is $.5 * \lambda * ||w||_2^2$ and it derives to $\lambda * w$.

- a. Changes the Gradient Descent Algorithm:
 - i. For each iteration:
 - 1. Calculation the regularization term before calculating d using λ and the current value of W .
 - 2. Once d is calculated, W is now $W + \eta d + \text{regularization term}$



- 4. We see as λ gets larger and larger, the accuracy of our predictions become more inaccurate. This is because the significance of λ is so large it outshines our actual prediction. In terms of training versus test data, the training data performs slightly better but has the same fallout rate as the test data.