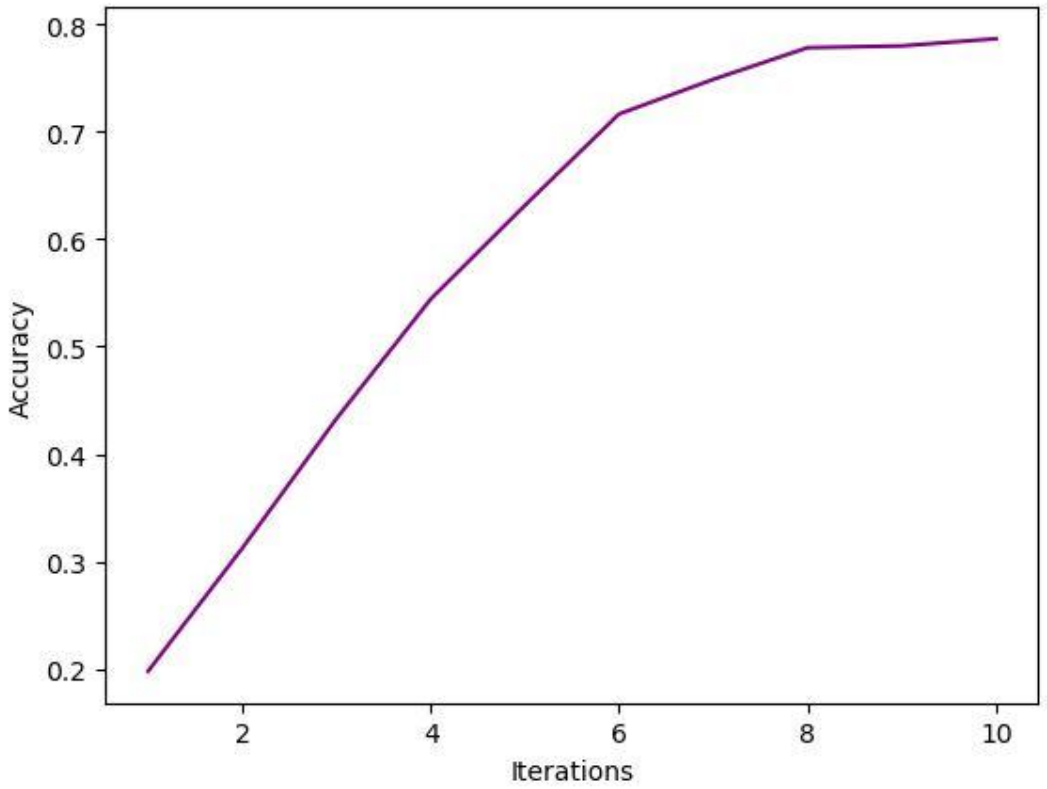
Assignment 1 Analysis

- Step 2:
  - How does the average cross validation error change for each of these depths?
    - The average cross validation error decreases as more depth is added to the decision tree. Here is a graph that depicts how the average accuracy rises as the depth of the tree gets larger. If the average accuracy rises this means the average cross validation error is decreasing as the depth of the decision tree grows.



    -
    - Here are the exact values (rounded) of the average cross-validation accuracy as the depth changes. It will become apparent that as the tree gains more depth, it will become more accurate. This means that as the depth increases the average cross validation error will decrease.
    - Depth 1: 20% accuracy
    - Depth 2: 31% accuracy
    - Depth 3: 43% accuracy
    - Depth 4: 54% accuracy
    - Depth 5: 63% accuracy
    - Depth 6: 71% accuracy
    - Depth 7: 74% accuracy
    - Depth 8: 77% accuracy
    - Depth 9: 78% accuracy

        - Depth 10: 78.6% accuracy
    - Overall to summarize, as the depth of the decision tree grows the average cross validation error decreases.
- Step 3:
    - Which digits had the maximum error? Briefly summarize your findings.
        - The digit with the maximum error is five at 31.0% error. The error percentage for digit one and digit eight have the second most error at 30.0% error. The lowest amount of error was 5.0% error and that was for the zero images. The second least error is 8.0% for the two images.The overall error percentage of the prediction is 19.0%. The total number of correct predictions is 725. Total total number of incorrect predictions is 174. The amount of guesses it had to make is 899. Overall, I don't think it was perfect, but I do think it was very accurate with over a 30% accuracy rate. If it was getting a grade it would get a B which is well above passing.
- Step 4:
    - Compare the 5-fold cross validation accuracy for the best results you obtained from the decision tree classifier to the 5-fold cross validation accuracy of a K-NN classifier (Experiment with different values of K).
        - From my results, I found that the K-NN classifier yielded better results than Decision trees. I also found that the number of neighbors does not really matter when comparing the K-NN classifier to decision trees. This is because the K-NN classifier in my testing always had higher accuracy than the decision trees; regardless of the amount of neighbors. So to summarize, the K-NN classifier performed better than decision trees regardless of the depth of the decision tree and the number of neighbors of the K-NN classifier.
    - Which value of K gave you the best results?
        - None of the values of K gave me the best results. I ran 4 different test changing K to 1, than 3, than 6, than 9. After testing each K-NN classifier with different neighbors, my testing revealed that in this testing case the number of neighbors did not affect the accuracy of the 5-fold cross valdiation. Overall, changing the number of neighbors did not affect the accuracy of my K-NN classifier and in all of my test the K-NN had higher accuracy than decision trees regardless of it's number of neighbors.