AMSC660 Homework #12

Kecai Xuan

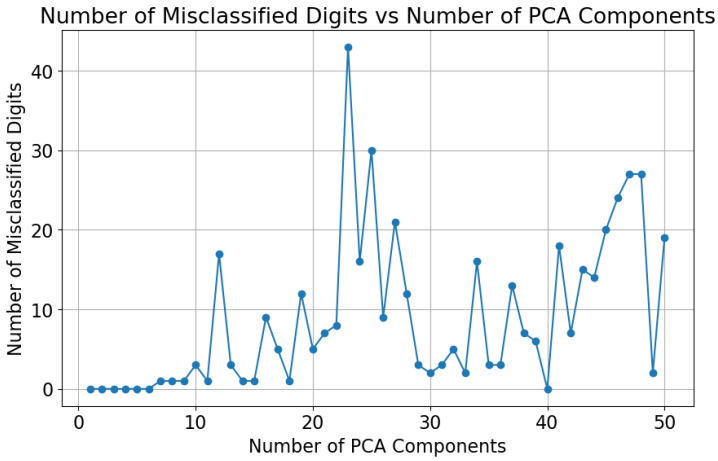
December 4, 2024

The code can be find at:

https://github.com/Bessgendre/minst-optimization-AMSC660-2024

Task 1

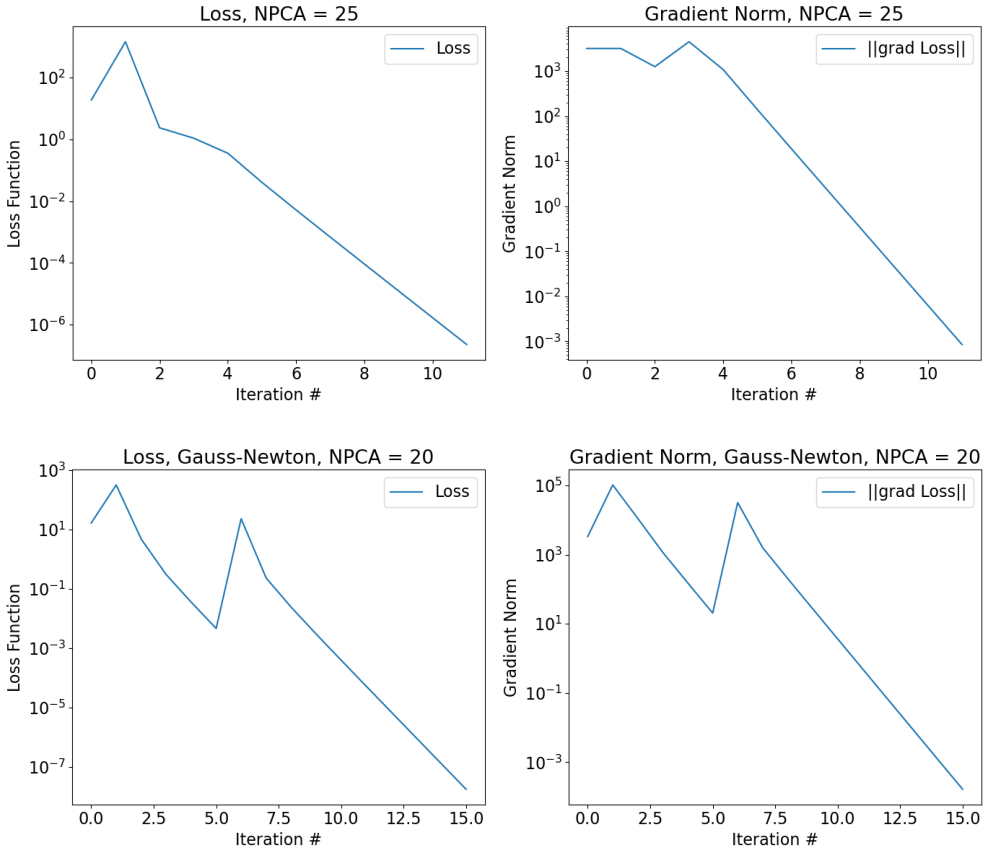
The number of misclassified digits first rises up when PCA components increases, and then decreases. The optimal number of PCA components is around 30, which gives the lowest number of misclassified digits.



This is the learning curves when number of PCA components equals to 25:

Task 2

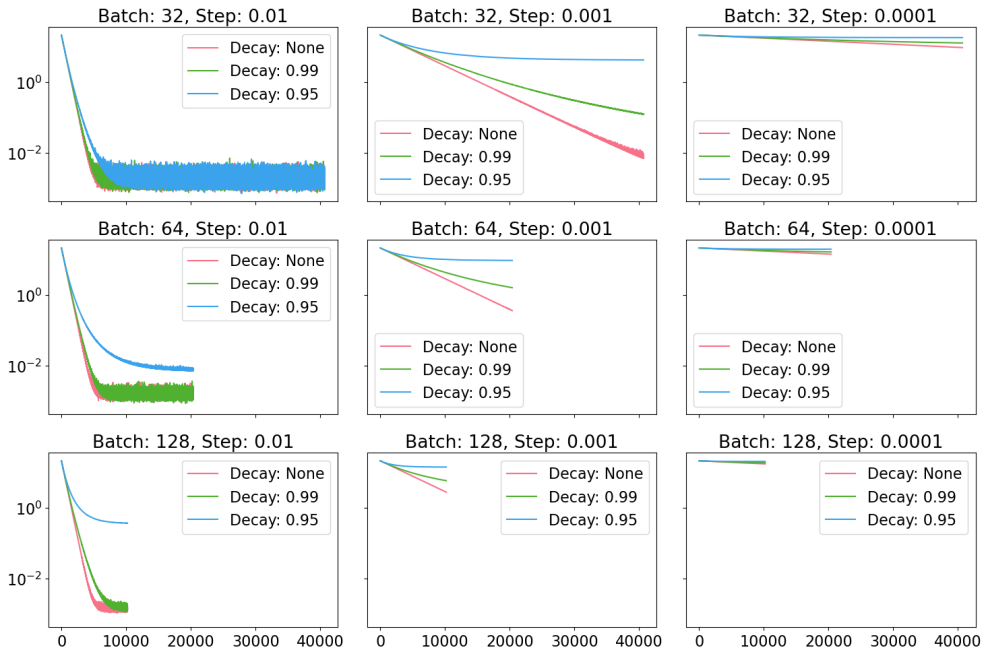
This is the learning curves when using Gauss-Newton method:



Task 3

Conclusions

- **Reasonable Batch Sizes:**
  - Batch sizes of 64 and 128 strike the best balance between computational cost, convergence speed, and stability.
- **Reasonable Step Sizes:**
  - Step sizes of 0.01 are ideal for faster convergence when paired with decay strategies (e.g., 0.99).
  - Smaller step sizes (e.g., 0.001) are also effective but slower, especially for larger batch sizes.
- **Number of Epochs:**
  - For batch sizes of 64 or 128, around 10,000 iterations (or approximately 150–200 epochs) are generally sufficient to reach convergence.
- **Effectiveness of Decreasing Step Size:**



– Using a decreasing step size is beneficial for this problem as it stabilizes convergence, particularly for larger step sizes and smaller batch sizes.

– Decay strategies (e.g., 0.99) strike a good balance between speed and stability.