

CS4361/5361 Machine Learning

Fall 2020

Dealing with small datasets

The attached program cell simulates the situation where we have a small dataset of labeled examples. Your task consists of implementing several techniques to improve accuracy in this situation.

Implement and test the following techniques (you may combine them):

1. Data augmentation for training. For each training image x , create 8 additional ones by shifting x by one row and/or one column in all 8 directions.
2. Data augmentation for training. Generate additional images by averaging two images of the same class using random weights, as described in the slides.
3. Data augmentation for testing. For each testing image x , create 8 additional ones by shifting x by one row and/or one column in all 8 directions. Average the prediction for the 9 images and return the maximum as the final prediction for x .
4. Semi-supervised learning. Use the unlabeled set (without using the y values) to try to improve your results
5. Adversarial learning.
6. Any other technique you want, as long as it doesn't involve accessing additional data.

Submit by the deadline a 1-2 page report describing your results and the program that generated your best results. We will give extra credit (50%, 30%, 15%) to the best 3 performing models. You may use Google colab or Spyder.