Malcolm Snyder | John Santaguida

Projects in ML – Fish Species Identification

Dr. Kunda

Assignment 4

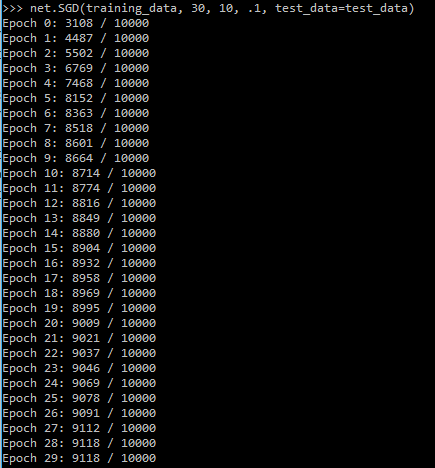
1. Description of Dataset

We researched the most common genomes of fish and found that 94% of fish are categorized under 8 general families, so we chose this as a starting place for our species classification. Having selected these 8 families to shape our target concept, we utilized a variety of recreational fisherman resources online to collect 30-80 pictures of each fish type (based on what data was available) until we ended up with right around 500 images to be used for this assignment. Most of these photos had the fish fairly clear to the camera, held up in “trophy” poses, although we included pictures of lower qualities/different angles as well to maximize our training variety.

**>Insert description of how images were split into training/testing/validation**

Prior to any data manipulation, we used random number generation to select indexes of images to pull into validation or test set and remove from training set.

1. Neural Network Experimentation, Analysis, and Results
2. MNIST Experiments
3. Kept number of epochs (30) and mini-batch size (10) constant
4. Learning Rate set to [.1, 1, 3, 5, 10, 30]
5. 8 GB ram, 64bit Windows OS, Intel i7 processor – took ~5 seconds per epoch
6. See screenshots on following pages
7. Even with a learning rate of .1, we can recognize 91% of digits by the 30th epoch. Conversely, a learning rate of 30 leads not to the highest learning but to overfitting and error



1. Our Experiments
   1. Kept Epochs Constant (50)
   2. Learning rates of [0.5, 1.0, 3.0, 5.0, 10.0, 30.0]  
      Mini batch sizes of [2, 3, 4, 5, 6, 7, 8, 9, 10]
   3. 8 GB ram, 64bit Windows OS, Intel i7 processor– took ~5 sec per SGD call
   4. SEE RESULTS ON LAST TWO PAGES – GROUPED BY BOTH HYPER PARAMETERS
   5. Batch size not huge effect, increasing learning rate started having an adverse effect at 10%

