DEEP LEARNING THE RELATIONSHIP BETWEEN VISUAL AND TEXTUAL DATA

Patrick Poirson
University of South Alabama
Wrote code to train a multimodal SDA
Trained a model with the code and flickr data
Model Trained

- First layer has 0.4 zero input mask
- Trained using 30 epochs
- With 20k training examples
Results

- Measured the euclidian distance between input and output
- Reconstruction error of 24.83% for the tags
- Reconstruction error of 32.24% for the images
Finish code for training and testing multimodal neural network
  - Have code done for the feed forward pass
  - Still need to write the code for the back propagation
Test neural nets and compute results
Start fine tuning hyperparameters