

A systematic neural net building process: Assuming you already built a simple model and perfected the engineering



Using transfer learning, vectorize text, images and audio data

Build a linear model and arrive at a bench mark

Based on the data decide how big the net can be and identify several distributions of nodes and layers

Randomly initialize the weights. Make predictions. Pick top three architectures based on error metric with random weights

With default weight decay and other weak hyper parameters, search for a good learning rate (logarithmic search and for 100 mini batches)

T-SNE the embeddings for visual perfection
Partial dependency plots

You can then experiment with momentum (pick 100 values on a log scale between 0.001 to 0.1 and momentum is 1-the value. Pick the best momentum based on loss after 100 mini batches.

Experiment on a log scale for the best weight decay using K-Fold validation on training data

Fix activation (RELU), initialization (Modified Xavier), Dropout (10% input layer and 50% hidden layers), logistic loss, ridge regularization and 128 mini batch. Adam update with 0.9 momentum and 0.999 RMS prop