Machine Learning - CSCI 5622

Data and Baseline Milestone

Exploring bayesian methods to improve neural network learning

Santhanakrishnan Ramani, Shane Grigsby[1], Shirly Montero, Shruthi Sukumar

Friday 18th November, 2016

Data Description

For this project we use two data sets:

(a) **World Development Indicators Dataset:** The World Development Indicators from the World Bank contain over a thousand annual indicators of economic development from hundreds of countries around the world. This dataset was used as downloaded from the cited source and the indicator we predict is life expectancy.

(b) **NASA ICESat/IceBridge Dataset:** The IceBridge DMS L3 Photogrammetric DEM (IODMS3) data set contains gridded digital elevation models and orthorectified images of Greenland and Antarctica derived from the Digital Mapping System (DMS)[1] [2]. This is a data set of waveforms acquired by the shuttle laser altimeter. We selected a sub-set and generated more input data by adding noise and shifting the waves along an horizontal axis.

Method

For predicting life expectency from the World Development indicators, a very simple neural network is used with a single hidden layers. The following hyperparameters can be passed to a function that generates the required architecture (this ability is required for hyperparameter tuning using GridsearchCV or Spearmint):

(a) Number of hidden neurons

(b) Weight Decay parameter

(c) Dropout ratio

(d) No. of training epochs

In order to obtain a baseline, we built a de-noising convolution auto-encoder (keras.layers) with hyperparameter tuned using grid search, and MSE as the loss function. In addition, we automated the process of building a neural net passing the following parameters for the Autoencoder network,

(a) number of kernels and dimensions (d,1)

(b) activation functions (relu, sigmoid)

(c) weight regularizers (l1, l2)

(d) weights initialization functions (glorot_normal, glorot_uniform)

[1]Not enrolled in CSCI 5622

Results

Just doing backpropogation on the simple neural network for the World Development Indicator and tuning parameters using GridSearchCV, we get the following results.
Best Negative Log-likelihood (5-fold cross validation Gridsearch) : -4057.612
Best Parameters :

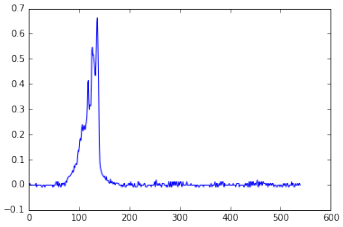
(a) l2 weight decay= 0.01

(b) No. of hidden neurons = 50

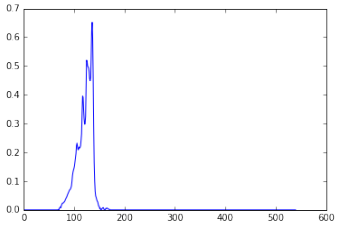
(c) Dropout ratio = 0.5

(d) No. of epochs = 50

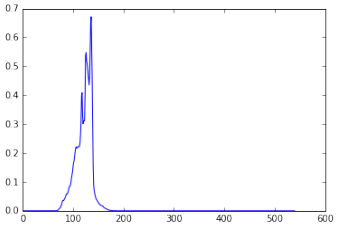
The following results are for the denoising convolutional autoencoder. Seen are figures of input, output and target images.



(a) Input



(b) Output



(c) Reference

Next Steps

The idea would be to implement a bayesian algorithm based on Gaussian Processes to tune hyperparameters of the network to show an improvement. In addition, bayesian algorithms like Markov chain Monte Carlo and Variational Inference will be used to substitute backpropogation to train neural networks in order to get rid of the black-box nature of training in these networks. The end results would be to show improvement over the baseline in performance using these techniques.

References

[1] Zwally J. et al. (2002). Icesats laser measurements of polar ice, atmosphere, ocean and land, j. geodyn., 34, 405 445.

[2] Dotson Ryan C. and John Arvesen 2014. Icebridge dms l3 photogrammetric dem, [indicate subset used]. boulder, colorado usa: Nasa daac at the national snow and ice data center.