

INDEPENDENT FINAL ASSIGNMENT

T81-558: Applications of Deep Neural Networks, Washington University

April 8, 2018

The final assignment is flexible. The primary objective is for you to apply what you have learned this semester to a new dataset. If you have your own idea for an application of deep learning to some objective, please discuss it with me. If there is something you are particularly interested in exploring, I would like to accommodate you.

First you must choose a dataset. This should be a dataset that you have access and are interested in. You can also choose any dataset from the University of California Irvine, Machine Learning Repository ([click for link](#)). Many of the datasets at UCI are not necessarily good choices for deep learning, if in doubt, ask me. Some datasets that I suggest are:

- Abalone Data Set ([click for link](#))
- Liver Disorders Data Set (BUPA) ([click for link](#))
- Covertypes Data Set ([click for link](#))
- Credit Approval Data Set (CRX)([click for link](#))
- Ecoli Data Set([click for link](#))
- Forest Fires Data Set)
- Glass Identification Data Set ([click for link](#))
- Heart Disease Data Set([click for link](#))
- Hepatitis Data Set([click for link](#))
- Horse Colic Data Set([click for link](#))
- Housing Data Set ([click for link](#))
- Pima Indians Diabetes Data Set([click for link](#))
- Soybean (Large) Data Set([click for link](#))
- Wine Data Set([click for link](#))

You should read the links above for more information on the datasets. I also have the above datasets properly formatted for Pandas at the following link: ([click for link](#))

Once you select a dataset create a Jupyter notebook or Python script that performs the following:

- Create a training and validation set.
- Train using a neural network that is large enough for your dataset.
- Choose a field to predict, decide if you are performing classification or regression.
- Choose (and use) an evaluation metric for your network (See Class 3).
- Choose (and use) a graphic evaluation for your network (See Class 4).

If you used Jupyter include a markdown section that provides a brief report on your model. If you used a Python script (no notepad) attach a Word document (please provide your output in either a text file or your notebook). This report will likely be 1/2 page to 1 page, but should include:

- How did you encode the input variables?
- How did you choose how to evaluate the accuracy of your model?
- Why is your model either classification or regression
- Were there any missing data? How did you handle it?
- Mention any other challenges or difficulties you had with the dataset.