

KERAS

- What is it?
 - API calls. Import into model python file. Has functions related to NN
 - API is a way to build a graph of layers, as a NN is an acyclic directed graph
- The INPUT
 - Need the dimension of the data you are using.
 - For example, a picture that is 32 pixels by 32 pixels has a shape of 32,32,3, the 3 being the RGB values
- DENSE
 - A new layer in the neural network.
 - Like drawing an arrow from one layer to another
 - Full understanding TBD
- SUMMARY
 - Outputs a table that explains the layer type, shape, and number of parameters of each layer
- PLOT_MODEL
 - An easier way to visualize the graph and the dense layers
- COMPILE
 - Understanding of optimizers TBD
 - Need a loss function (I think several are provided by keras API)
 - Need a metrics function (I think also provided by keras)
- FIT
 - Needs your input data like a numpy array

- Needs target data structure, again like a numpy array
- Batch size understanding TBD
- Epochs just seems to be a unit of measurement or number of cycles to train the data, an “iteration of the entire x and y datasets”
- Validation split sets aside a certain amount of the training data to use to measure loss after training
- SAVE
 - Saves the model as a file path
 - Overwrite variable is boolean and says if we should overwrite an existing model at location
 - Load_model looks for a file and loads a NN model from that file

Revelations Related to KERAS

- The .keras files are saved, trained models. It makes sense we cannot open them, as it is information we are letting keras take care of for us.
- You can use .save to save a model, delete it from current memory, and reload it with .load_model
- Start by creating an input layer. Create multiple layers with keras.layers.Dense until you define an output, then create the “Model” class that uses the input variable as input and the output variable as output

JOBLIB

- What is it?
 - A set of pipelining tools to help with disk-caching and parallel computing
 - From what I can tell, only used for .dump function
- DUMP
 - Dumps a python object into a file
 - Used to save the model
 - This is where the pkl file comes from. It saves the PCA model

SKLEARN (SciKit Learn)

- What is it?
 - Simple predictive tools for data analysis
 - This is our data science package
- PCA
 - Principal component analysis
 - Uses singular value decomposition to reduce dimensionality of datasets. This is how we narrow down something with a lot of dimensions to just a few for analysis
- StandardScaler
 - Understanding TBD
- Train Test Split
 - Splits an array randomly into a training set and a testing set