Three use cases for arrays.

1. Store statistical metrics for a series of candidate models

If you want to optimize the performance of a model you might want to do a grid search of hyperparameters. In the mocked-up code below I’m doing a grid search of two hyper parameters for an XGBoost model. Depthlist is a list of potential depths of the trees that make up the model. EtaList is a list of potential shrinkage factors or learn rates for the model.

I create an array initialized to zero. The array has dimension 3 horizontally and dimension len(DepthList)\*len(EtaList) = 6\*7 = 42 in the vertical dimension.

For each combination of depth and eta I fit an XGBoost model and calculate the AUC statistic. The AUC statistic is an omnibus measure of how well a model that predicts a binary outcome sorts the data from most likely to have an outcome of 1 to least likely.

Then starting with row i=0, I enter the value of Depth in column 0, eta in column 1 and auc in column 2.

I increment the value of I by 1, fit a new model on the next pair of hyperparameters and then store the hyper parameters and auc in the next row. And so on.

DepthList = [2,3,4,5,6,7]

EtaList =[0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, 0.0001]

arrayYDim = len(DepthList)\*len(EtaList)

AucArray = [[0]\*3 for j in range(arrayYDim)]

i=0

For Depth in DepthList:

For eta in etalist:

model = XGBRegressor(n\_estimators=400, max\_depth=7, eta=0.01,)

model.fit(X\_train[KeepVariables], y\_train)

y\_pred = model.predict(X\_test[KeepVariables])

auc = metrics.roc\_auc\_score(y\_test, y\_pred)

arrayYDim[i,0] = Depth

arrayYDim[i,1] = eta

arrayYDim[i,2] = auc

i += 1

Important caveat. I do not have a good data set to test this code on so, given my tendency for typos, it probably will not work on the first try, but something like this should work.

1. Store information about pixels in a picture

Another application is to take a picture in grey scale, create an array that has the same dimension as the horizontal and vertical pixel count of the picture and then for each position in the array store the integer that describes the darkness of the pixel in the same location in the picture. For example, if the pixel in the third row and second column had a grey scale value of 120, I would put the number 120 in the third row, second column of the array.

1. Store values of variables to be used to fit a model

Suppose I have 1000 rows of data and each row has 50 numeric valued variables. I could fix the order of the variables for each row and then create a 1000 row by 50 column arrays. Then for each row I could enter the values of the variales in the columns in the order I create. So now the data is stored in an array and I can use it to fit a model.