Predictive Analytics World / Deep Learning World
Exercises - The Boston Housing Dataset

1. Write a Keras/TensorFlow program to create a neural network regression model for the Boston Housing Dataset. Use a 90%-10% split for training and test data. Use a 13-(10-10)-1 architecture, tanh hidden activation. Use Glorot uniform weight initialization and zero initialization for biases. Train using Adam with a learning rate of 0.005, beta1 = 0.90, beta_2 = 0.999 with no momentum or weight decay. Use online training rather than batch training.

2. Modify your program to see if you can improve prediction accuracy (within 15% of actual) of the test data to greater than 90% by modifying one or more of the following hyperparameters. Be sure to save your models.

* architecture: change number of hidden layers and/or number hidden nodes
* architecture: add one or more dropout layers
* initialization: change to RandomUniform or RandomNormal or he_normal
* optimizer: change max epochs and/or batch size
* optimizer: change the parameters of the Adam optimizer
* optimizer: add momentum
* optimizer: use SGD or RMSprop or Adagrad or Adadelta

3. Which statement is most accurate?

a.) Dropout is a technique to both increase training speed and combat overfitting.
b.) Dropout is a technique to increase training speed, often at the expense of overfitting.
c.) Dropout is a technique to combat overfitting, often at the expense of training speed.

4. Which statement is most accurate?

a.) Dropout rates are typically between 0.90 and 0.95.
b.) Dropout can be thought of as a form of model averaging.
c.) A dropout layer is placed before the layer you wish to influence.

5. Which statement about regularization is most accurate?

a.) L1 often leads to many weights that are close to, but not equal to 0.0.
b.) L2 often leads to many weights that are close to, but not equal to 1.0.
c.) L1 penalizes squared weight values, and L2 penalizes absolute weight values.

6. Which statement is most accurate?

a.) A sparse network is one where many node values are close to 100.0 (can be good or bad)
b.) A saturated network is one where many nodes are at the limits of their activation range (usually bad).
c.) Feature selection is a general term used to describe algorithmic techniques to tune hyperparameters.