Predictive Analytics World / Deep Learning World
Exercises - The Iris Dataset

1. Write a Keras/TensorFlow program to create a neural network prediction model for Fisher's Iris Dataset. Use a 120-30 item split for training and test data. Use a 4-5-3 architecture, tanh hidden activation, softmax output activation. Use uniform weight initialization in [-0.01, +0.01] and zero initialization for biases. Train using stochastic gradient descent with a learning rate of 0.01, no momentum or weight decay. Use online training rather than batch training.

2. Modify your program to use 7 hidden nodes with logistic sigmoid activation, and to save the trained model as "iris_model.h5".

3. Modify your original program to use SGD with a learning rate of 0.05, momentum = 0.50, and a batch size = 4.

4. Which statement is most accurate?

a.) Cross entropy error/loss is generally preferred for classification problems, and mean squared error is preferred for regression problems.
b.) Cross entropy error/loss is generally preferred for regression problems, and mean squared error is preferred for classification problems.
c.) Cross entropy error/loss is generally preferred over mean squared error for both classification and regression problems.

5. Which statement is most accurate?

a.) Stochastic gradient descent (SGD) optimization is sometimes called back-propagation and is most often used for relatively simple neural networks.
b.) Stochastic gradient descent (SGD) is generally preferred for very deep neural networks, and Adam (adaptive moment estimation) is generally preferred for simple neural networks.
c.) RMSProp (root mean squared propagation) and Adagrad (adaptive gradient) are generally preferred over Nadam (Nesterov Adam) and Adamax (Adam infinity norm).

6. Which statement is most accurate?

a.) Momentum is optional and is used to prevent model overfitting.
b.) Nesterov momentum is optional and is the most basic form of momentum.
c.) Momentum is optional and is used to speed up training.

7. If N is the number of training items, which statement is most accurate?

a.) Using a batch size of N / 2 is often called batch duplex training.
b.) Using a batch size of 1 is often called online training.
c.) Using a batch size of N is often called dropout training.