Predictive Analytics World / Deep Learning World
Exercises - Fundamentals

1. Write a Python program that creates a 4-5-3 single hidden layer neural network classifier. Use tanh hidden layer activation and softmax output layer activation. Set input-hidden weights to 0.01 to 0.20. Set hidden biases to 0.21 to 0.25. Set the hidden-output weights to 0.26 to 0.40. Set the output biases to 0.41 to 0.43. Compute the output values of [0.3151, 0.3330, 0.3519] for input values = [1.0, 2.0, 3.0, 4.0].

2. If you modify your program to use logistic sigmoid hidden node activation, using the same weights and input values, what is the output?

a.) [0.3151, 0.3330, 0.3519] (the same as using tanh)
b.) [0.3165, 0.3330, 0.3505]
c.) [0.2876, 0.2243, 0.6001]

3. Suppose a deep neural network has a 5-(6-4)-3 architecture. How many weights and biases does the network have?

a.) 5 * (6 + 4) + (3 * 3) = 59
b.) (5 + 6 + 4) * 3 = 45
c.) (5 * 6) + 6 + (6 * 4) + 4 + (4 * 3) + 3 = 79

4. Suppose a NN classifier has three pre-softmax output values = [1.50, -1.00, 0.50]. What are the softmax values?

a.) exp(xi) / exp(1.50 + (-1.00) + 0.50) = [0.55, -0.37, 0.18]
b.) xi / (1.50 + (-1.00) + 0.50) = [0.50, 0.40, 0.10]
c.) exp(xi) / (exp(1.50) + exp(-1.00) + exp(0.50)) = [0.69, 0.06, 0.25]

5. Suppose you have a predictor variable age, with three values = [27, 38, 49]. If you apply min-max normalization, what are the normalized values?

a.) (x - min) / (max - min) => [0.00, 0.50, 1.00]
b.) (x - mean) / sd => [-0.76, -0.66, 1.41)
c.) x / max => [0.55, 0.57, 1.00]

6. Suppose you have a predictor variable color, that can be either red, blue, green, or yellow. Which is the 1-of-(N-1) encoding of the four colors?

a.) (0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0)
b.) (1, 0, 0), (0, 1, 0), (0, 0, 1), (-1, -1, -1)
c.) 1, 2, 3, -1

7. Which of the following very roughly summarizes the UAT / Cybenko theorem?

a.) A single hidden layer can compute anything, but in practice multiple hidden layers can work better.
b.) Two hidden layers can compute twice the entropy of a single hidden layer.
c.) The number of hidden nodes must be greater than twice the dimension of the output variable.