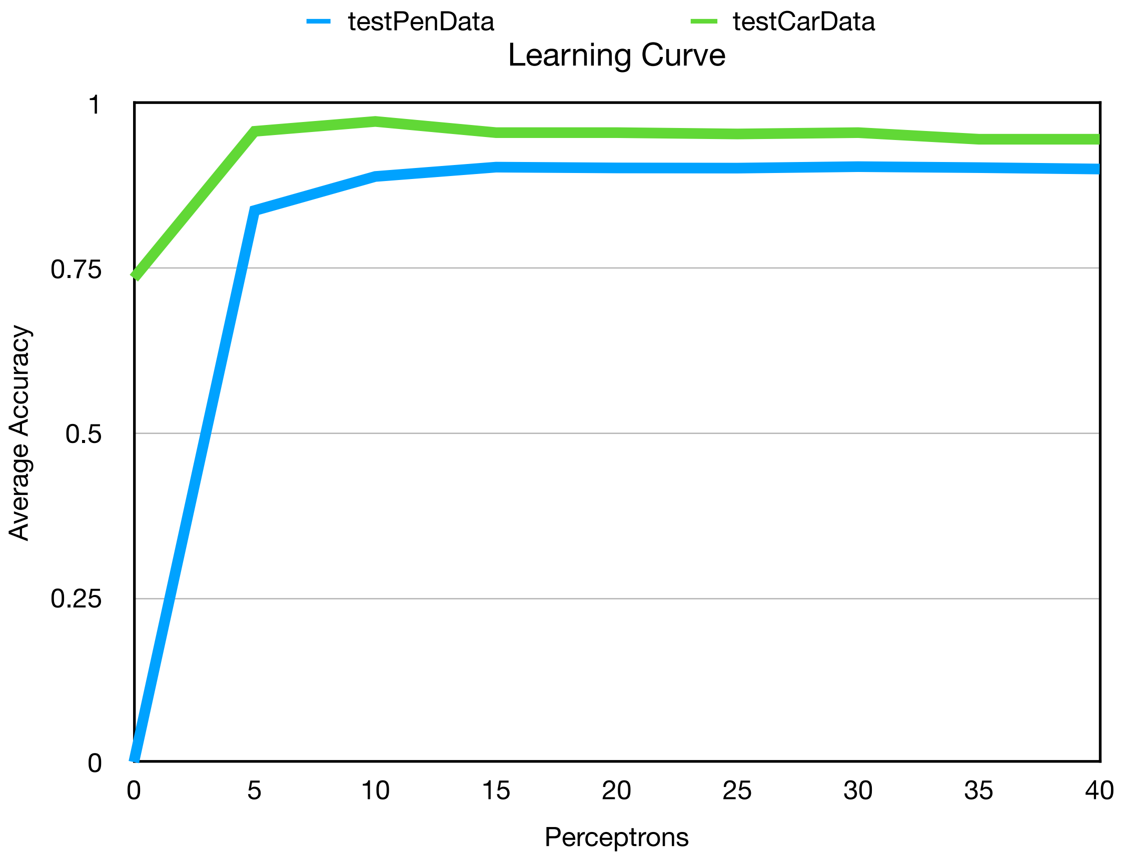
Question 5:

	max	average	standard deviation
testPenData	0.91080617495711800	0.90303030303030300	0.00770364667968082
testCarData	0.99000000000000000	0.98000000000000000	0.00547722557505167

Question 6:

testPenData	max	average	standard deviation
0	0.00000000000000000	0.00000000000000000	0.00000000000000000
5	0.85334476843910800	0.83767867352773000	0.01379743960061950
10	0.89622641509434000	0.88942252715837600	0.00673216056239451
15	0.91052029731275000	0.90360205831904000	0.00478501451181423
20	0.90994854202401400	0.90240137221269300	0.00904675151804162
25	0.90794739851343600	0.90211549456832500	0.00900691534248835
30	0.90909090909090900	0.90434534019439700	0.00273965661537128
35	0.90451686678101800	0.90285877644368200	0.00188245599006910
40	0.90394511149228100	0.90074328187535700	0.00315451440231761

testCarData	max	average	standard deviation
0	0.73500000000000000	0.73500000000000000	0.00000000000000000
5	0.96500000000000000	0.95800000000000000	0.00678232998312527
10	0.98500000000000000	0.97300000000000000	0.01288409872672510
15	0.97000000000000000	0.95600000000000000	0.00969535971483267
20	0.96500000000000000	0.95600000000000000	0.00489897948556636
25	0.96000000000000000	0.95400000000000000	0.00734846922834954
30	0.96500000000000000	0.95600000000000000	0.00583095189484531
35	0.95000000000000000	0.94600000000000000	0.00374165738677394
40	0.95500000000000000	0.94600000000000000	0.00799999999999998



For the above graph, the average accuracy of both pen and car data shows an increasing and converging(at around 0.9) trend when the number perceptron increases. The accuracy in both dataset skyrocketed during 0 to 5 perceptron. The low accuracy at the very beginning is because of lack of perceptron, the model do not have enough feature to predict. Hence, an underfitting occurs. Also both data set in the graph shows a slight decrease at the tail, it is because having too much perceptron can overly complex the model. Hence, an overfitting happens. The maximum accuracy of the pen data is at 30 perceptrons. The maximum accuracy of the car data is at 10 perceptrons.

Question 7:
Code:

```
xorData = [[([0, 0], [0]), ([0, 1], [1]), ([1, 0], [1]), ([1, 1], [0])],
           [([0, 0], [0]), ([0, 1], [1]), ([1, 0], [1]), ([1, 1], [0])]]
def testXorData(hiddenLayers=[10]):
    return buildNeuralNet(xorData, maxItr=500, hiddenLayerList=hiddenLayers)
xorPerceptron = [0]
k = 0
while k != 1.0:
    xorPerceptron[0] += 1
    nnet, k = testXorData(xorPerceptron)
nnet, p = testXorData([0])
print([0], p)
print(xorPerceptron, k)
```

Accuracy without hidden layer: 0.5
Max Accuracy with 21 perceptron: 1.0

The result is same as expected. Since xor is not linearly separable, hidden layers are needed to model the network. Therefore, A network with 21 perceptron is best to learn the behaviour and gives the best result. While the accuracy without a hidden layer is similar to randomly guessing the result (0.5).

Question 8:

	max	average	standard deviation
testQ8Data	0.98475609756097600	0.97591463414634100	0.00504662968208747

The data I have used is a banknote dataset by (Brownlee, 2021). It is a binary (0 for authentic, 1 for inauthentic) classification dataset which takes the following attributes:
1. Variance of Wavelet Transformed image (continuous).
2. Skewness of Wavelet Transformed image (continuous).
3. Kurtosis of Wavelet Transformed image (continuous).
4. Entropy of image (continuous).

Firstly, I use getNNQ8Data() to shape the data into a readable format by the model. Secondly, I used buildExamplesFromExtraData() to get both training and testing data from its respective txt file. After that, I use the follow code to call the above functions to shape the data and to execute the learning model with some parameters:

```
xorData = buildExamplesFromExtraData()
def testXorData(hiddenLayers=[25]):
    return buildNeuralNet(xorData, maxItr=500, hiddenLayerList=hiddenLayers)
q8Result = []
for _ in range(5):
    nnet, j = testXorData()
    q8Result.append(j)
print('q8Result, max: ', max(q8Result), ' average: ',
      average(q8Result), ' sd: ', stDeviation(q8Result))
```

The above code will print the max, average and sd of the accuracy in the 5 restarted network.

Reference:

1. Brownlee, J. (2021, October 19). 10 standard datasets for practicing Applied Machine Learning. Machine Learning Mastery. Retrieved April 24, 2022, from https://machinelearningmastery.com/standard-machine-learning-datasets/