Question 5:

See results below. The pen dataset performed slightly better than the car dataset, which may be because neural nets are more suited to continuous data such as the pen data rather than discrete data.

Dataset	Max	Average	Standard deviation
Pen	0.9088050314465409	0.902687249857	0.00598734538208
Car	0.8488219895287958	0.843062827225	0.00296748273528

Question 6: See results and discussion below.

Car data results:

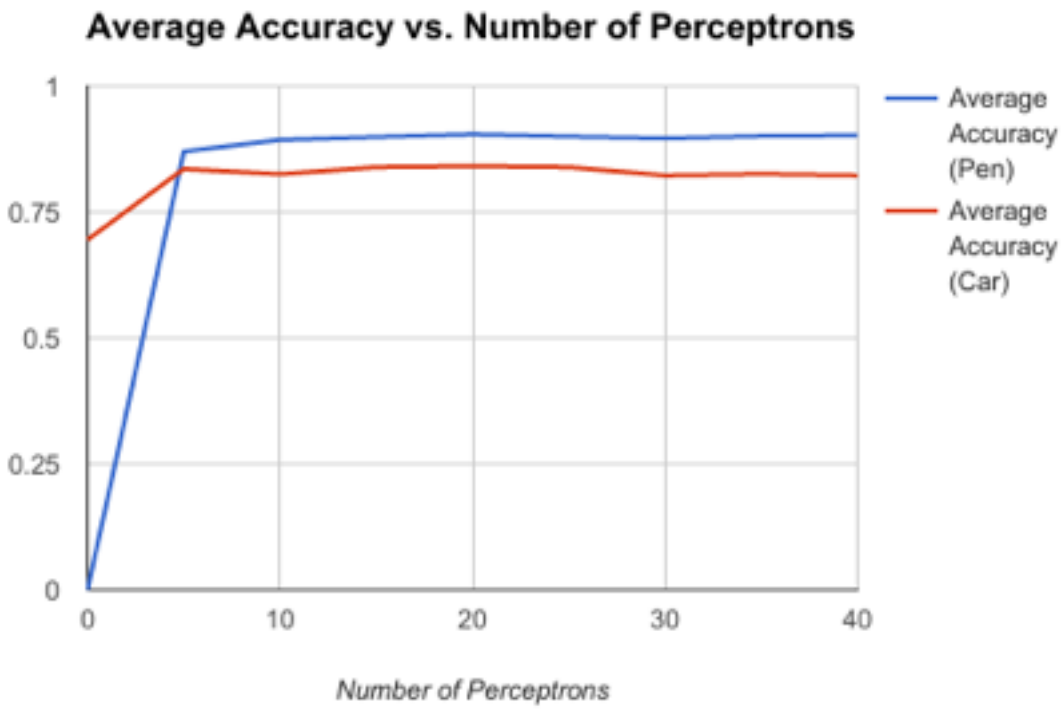
Number of perceptrons	Max	Average	Standard deviation
0	0.0	0.0	0.0
5	0.9088050314465409	0.870268724986	0.0331841765208
10	0.9345340194396798	0.893367638651	0.022029279416
15	0.9082332761578045	0.899142367067	0.0098142257478
20	0.9108061749571184	0.904402515723	0.0061116334667
25	0.9070897655803316	0.900057175529	0.0065729481364
30	0.9073756432246999	0.896512292739	0.00972488349957
35	0.9085191538021726	0.900914808462	0.0067447739751
40	0.9048027444253859	0.902344196684	0.00184916656704

Pen data results:

Number of perceptrons	Max	Average	Standard deviation
0	0.69502617801	0.69502617801	0.0
5	0.8514397905759162	0.835602094241	0.0121340090941
10	0.8324607329842932	0.825130890052	0.0150460947597
15	0.9082332761578045	0.839136125654	0.00717392125971
20	0.8710732984293194	0.841230366492	0.0176512406018
25	0.8534031413612565	0.839136125654	0.0106045178537

30	0.8337696335078534	0.822643979058	0.00767689963278
35	0.8481675392670157	0.825654450262	0.0171740517978
40	0.8298429319371727	0.82277486911	0.0061169222228

Learning curve:



Discussion: In general, the neural network performed better with the pen data than the car data. This is likely because the car data is discrete and therefore not well suited to a neural network (as evidenced by the fact that the car dataset performed significantly better with the decision tree than with the neural network). The pen data, on the other hand, is continuous and therefore appropriate for a neural network classifier.

I would have expected the performance for both datasets to improve consistently as the number of perceptrons increased, which was generally the case for the pen data. However, the performance of the car data fluctuated as more perceptrons were used, which could be because of overfitting to the training set or because of the gradient descent algorithm getting stuck in local minima or maxima.

Question 7: With one perceptron, the neural network had an accuracy between .5 and .75 for five each of five trials. With two perceptrons, the accuracy was between .75 and 1.0. It took three perceptrons to consistently get an accuracy of 1.0. This makes sense given that XOR is a nonlinear function and therefore cannot be learned correctly with one perceptron or particularly reliably with two.

Question 8:

Max	Average	Standard deviation
0.9682539682539683	0.873015873016	0.0481452409146

Discussion: This dataset performed better with the neural net than with the decision tree, which could be because it has numerical (although not continuously valued) attributes. The standard deviation was fairly high; this was mostly a result of the max value, which appeared to be an outlier.