CSE 3521 HOMEWORK 5

BRENDAN WHITAKER

1. (a) This is classification, since there is an existing, categorical (discrete and finite) label for our data which we wish to approximate.
   (b) You are asking the librarians for a clustering. You are not giving them any specific label which says how to specify "similarity" within the first 100 words. We are simply looking to identify a group of books which correlate in some way.
   (c) You have not changed machine learning problems, because the variable we are approximating, time, is still fundamentally continuous. Instead, we are only changing our interpretation of the regression results.

4. (a) For the separable data, the number of iterations increased as the learning rate approached zero. For a very small learning rate, the number of iterations was more than 200, and for a large learning rate of 0.1, we had 8 iterations. The SSE was also largest when the learning rate was lowest, giving a value of more than 20. For the separable data, the test set accuracy was 100% in every run. For the medium data, the test set accuracy dropped to around 90%. For a learning rate of around 0.05, we had a larger number of iterations at around 17, compared to 10 for the separable. At a learning rate of 0.00001, we had only 2 iterations, and very low accuracy on both training and test sets, around 50%. The training error was over 150! This suggests the medium data was more difficult to fit than the separable data. The inseparable data was bad no matter what the learning rate was set to. We had around 50% accuracy for the training and test datasets, and the error was above 100 for all runs. The iterations again were puzzlingly low, all in the single digits.
   (b) When varying the convergence delta on the inseparable data, the number of iterations varied from test to test, but were in the range of 6-40, and the higher numbers correlated with lower convergence values. The SSE was around 100% for all tests, and the accuracy around 45-50% for most tests. For the separable data, the number of iterations, computation time, and accuracy all clearly negatively correlated with the convergence delta values. The SSE decreased to close to 0 as the convergence delta dropped. I observed iterations as high as 1252 as the convergence delta was around 0.0001. For the medium data, the number of iterations dropped from the separable, as well as the accuracy, and the training error increased. However, the correlations remained the same. I noticed that the decision boundary barely moves at all, but the points seem to diffuse over the course of the regression.

5. (a) No features are perfectly discriminative, the best is (1), which gave a 91% test set accuracy reading.
   (b) The best combination I could find was using features 1 and 2, which yielded a test set accuracy of 93%.
   (c) Yes, features 0 and 4 had runs which had test set accuracy as low as 38%, and these features averaged around 53%, which is pretty bad. So I would say throw them out.