Assignment three

- QA1): SVM is a type of supervised machine learning algorithm which is utilized for regression and classification tasks. The difference between SVM with a soft and hard margin is the following:
  Hard margin: The algorithm will search for a hyperplane which efficiently distinguishes the data points of the different classes, but the hard margin only operates whenever the data is linearly separable. This means when there is a clear boundary between the classes. The hard margin model will attempt to maximize the margin between the nearest data point and hyperplane of two classes, but the hard margin will have a low performance if there are outliers in the data.
  Soft margin: This algorithm will permit some misclassification of data points and search for a hyperplane which makes the smallest number of classification errors and has a small margin, the soft margin is utilized when the data is not separable. Another key difference between the two is that the soft margin adds a penalty term to the objective function, which is minimized during training, the point of the penalty is to improve the classification accuracy by classifying some of the data points incorrectly.
  Overall, the main difference between the two is that the hard margin attempts to perfectly separate the data points while the soft margin will allow some misclassification.

- QA2) The C is a hyperparameter which controls the barter between achieving a wider margin and permitting some misclassification of data points. The C balances the margin width and misclassification penalty and thus operates an important role in the performance of the classifier. A large C value has a high bias and a low variance model, this means the classifier will attempt to reduce the number of misclassified data points, even though it could result in a smaller margin. A small C value has a low bias and high variance model, which means that it has a wider margin but will permit for a larger number of misclassifications. The complexity of the problem and the nature of the data will determine the optimal value of the C parameter, should the data points not be well separated a small C value is best and if the data is well separated a large C value may be best. Overall, the C value can have a significant impact on the performance of a soft margin SVM classifier.
-
- QA3) The Preceptron will not be activated due to the weighted sum being smaller than the 2.8 activation threshold.

- QA4) The delta rule is a learning rule which is utilized in supervised learning algorithms to update the weights of a neural networking during the training. Alpha which is known as the learning rate determines the size of the step taken in the direction of the steepest descent of the cost function. When the learning rate is too big, the updates of the weight can become unstable and lead to a slow convergence or divergence, and on the opposite end if the learning rate is too small the updates of the weight can also be too small and will lead to a slow convergence and longer running times. This displays the importance of the learning rate and how crucial it is to the performance of the delta rule. An excellent learning rate choice will speed up the convergence, while a bad learning rate choice can result in the failure of a convergence altogether. Choosing the best learning rate is dependent on the data and architecture of the neural network.

- QB1) The R Squared value of the model is 0.34%
- QB2) The best C value is 0.5 and this is where the RMSE was at 2.29.
- QB3) The Value of the R-Squared with the highest parameter is 0.34% and that is when the size is equal to 1 and the decay value is equal to 0.1.
- QB4) Using the given record the estimated sales value is 4.97.