ML Assignment 2

Harshvardhan Kalra
2014043

Maximum iterations for every part were 10000. Since random initialization of theta was giving a slightly worse performance than setting it to all zeros, the latter was done.
1) Linear regression on lin.txt and sph.txt was carried out with the following parameters:
    a) lin.txt : Learning rate - $2.5 * 10 \char94 -4$
    b) sph.txt : Learning rate - $2.5 * 10 \char94 -6$
2) Polynomial and Gaussian kernels were used with the following parameters:
    a) lin.txt : Learning rate (polynomial, degree 2) - $5 * 10 \char94 -7$
    b) lin.txt : Learning rate (Gaussian) - $1 * 10 \char94 -3$
    c) sph.txt : Learning rate (polynomial, degree 2) - $5 * 10 \char94 -10$
    d) sph.txt : Learning rate (Gaussian) - $5 * 10 \char94 -5$
3) For all kernels above, the error decreased as the amount of training data was increased, which coincides with the expected behaviour. The plots also indicate the same.
    However, the Gaussian kernel doesn't fit well for sph.txt, probably because a single Gaussian was being fit over the curve. The learning rates were chosen after a lot of trial and error, and degrees were taken as 2, as they gave a slightly better fit than degree 3.
    For lin.txt, the Gaussian kernel seemed to approximate the data best, followed closely by both the linear and 2-degree polynomial kernels.
    For sph.txt, The 2 and 3-degree polynomial kernels gave a nearly perfect fit, while the Gaussian kernel gave an extremely poor fit. The linear kernel gave an approximate fit, much better than the Gaussian kernel.
4) Since ridge regression imposes stricter constraints on the curve being fit, increasing the value of initial_del worsens the MSE. Hence, the optimal value should be 0, even though that would allow underfitting/overfitting to occur, at the cost of lower MSE. The plots for lin.txt and sph.txt were computed with initial_del as 1, while the UCI datasets were computed with 0.1.
5) The results from 10-fold cross validation were as follows:
    a) iris.data : Learning rate - $8.5 * 10 \char94 -10$, mean - 1.85492441771, SD - 0.775307560832
    b) seeds_dataset.txt - Learning rate - $8.75 * 10 \char94 -10$, mean - 0.869637747166, SD - 0.854762387454
    c) AirQualityUCI.csv - Learning rate - $5 * 10 \char94 -13$, mean - -6.46732391725, SD - 29.848106503

All associated plots are provided in the corresponding folders, along with the curve for Q3.