E0-214 Applied Linear Algebra and Optimization

Assignment 4

The aim of this assignment is to study and compare different optimization algorithms for solving the scaled ridge regression problem discussed in the class.

The dataset "Assignment4Dataset.csv" is provided. Every instance of the dataset is two dimensional. The last column denotes the class label ($+1$ or -1) for each instance. You can generate new features from the given features using some function $\phi(\boldsymbol{x})$. Use the first 400 examples as a training set and the remaining 4900 examples form the test set ($n = 400, ntest = 4900$). The classifier model to be designed is $f(\boldsymbol{x}; \boldsymbol{w}) = \boldsymbol{w} \cdot \phi(\boldsymbol{x})$ where the model parameter $\boldsymbol{w}$ is obtained by solving the scaled Ridge Regression problem for the training data set, $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$

Compare the following methods (after tuning the hyperparameter λ using k-fold cross-validation technique):

1. Steepest Descent Method (using backtracking line-search)
2. Mini-batch gradient method (Try different batch sizes)
3. Stochastic gradient method

Plot the objective function value vs number of iterations for each method and compare the speed of the methods. Compare the performance of the models (obtained using each of the above-mentioned methods) on the test set in terms of classification accuracy. For example, if $\boldsymbol{w}_{SD}$ denotes the model parameter when the steepest descent algorithm terminates, then the test set accuracy can be calculated as:

$$TestAccuracy_{SD} = \frac{1}{ntest} \sum_{i=1}^{ntest} I(y_i = sign(\boldsymbol{w}_{SD} \cdot \phi(\boldsymbol{x}_i))$$

where $I(\cdot)$ is the indicator function which outputs 1 if the argument is true and 0 otherwise.