Lab Course Machine Learning

Exercise Sheet 6

Prof. Dr. Dr. Lars Schmidt-Thieme
Kiran Madhusudhanan

HiWi: Basharat Mubashir Ahmed
Sophia Damilola Lawal
Submission deadline : December 18, 2022

General Instructions

1. Perform a data analysis, deal with missing values and any outliers.
2. Unless explicitly noted, you are not allowed to use scikit, sklearn or any other library for solving any part.
3. Data should be normalized.
4. Train to Test split should be 80-20.

1 Implementing Coordinate Descent (7 points)

This week the main task is to implement Coordinate Descent that has been covered in the lecture. To make things a bit more interesting, we will be implementing Lasso Regression along with the Coordinate Descent. You should NOT use scikit-learn for this question. We will use the regression.npy dataset for this question.

1. Coordinate Descent.
   a) Implementing the Coordinate Descent algorithm.
   b) Maintain a history of your β values. After training, plot them against iterations; plot the $betas$ in a single plot. This should show you the progression of your feature values($betas$) as your train the model.
   c) Coordinate Descent with L1 Regularization
      i. Implement CD with L1 regularization
      ii. Maintain a history of your β values. After training plot them against iterations.
2. Compare the plots of the unregularized and regularized CD
3. How is the Coordinate Descent method different from SGD and Newton's method? In which case is it advisable to use the CD method?
4. **Note:** Ensure you provide adequate comments where necessary in your code, algorithm implementation without comments will be penalized.

2 Accelerating K-Nearest Neighbour Classifier (13 points)

Load the dataset **classification.npy**, the dataset consists of over 100 predictors.

1. Implement the following versions of the Nearest Neighbour Classifier
   a) Vanila KNN algorithm
   b) Partial Distances/Lower Bounding(slide 32-38)
   c) Locality Sensitive Hashing(slide 43)

2. Experiment with k = [1, 2, 3, 4, 5, 7] and report your accuracy on the test set.

	1	2	3	4	5	7
Vanila KNN						
Partial Distances						
Locality Sensitive Hashing						

   Furthermore, report the runtime.

	1	2	3	4	5	7
Vanila KNN						
Partial Distances						
Locality Sensitive Hashing						

3. Experiment with distance=[cosine,euclidean,cityblock]. You can use scipy for the distance, report accuracy on the test set

	Cosine	Euclidean	Cityblock
Vanila KNN			
Partial Distances			
Locality Sensitive Hashing			

4. How is the NN algorithm different from the algorithms we have studied so far?

5. It is advisable to use class implementation for this task.

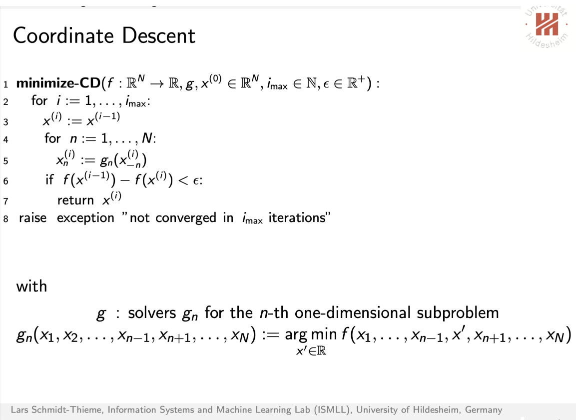

Figure 1: Coordinate Descent

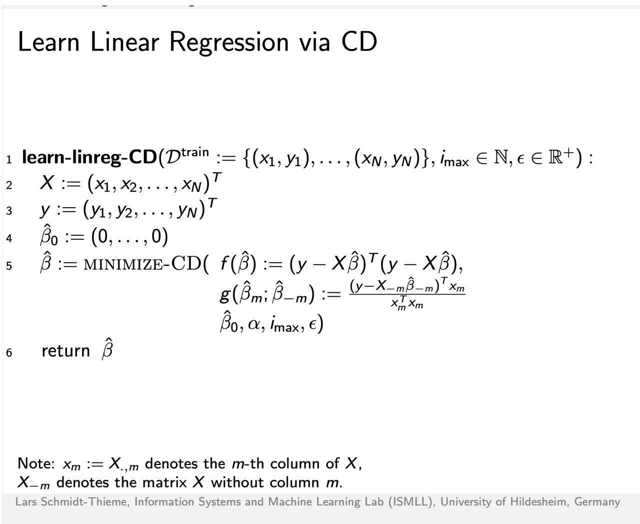

Figure 2: Linear Regression via CD

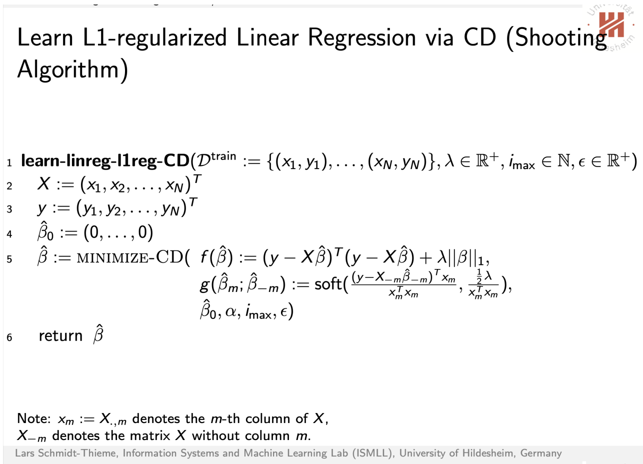

Figure 3: Learn L1-regularized Linear Regression via CD

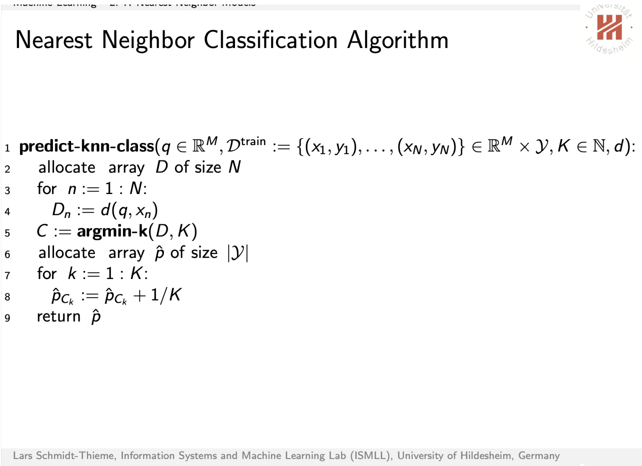

Figure 4: Nearest Neighbor Classification Algorithm