Linear classification and coordinate descent

1 Introduction

Let $\mathbf{X}$ be an $n \times m$, whose rows correspond to the records and the columns to the attributes. Let y be an $n \times 1$ vector of the known target values of the records in $\mathbf{X}$. For this assignment, we will consider binary classification. We will assign the label 1 to one class and -1 to the other. So the target values will either be 1 or -1. The goal is to learn a linear model, that will predict the label of a record. More specifically, the following minimization problem must be solved:

$$minimize_w(||\mathbf{X}w - y||^2)$$

where the squared norm is the loss that we try to minimize, and it is called the least squares problem.

2 Minimization

To minimize our loss, we can use **coordinate descent**. It starts with an initial random guess for w, and there is a set of k outer iterations. For the purpose of this project, we will initialize w as a vector of $0's$. In each outer iteration, it iterates m times in order to optimize the value of the objective function with respect to w_i variable only, while keeping all the rest fixed.

$$w_i^{new} = argmin_{w_i}(||\mathbf{X}w - y||^2), for \quad i = 1, 2, ...m$$

This optimization is performed by taking the partial derivative of our loss function with respect to w_i and setting it to 0. Then we solve that for w_i and that will become the updated value for w_i. More specifically, the updates are derived as follows:

$$\nabla_{w_i}(||\mathbf{X}w - y||^2) = 2\mathbf{X}_i^T(\mathbf{X}w - y) = 0$$
$$\Rightarrow \mathbf{X}_i^T(\mathbf{X}_i w_i + \mathbf{X}_{-i}w_{-i} - y) = 0$$
$$\Rightarrow w_i = \frac{\mathbf{X}_i^T(y - \mathbf{X}_{-i}w_{-i})}{\mathbf{X}_i^T\mathbf{X}_i}$$

The $\mathbf{X}_i$ is the ith column of $\mathbf{X}$, and w_i is the ith element of w. The subscript $-i$ indicates that the $n \times (m-1)$ matrix $\mathbf{X}_{-i}$ is the matrix $\mathbf{X}$ if we exclude column i, and correspondingly, the $(m-1) \times 1$ vector w_{-i} is the w without the ith element.

For the purpose of this assignment, we will run a fixed k number of outer iterations, which will be given as a command line argument. Your program should also output the model loss ($||\mathbf{X}w - y||^2$) to the standard output after each outer iteration.