Homework 4 for Machine learning: Theory and Algorithms

Gabriele Genovese

11 October 2024

1 Exercise 1

1.1 Point 1

The first term of $\overline{L}_S(x)$ is the squared loss function and we already know it's convex. The second term is the regularization term is also a quadratic function and quadratic function are convex. The sum of the two terms will result also in a quadratic function. So, the minimization of $\overline{L}_S(x)$ is a convex optimization problem.

1.2 Point 2

To show this, let's compute the gradient of $\overline{L}_S(x)$:

$$\nabla \overline{L}_S(x) = \frac{2}{m} \sum_{i=1}^{m} (w^T x_i - y_i) x_i + 2\lambda w$$

We need to manage $\nabla \overline{L}_S(x) = 0$ to be able to see if it's a linear system $Aw = b$.

$$\frac{2}{m} \sum_{i=1}^{m} (w^T x_i - y_i) x_i + 2\lambda w = 0$$

$$\sum_{i=1}^{m} w x_i x_i^T + \lambda w = \sum_{i=1}^{m} y_i x_i$$

$$\left(\sum_{i=1}^{m} x_i x_i^T + \lambda I \right) w = \sum_{i=1}^{m} y_i x_i$$

A is equal to $\sum_{i=1}^{m} x_i x_i^T + \lambda I$ and b is equal to $\sum_{i=1}^{m} y_i x_i$.

1.3 Point 3

If A is invertible then the solution is unique and it's $w = A^{-1} b$. Since the first term of A is a positive semidefinite matrix and the second one is a strictly positive diagonal matrix, A is invertible, then it exist a unique solution to the minimization problem.

2 Exercise 2

2.1 Point 1

The empirical risk is:

$$L_S(w) = \frac{1}{m} \sum_{i=1}^{m} \mid w^T x_i - y_i \mid$$

The absolute value function is convex and since the sum of convex functions is also convex, the empirical risk $L_S(w)$ is convex in w. Therefore, the empirical risk minimization problem is a convex optimization problem.

2.2 Point 2

The loss is defined as $\ell(w, (x_i, y_i)) = |wx_i - y_i|$ and is always differentiable except for $wx_i = y_i$. The empirical loss over the dataset S will not be differentiable if $wx_i = y_i$ for any i in the sum.

2.3 Point 3

To find w^*, we need to ensure that:

$$\frac{dL_S(w)}{dw} \leq 0 \text{ for } w \leq w^*$$

$$\frac{dL_S(w)}{dw} \geq 0 \text{ for } w \geq w^*$$

Since $L_S(w)$ is a convex function and absolute values are involved, the optimal solution w^* will be located where the sum of the absolute deviations is minimized. This is typically the median of the values $\frac{y_i}{x_i}$, assuming $x_i \neq 0$. So, the algorithm proposed is the following:

1. For each i where $x_i \neq 0$, calculate $\frac{y_i}{x_i}$.
2. Identify the median of the set $\left\{\frac{y_i}{x_i}\right\}$.
3. Set w^* to this median. In convex problems with absolute value terms, the median minimizes the overall deviation.

This approach is effective because in convex optimization problems involving absolute values, the median provides the best fit by minimizing the sum of the absolute differences.

2.4 Point 4

The most computationally expensive part of the proposed algorithm is sorting the set $\left\{\frac{y_i}{x_i}\right\}$. Sorting this set requires $O(m \log m)$ time, where m is the number of data points.