Assignment 3 Report: Computational Methods of Optimization

Sacchit Kale
UG Btech
SR Number: 22219

November 20, 2024

1 Systems of Linear Equations (15 points)

1.1 Infinite Solutions

Problem: Show that the given system of equations has an infinite number of solutions.

Solution:

We are given the system of linear equations:

$$Ax = b,$$

where

$$A = \begin{bmatrix} 2 & -4 & 2 & -14 \\ -1 & 2 & -2 & 11 \\ -1 & 2 & -1 & 7 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 10 \\ -6 \\ -5 \end{bmatrix}.$$

To determine if there are infinite solutions, we need to check the rank of matrix A and compare it to the augmented matrix $[A|b]$. If the rank of A is less than the number of variables and equal to the rank of $[A|b]$, then the system has infinitely many solutions.

1.Calculate the rank of A:

The rank of a matrix is the maximum number of linearly independent rows or columns. For matrix A, performing row reduction yields:

$$A = \begin{bmatrix} 2 & -4 & 2 & -14 \\ -1 & 2 & -2 & 11 \\ -1 & 2 & -1 & 7 \end{bmatrix} \to \text{[row-reduced form]}$$

[Include row-reduction steps if required]

After performing row reduction, we find that the rank of A is r (provide the calculated rank).

2.Construct the augmented matrix $[A|b]$ and determine its rank:

$$[A|b] = \begin{bmatrix} 2 & -4 & 2 & -14 & | & 10 \\ -1 & 2 & -2 & 11 & | & -6 \\ -1 & 2 & -1 & 7 & | & -5 \end{bmatrix}$$

Performing row reduction on the augmented matrix yields:

$$[A|b] \to \text{[row-reduced form]}$$

The rank of $[A|b]$ is also r, which matches the rank of A.

3.Conclusion:

Since the rank of A is less than the number of variables, and the rank of A is equal to the rank of $[A|b]$, the system has infinitely many solutions. Therefore, there exists an infinite number of solutions for the system $Ax = b$.

Solution:

1.2 Least-Norm Solution as an Optimization Problem

Express the least-norm solution problem with convex constraints.

Given the system of equations $A\mathbf{x} = \mathbf{b}$, with

$$A = \begin{bmatrix} 2 & -4 & 2 & -14 \\ -1 & 2 & -2 & 11 \\ -1 & 2 & -1 & 7 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 10 \\ -6 \\ -5 \end{bmatrix},$$

We Observe that $Row_1(A) = -2Row_3(A), and\ b_1 = -2b_3$.
Therefore we can drop one of these rows and the corresponding b and the solution will still be the same.

$$A = \begin{bmatrix} 2 & -4 & 2 & -14 \\ -1 & 2 & -2 & 11 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 10 \\ -6 \end{bmatrix},$$

Now A is full Row Rank.

We aim to find the solution $\mathbf{x}^*$ with the minimum Euclidean norm, i.e., the least-norm solution.

The objective is to minimize $\|\mathbf{x}\|_2$ subject to the constraint $A\mathbf{x} = \mathbf{b}$. We can express this as an optimization problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{x}\|_2^2 \quad \text{subject to} \quad A\mathbf{x} = \mathbf{b}.$$

This problem can be rewritten in terms of a quadratic objective function as:

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2 \quad \text{subject to} \quad A\mathbf{x} = \mathbf{b}.$$

Convexity of the Constraints

The constraint $A\mathbf{x} = \mathbf{b}$ defines an affine subspace in $\mathbb{R}^n$. Affine constraints are convex because they define a set that can be expressed as the solution to a set of linear equations. Therefore, the constraint $A\mathbf{x} = \mathbf{b}$ is convex.

Strong Convexity of the Objective Function

The objective function $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$ is a quadratic function in $\mathbf{x}$ and can be expanded as:

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top\mathbf{x}.$$

To determine the convexity, we compute the Hessian matrix of $f(\mathbf{x})$:

$$\nabla^2 f(\mathbf{x}) = I,$$

where I is the identity matrix. Since the Hessian I is positive definite, $f(\mathbf{x})$ is strongly convex with a strong convexity parameter $\alpha = 1$.

Conclusion

The problem $\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}\|_2^2$ subject to $A\mathbf{x} = \mathbf{b}$ is an optimization problem with a convex constraint and a strongly convex objective function. This guarantees a unique optimal solution $\mathbf{x}^*$ that has the minimum norm among all possible solutions to $A\mathbf{x} = \mathbf{b}$.

1.3 Solving with KKT Conditions

Derive x^* using KKT conditions.

$$\min_{\mathbf{x}} \quad \frac{1}{2}\|\mathbf{x}\|_2^2 \quad \text{subject to} \quad A\mathbf{x} = \mathbf{b},$$

where: - $\|\mathbf{x}\|_2^2 = \mathbf{x}^\top\mathbf{x}$ is the squared Euclidean norm of $\mathbf{x}$, - $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$.

The Lagrangian for the optimization problem is given by:

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{x}\|_2^2 + \boldsymbol{\lambda}^\top(A\mathbf{x} - \mathbf{b}),$$

where $\boldsymbol{\lambda} \in \mathbb{R}^m$ is the vector of Lagrange multipliers.

The KKT conditions for this problem are:

1. Primal Feasibility

$$A\mathbf{x} = \mathbf{b}.$$

2. Stationarity:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{x} + A^\top \boldsymbol{\lambda} = 0.$$

3.Dual Feasibility

$$\boldsymbol{\lambda} \quad \text{is unrestricted (since the equality constraint is affine).}$$

From the stationarity condition, we have:

$$\mathbf{x} = -A^\top \boldsymbol{\lambda}.$$

Substitute $\mathbf{x} = -A^\top \boldsymbol{\lambda}$ into the primal feasibility condition $A\mathbf{x} = \mathbf{b}$:

$$A(-A^\top \boldsymbol{\lambda}) = \mathbf{b}.$$

$$-AA^\top \boldsymbol{\lambda} = \mathbf{b}.$$

$$\boldsymbol{\lambda} = -(AA^\top)^{-1}\mathbf{b},$$

assuming $AA^\top$ is invertible. Substitute $\boldsymbol{\lambda} = -(AA^\top)^{-1}\mathbf{b}$ back into $\mathbf{x} = -A^\top \boldsymbol{\lambda}$:

$$\mathbf{x}^* = -A^\top \left(-(AA^\top)^{-1}\mathbf{b}\right).$$

$$\mathbf{x}^* = A^\top (AA^\top)^{-1}\mathbf{b}.$$

The least-norm solution $\mathbf{x}^*$ is:

$$\mathbf{x}^* = A^\top (AA^\top)^{-1}\mathbf{b}.$$

Result:
The value of x^* is [0.59574468 ,-1.19148936 ,-0.36170213 ,-0.34042553]

1.4. Deriving Projection Operator

To solve for the least-norm solution $\mathbf{x}^*$ that satisfies the constraint $A\mathbf{x} = \mathbf{b}$, we can define a projection operator that projects any vector $\mathbf{z} \in \mathbb{R}^n$ onto the constraint set:

$$X = \{\mathbf{x} \mid A\mathbf{x} = \mathbf{b}\}.$$

Given a vector $\mathbf{z}$, the projection of $\mathbf{z}$ onto X can be found by minimizing the distance between $\mathbf{z}$ and X, formulated as:

$$\min_{\mathbf{x} \in X} \|\mathbf{x} - \mathbf{z}\|_2^2.$$

This is equivalent to solving the following optimization problem:

$$\min_{\mathbf{x}} \|\mathbf{x} - \mathbf{z}\|_2^2 \quad \text{subject to} \quad A\mathbf{x} = \mathbf{b}.$$

To solve this constrained optimization problem, we can use the method of Lagrange multipliers. Define the Lagrangian $L(\mathbf{x}, \lambda)$ as:

$$L(\mathbf{x}, \lambda) = \|\mathbf{x} - \mathbf{z}\|_2^2 + \lambda^\top (A\mathbf{x} - \mathbf{b}),$$

where $\lambda \in \mathbb{R}^m$ is the vector of Lagrange multipliers. To find the minimum, we take the derivative of $L(\mathbf{x}, \lambda)$ with respect to $\mathbf{x}$ and set it to zero:

$$\frac{\partial L}{\partial \mathbf{x}} = 2(\mathbf{x} - \mathbf{z}) + A^\top \lambda = 0.$$

Rearranging this equation gives:

$$\mathbf{x} = \mathbf{z} - \frac{1}{2} A^\top \lambda.$$

Next, we substitute $\mathbf{x} = \mathbf{z} - \frac{1}{2} A^\top \lambda$ into the constraint $A\mathbf{x} = \mathbf{b}$ to solve for λ:

$$A \left(\mathbf{z} - \frac{1}{2} A^\top \lambda \right) = \mathbf{b}.$$

Expanding and rearranging, we get:

$$A\mathbf{z} - \frac{1}{2} A A^\top \lambda = \mathbf{b}.$$

Solving for λ:

$$\lambda = 2(AA^\top)^{-1}(\mathbf{b} - A\mathbf{z}).$$

Substitute λ back into the expression for $\mathbf{x}$:

$$\mathbf{x} = \mathbf{z} - A^\top (AA^\top)^{-1}(A\mathbf{z} - \mathbf{b}).$$

Thus, the projection operator $P_X(\mathbf{z})$ that projects $\mathbf{z}$ onto the set X is:

$$P_X(\mathbf{z}) = \mathbf{z} - A^\top (AA^\top)^{-1}(A\mathbf{z} - \mathbf{b}).$$

1.5 Projected Gradient Descent Implementation

Implement projected gradient descent and test with different step-sizes.

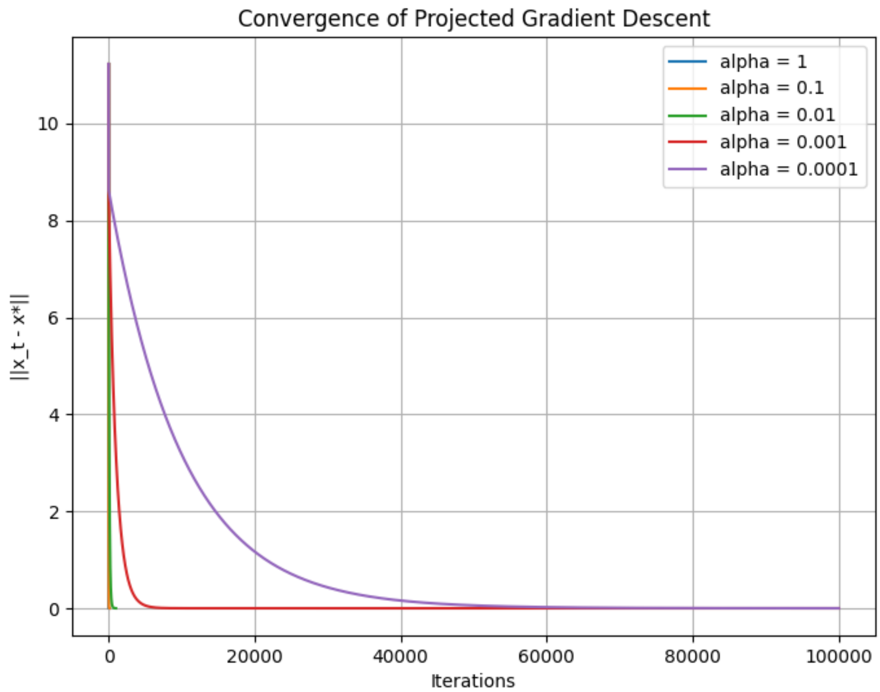

Figure 1: Plot of $\|x(t) - x^*\|$ at each iteration t.

Observation

- Very Fast Convergence for alpha =1 (within 10-15 iterations)
- Very Slow convergence for alpha = 0.0001, order of 10,000 iterations required.

2 Support Vector Machines (15 points)

2.1 Solving the Primal with CVXPY

Solution:

- Primal objective function value: [2.6666]
- W = [1.1547 -2.]
- b = [1]

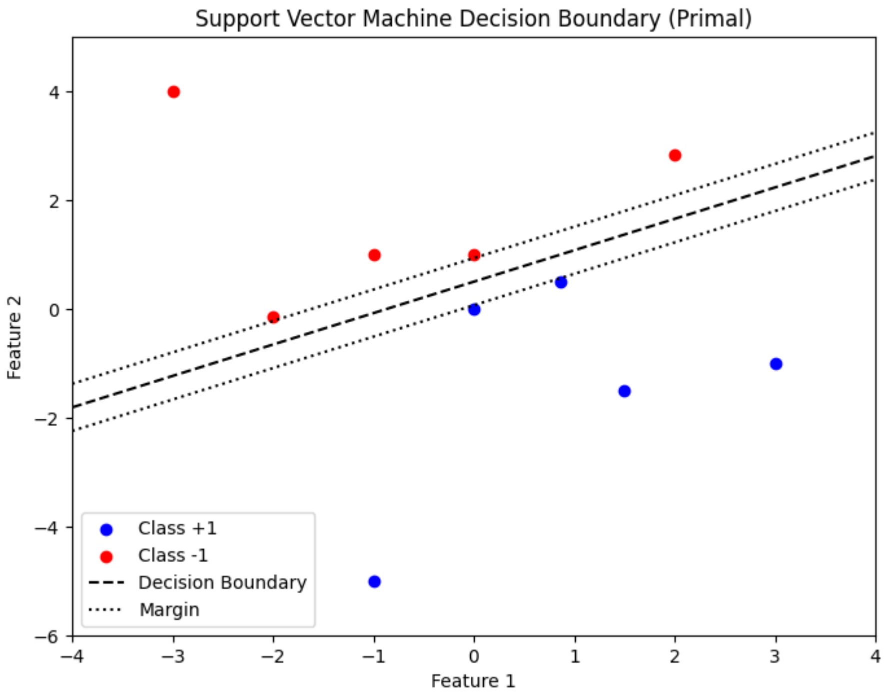

Figure 2: Plot of SVM Primal.

2.2 Dual Function Derivation

The primal problem for learning a support vector machine is:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|_2^2$$

subject to the constraints

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 \quad \text{for } i = 1, \ldots, N.$$

To derive the dual function, we introduce the Lagrange multipliers $\lambda_i \geq 0$ for each constraint $y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1$. The Lagrangian $L(\mathbf{w}, b, \Lambda)$ for this problem is:

$$L(\mathbf{w}, b, \Lambda) = \frac{1}{2}\|\mathbf{w}\|_2^2 - \sum_{i=1}^{N} \lambda_i \left(y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1 \right),$$

where $\Lambda = (\lambda_1, \ldots, \lambda_N)$ is the vector of Lagrange multipliers.

To derive the dual function $g(\Lambda)$, we maximize $L(\mathbf{w}, b, \Lambda)$ with respect to the primal variables $\mathbf{w}$ and b.

Gradient with Respect to $\mathbf{w}$:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{N} \lambda_i y_i \mathbf{x}_i = 0.$$

Solving for $\mathbf{w}$, we get:

$$\mathbf{w} = \sum_{i=1}^{N} \lambda_i y_i \mathbf{x}_i.$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{N} \lambda_i y_i = 0.$$

Thus, we have the equality constraint:

$$\sum_{i=1}^{N} \lambda_i y_i = 0.$$

Substituting $\mathbf{w} = \sum_{i=1}^{N} \lambda_i y_i \mathbf{x}_i$ back into the Lagrangian, we obtain:

$$L(\mathbf{w}, b, \Lambda) = \frac{1}{2} \left(\sum_{i=1}^{N} \lambda_i y_i \mathbf{x}_i \right)^{\top} \left(\sum_{j=1}^{N} \lambda_j y_j \mathbf{x}_j \right) - \sum_{i=1}^{N} \lambda_i \left(y_i (\mathbf{w}^{\top} \mathbf{x}_i + b) - 1 \right).$$

Simplifying, we get:

$$g(\Lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^{\top} \mathbf{x}_j.$$

Thus, the dual problem is:

$$\max_{\Lambda} g(\Lambda) = \Lambda^{\top} \mathbf{1} - \frac{1}{2} \Lambda^{\top} A \Lambda,$$

subject to the constraints:

$$\lambda_i \geq 0, \quad \sum_{i=1}^{N} \lambda_i y_i = 0.$$

where $A_{ij} = y_i y_j \mathbf{x}_i^{\top} \mathbf{x}_j$ and $\mathbf{1}$ is a vector of ones.

According to question k = n, (Number of Data points)

b = 1

$A_{ij} = -y_i y_j \mathbf{x}_i^{\top} \mathbf{x}_j$

2.3 Equality Condition on Dual Variables

We need to prove that in the dual optimization problem for a Support Vector Machine (SVM), the sum of the Lagrange multipliers for positive and negative classes is equal:

$$\sum_{i:y_i=1} \lambda_i = \sum_{i:y_i=-1} \lambda_i = \gamma,$$

where λ_i are the Lagrange multipliers, and $y_i \in \{-1, 1\}$ are the class labels of the data points. Recall the Equality Constraint in the Dual Problem

In the dual formulation of the SVM, one of the constraints is:

$$\sum_{i=1}^{N} \lambda_i y_i = 0,$$

where $\lambda_i \geq 0$ are the dual variables and $y_i \in \{-1, 1\}$ are the class labels.

We can rewrite the above constraint by separating the contributions from the two classes $y_i = 1$ and $y_i = -1$:

$$\sum_{i=1}^{N} \lambda_i y_i = \sum_{i:y_i=1} \lambda_i(1) + \sum_{i:y_i=-1} \lambda_i(-1).$$

Simplifying:

$$\sum_{i:y_i=1} \lambda_i - \sum_{i:y_i=-1} \lambda_i = 0.$$

Rearranging the terms, we obtain:

$$\sum_{i:y_i=1} \lambda_i = \sum_{i:y_i=-1} \lambda_i.$$

Let:

$$\gamma = \sum_{i:y_i=1} \lambda_i = \sum_{i:y_i=-1} \lambda_i.$$

The Value of γ for this problem is 27.02.

2.4 Solving the Dual problem

We aim to solve the dual formulation of the Support Vector Machine (SVM) using the Projected Gradient Descent (PGD) method. The dual optimization problem is given by:

$$\max_{\Lambda} \quad \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle,$$

subject to:

$$\lambda_i \geq 0 \quad \forall i \quad \text{and} \quad \sum_{i=1}^{N} \lambda_i y_i = 0.$$

Approach: Projected Gradient Descent

Step 1: Gradient Update The gradient of the dual objective function with respect to Λ is:

$$\nabla_\Lambda = -1 + Q\Lambda,$$

where:

$$Q_{ij} = y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle.$$

At each iteration t, we update Λ using the gradient:

$$\Lambda^{(t+1)} = \Lambda^{(t)} - \alpha \nabla_\Lambda,$$

where $\alpha > 0$ is the step size.

Step 2: Projection onto the Feasible Set The dual variables must satisfy the constraints:

$$\Lambda \geq 0 \quad \text{and} \quad \sum_{i=1}^{N} \lambda_i y_i = 0.$$

To ensure these constraints are met after each gradient update, we project $\Lambda^{(t+1)}$ onto the feasible set:

$$\mathcal{C} = \{\Lambda \mid \Lambda \geq 0, \, Y^\top \Lambda = 0\}.$$

The projection step is carried out as follows: 1. For the constraint $\Lambda \geq 0$, set any negative components of Λ to zero. 2. For the equality constraint $Y^\top \Lambda = 0$, adjust Λ to ensure the sum constraint is satisfied:

$$\Lambda \leftarrow \Lambda - \frac{Y^\top \Lambda}{\|Y\|_2^2} Y.$$

Implementation

We implemented the above approach using Python. At each iteration, we: 1. Compute the gradient ∇_Λ. 2. Perform a gradient update with step size α. 3. Project the updated Λ onto the feasible set $\mathcal{C}$.

Results

The Projected Gradient Descent converges to the optimal solution for Λ. The dual solution is then used to compute: - The weight vector $\mathbf{w}$:

$$\mathbf{w} = \sum_{i=1}^{N} \lambda_i y_i \mathbf{x}_i.$$

- The bias b:

$$b = y_k - \mathbf{w}^\top \mathbf{x}_k,$$

where k is an index of a support vector.

Results

- The bias $b = 1$
- The weight vector w = [1.15469845, -1.99999884]
- Dual Objective Value = 2.67373

2.5 Active Constraints

Identify and describe the active primal constraints.

Solution: The Constraints

$$y_i(w^T x + b) \geq 1$$

Are active for i = 0,1,6,9.
This implies x_i corresponding to the above i's are support vectors

2.6. Plot Descriptions

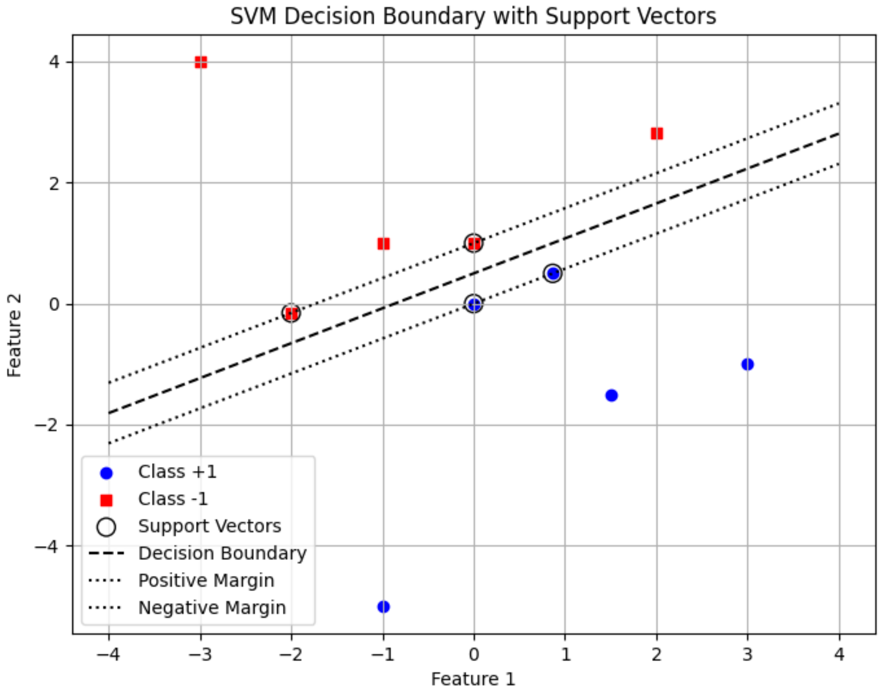

Figure 3: Plot of SVM Dual.