भारतीय विज्ञान संस्थान

# Assignment 3

# E0 230

## Computationl Methods of Optimization

**Professor Chiranjib Bhattacharyya**

**Submitted by:**

Pratyush Kant

# 1 Systems of Linear Equations (15 points)

Let us revisit the case where solutions to a linear system of equations, $\mathbf{Ax} = \mathbf{b}$, need to be found. In the last assignment, this was cast as the convex optimisation problem given by:

$$\min_{\mathbf{x}\in\mathbb{R}^n} ||\mathbf{Ax} - \mathbf{b}||^2.$$

While this method has its own merits (such as finding a point which is "closest" to being a solution if the system is indeterminate, it does not find solutions with a particular property. In this assignment, we will assume that there exists at least one solution, and try to find the particular solution $\mathbf{x}^*$ that has the least norm. For the purpose of this question, let:

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

1. Show that the given system of equations has an infinite number of solutions.
2. Express the problem of finding the least-norm solution as an optimisation problem (ConvProb) with convex constraints and a strongly convex objective function. Show that the constraints and the objective are convex and strongly convex, respectively.
3. Use the KKT conditions to solve ConvProb, and arrive at an expression for $\mathbf{x}^*$ and show the intermediate steps. Write code to evaluate the expression and report $\mathbf{x}^*$.
4. Derive a projection operator for the constraint set of ConvProb.
5. Use the derived projection operator and implement projected gradient descent to solve ConvProb. Test with different step-sizes and plot $||\mathbf{x}^{(t)} - \mathbf{x}^*||$ at each iteration $t$.

# Solution

We claim that any system of equation has (a) no solution, (b) a unique solution, or (c) infinitely many solutions. Suppose $\mathbf{Ax} = \mathbf{b}$ has a finite number of solutions say $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$. Then, the set of solutions is given by $\mathbf{x} = \mathbf{x}_1 + \sum_{i=1}^{k} c_i(\mathbf{x}_i - \mathbf{x}_1)$ where $c_i \in \mathbb{R}$. This set of solutions is infinite. Hence, the given system of equations has an infinite number of solutions.

1. The optimization problem can be written as:

$$\begin{aligned} ||\mathbf{Ax} - \mathbf{b}||^2 &= (\mathbf{Ax} - \mathbf{b})^T(\mathbf{Ax} - \mathbf{b}) \\ &= \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} - \mathbf{b}^T\mathbf{Ax} + \mathbf{b}^T\mathbf{b} \\ &= \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} - 2\mathbf{b}^T\mathbf{Ax} + \mathbf{b}^T\mathbf{b} \end{aligned}$$

In the script the determinant of $\mathbf{A}^\top\mathbf{A}$ has bee reported which is zero. Hence, $\mathbf{A}^\top\mathbf{A}$ cannot be a positive definite matrix. Therefore, the minimal value of the function is $-\infty$ and the system has an infinite number of solutions. (If the system has a unique solution, the function would have a minimum value, namely 0 at the unique solution.)

2. The ConvProb can be written as:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}||\mathbf{x}||^2 \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{b}$$

The objective function is strongly convex since the Hessian of the objective function is the identity matrix which is positive definite. The constraints are convex since the Hessian of the function function $\mathbf{Ax} - \mathbf{b}$ is the zero matrix which is positive semi-definite. Hence, the constraints and the objective are convex and strongly convex, respectively.

3. The Lagrangian for the problem is given by:

$$\mathcal{L}(\mathbf{x}, \mathbf{\Lambda}) = \frac{1}{2}||\mathbf{x}||^2 + \mathbf{\Lambda}^T(\mathbf{Ax} - \mathbf{b})$$

The KKT conditions are given by:

(a) Stationarity: $\nabla_x \mathcal{L} = \mathbf{x} + \mathbf{A}^T\mathbf{\Lambda} = 0$

(b) Primal feasibility: $\mathbf{Ax} = \mathbf{b}$

From the stationarity condition, we have $\mathbf{x} = -\mathbf{A}^T\mathbf{\Lambda}$. Substituting this in the primal feasibility condition, we get $\mathbf{AA}^T\mathbf{\Lambda} = -\mathbf{b}$. The matrix $\mathbf{AA}^T$ is non-invertible since the determinant of $\mathbf{AA}^\top$ is also reported as zero in the script. However as instructed we will use np.linalg.pinv() to get $(\mathbf{AA}^\top)^{-1}$. Using this we compute $\mathbf{\Lambda} = -(\mathbf{AA}^\top)^{-1}\mathbf{b}$ and $\mathbf{x}^* = -\mathbf{A}^T\mathbf{\Lambda}$. The reported values are:

$$\mathbf{\Lambda} = \begin{pmatrix} -0.6212766 \\ -0.95744681 \\ 0.3106383 \end{pmatrix}$$

$$\mathbf{x}^* = \begin{pmatrix} 0.59574468 \\ -1.19148936 \\ -0.36170213 \\ -0.34042553 \end{pmatrix}$$

The value of the primal objective function at $\mathbf{x}^*$ is 1.0106382978723247 and the value of $||\mathbf{Ax}^* - \mathbf{b}||^2$ is $3.62560472039577e - 27$.

4. Let $\mathbf{z} \in \mathbb{R}^n$. The projection operator is given by solving the following optimization problem:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}||\mathbf{x} - \mathbf{z}||^2 \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{b}$$

The Lagrangian for the problem is given by:

$$\mathcal{L}(\mathbf{x}, \mathbf{\Lambda}) = \frac{1}{2}||\mathbf{x} - \mathbf{z}||^2 + \mathbf{\Lambda}^T(\mathbf{Ax} - \mathbf{b})$$

From the KKT conditions, we have:

(a) Stationarity: $\nabla_x \mathcal{L} = \mathbf{x} - \mathbf{z} + \mathbf{A}^T\mathbf{\Lambda} = 0$

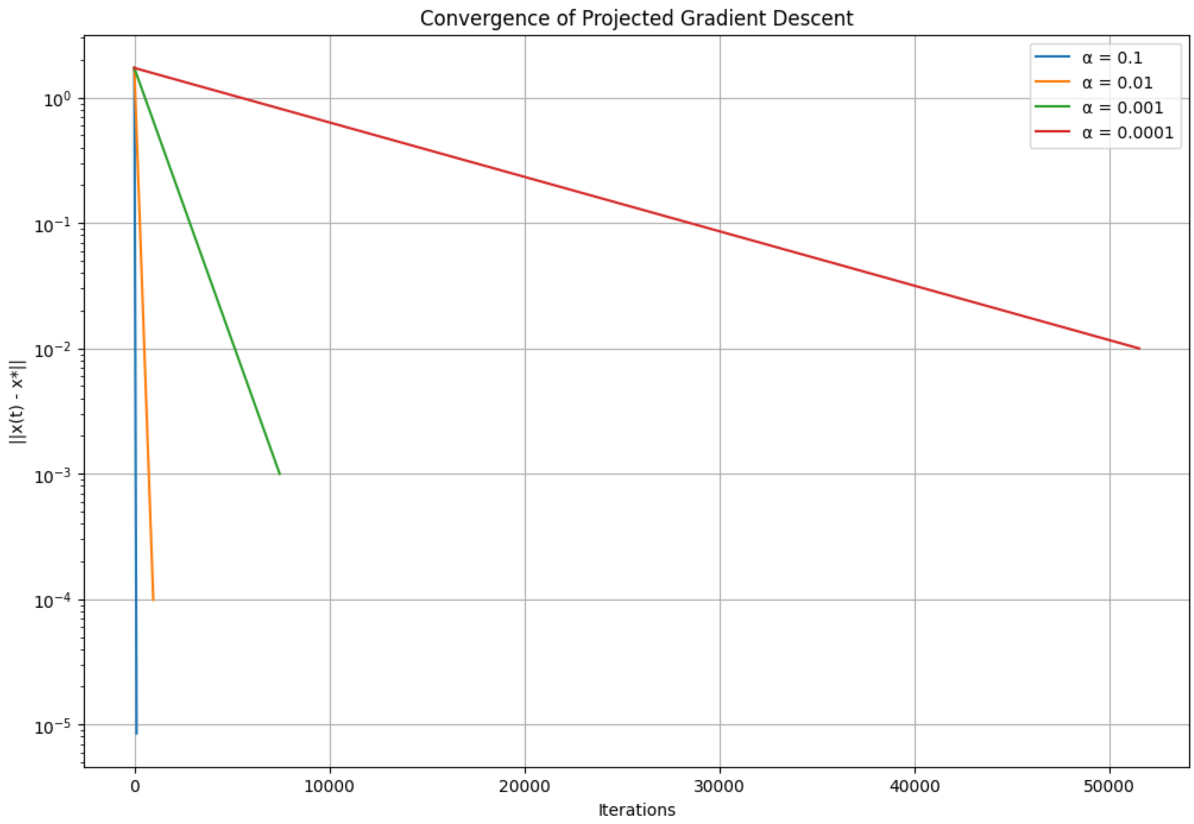



Figure 1: Projected Gradient Descent

(b) Primal feasibility: $\mathbf{Ax} = \mathbf{b}$

From the first condition, we have $\mathbf{x} = \mathbf{z} - \mathbf{A}^T\mathbf{\Lambda}$. Substituting this in the primal feasibility condition, we get $\mathbf{A}(\mathbf{z} - \mathbf{A}^T\mathbf{\Lambda}) = \mathbf{b}$. Again using the pseudo-inverse, we get $\mathbf{\Lambda} = (\mathbf{AA}^T)^{-1}(\mathbf{Az} - \mathbf{b})$. The projection operator is then given by:

$$P_{\text{Ax=b}}(\mathbf{z}) = \mathbf{z} - \mathbf{A}^T(\mathbf{AA}^T)^{-1}(\mathbf{Az} - \mathbf{b})$$

5. The scripts projection_operator and projected_gradient_descent do the job. Let us find a $\mathbf{x}^{(0)}$ to use in the script such that $\mathbf{Ax}^{(0)} = \mathbf{b}$.

$$\begin{aligned} 2x_1^0 - 4x_2^0 + 2x_3^0 - 14x_4^0 &= 10 \\ -x_1^0 + 2x_2^0 - 2x_3^0 + 11x_4^0 &= -6 \\ -x_1^0 + 2x_2^0 - x_3^0 + 7x_4^0 &= -5 \end{aligned}$$

Since the first and the third equations are the same, we can ignore the first equation. Take $x_1^0 = x_4^0 = 0$ to get the system of equations:

$$\begin{aligned} 2x_2^0 - 2x_3^0 &= -6 \\ 2x_2^0 - x_3^0 &= -5 \end{aligned}$$

Solving, we get $x_2^0 = -2$ and $x_3^0 = 1$. Hence, $\mathbf{x}^{(0)} = \begin{pmatrix} 0 & -2 & 1 & 0 \end{pmatrix}^T$. The plot for different step sizes is given in the figure 1.

# 2 Support Vector Machines (15 points)

Support vector machines (SVMs) are among the most widely used techniques for classifying data, and are very well studied. The SVM is a linear classifier; that is, we aim to find a function $y = f(\mathbf{x}) = \text{sign}(\mathbf{w}^T\mathbf{x} + b)$. We are given data of the form $\{\mathbf{x}_i, y_i\}_N$, where the pair $(\mathbf{x}_i, y_i) \in \mathbb{R}^n \times \{-1, 1\}$. To learn a support vector machine, we need to solve the following convex optimization problem.

$$\mathbf{w}^*, b^* = \arg\min_{\mathrm{w}, b} \frac{1}{2}\|\mathbf{w}\|^2$$

subject to

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

1. Use either CVXPY (python) to solve the primal for the data given in "Data.csv" and "Labels.csv", which can be found under the Files tab. What is the value of the primal objective function?
2. Show that the dual function is of the form
$$g(\mathbf{\Lambda}) = \mathbf{\Lambda}^T\mathbf{b} + \frac{1}{2}\mathbf{\Lambda}^T\mathbf{A}\mathbf{\Lambda},$$
$\mathbf{\Lambda} = (\lambda_1, \ldots, \lambda_k)$. What are $\mathbf{A}_{ij}$ and $\mathbf{b}_i$? What is $k$?
3. Show that
$$\sum_{i:y_i=1} \lambda_i = \sum_{i:y_i=-1} \lambda_i = \gamma$$
What is the value of $\gamma$ for the given problem?
4. Write a program to solve the dual problem. What is the value of the dual objective at optimality?
5. Which of the primal constraints are active? Describe your answer.
6. Plot the following:
   (a) The line $\mathbf{w}^T\mathbf{x} + b = 0$.
   (b) The data points provided. For $y = 1$, mark the points with circles, and for $y = -1$, mark the points with squares.
   (c) The points corresponding to active constraints. Mark these points in red.

# Solution

1. Primal objective value has bee found as 2.666463 using CVXPY.

2. To derive the dual function of the given support vector machine (SVM) optimization problem, let's begin by writing the Lagrangian $L(\mathbf{w}, b, \mathbf{\Lambda})$ for the primal problem:

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

Here, $\mathbf{\Lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ are the dual variables associated with the inequality constraints.

To find the dual function $g(\mathbf{\Lambda})$, we first compute the gradients of $L$ with respect to $\mathbf{w}$ and $b$:

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

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

The dual function is then given by substituting these conditions in the Lagrangian.

$$\begin{aligned}
g(\mathbf{\Lambda}) &= \frac{1}{2}\left(\sum_{i=1}^{N} \lambda_i y_i \mathbf{x}_i\right)^\top \left(\sum_{j=1}^{N} \lambda_j y_j \mathbf{x}_j\right) - \sum_{i=1}^{N} \lambda_i [y_i \left(\left(\sum_{j=1}^{N} \lambda_j y_j \mathbf{x}_j\right)^\top \mathbf{x}_i + b\right) - 1] \\
&= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j - \sum_{i=1}^{N} \lambda_i y_i \left(\sum_{j=1}^{N} \lambda_j y_j \mathbf{x}_j^\top \mathbf{x}_i + b\right) + \sum_{i=1}^{N} \lambda_i \\
&= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j - b\sum_{i=1}^{N} \lambda_i y_i + \sum_{i=1}^{N} \lambda_i \\
&= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j + \sum_{i=1}^{N} \lambda_i
\end{aligned}$$

Hence the dual function is of the form $g(\mathbf{\Lambda}) = \mathbf{\Lambda}^\top \mathbf{b} + \frac{1}{2}\mathbf{\Lambda}^\top \mathbf{A}\mathbf{\Lambda}$, where $\mathbf{b}_i = 1$ and $\mathbf{A}_{ij} = -y_i y_j \mathbf{x}_i^\top \mathbf{x}_j$. The number of dual variables is $k = N$ and the dual optimization problem is given by:

$$\max_{\Lambda} \mathbf{\Lambda}^\top \mathbf{b} + \frac{1}{2}\mathbf{\Lambda}^\top \mathbf{A}\mathbf{\Lambda} \quad \text{subject to} \quad \mathbf{\Lambda} \geq 0, \quad \sum_{i=1}^{N} \lambda_i y_i = 0$$

3. Let $A = \{i : y_i = 1\}$ and $B = \{i : y_i = -1\}$. Since $A$ and $B$ are partitions of $\{1, 2, \ldots, N\}$, we have

$$\sum_{i=1}^{N} \lambda_i y_i = \sum_{i \in A} \lambda_i - \sum_{i \in B} \lambda_i$$

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

$$\sum_{i \in A} \lambda_i = \sum_{i \in B} \lambda_i = \gamma$$

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

The dual objective function is

$$\sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{N} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j$$

Since $\mathbf{w} = \sum_{i=1}^{N} \lambda_i y_i \mathbf{x}_i$, the dual objective function can be written as

$$\sum_{i=1}^{N} \lambda_i - \frac{1}{2} ||\mathbf{w}||^2$$

If strong duality holds, then the primal and dual objectives are equal. Hence,

$$\sum_{i=1}^{N} \lambda_i - \frac{1}{2} ||\mathbf{w}||^2 = \frac{1}{2} ||\mathbf{w}||^2 \implies \sum_{i=1}^{N} \lambda_i = ||\mathbf{w}||^2$$

Therefore, the value of $\gamma$ is $\frac{1}{2} \sum_{i=1}^{N} \lambda_i = \frac{1}{2} ||\mathbf{w}||^2$.

From the subquestion 1, the primal objective value is 2.666463. Hence, the value of $\gamma$ is 2.666463.

4. We will use projected gradient descent to solve the dual problem. Let $\mathbf{z}$ be a vector. The projection is then given by

$$\min \frac{1}{2} ||\mathbf{z} - \mathbf{\Lambda}||^2 \quad \text{subject to} \quad \sum_{i=1}^{N} \mathbf{y}^\top \mathbf{\Lambda} = 0$$

where $\mathbf{z} \in \mathbb{R}^N$, $\mathbf{y} \in \{-1, 1\}^N$, and $\mathbf{\Lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N)^\top$. Introduce a Lagrange multiplier $\mu$ for the constraint:

$$\mathcal{L}(\mathbf{\Lambda}, \mu) = \frac{1}{2} \|\mathbf{z} - \mathbf{\Lambda}\|^2 + \mu \mathbf{y}^\top \mathbf{\Lambda}.$$

$$\nabla_\Lambda \mathcal{L} = -(\mathbf{z} - \mathbf{\Lambda}) + \mu \mathbf{y} = 0 \implies \mathbf{\Lambda} = \mathbf{z} + \mu \mathbf{y}.$$

Substitute $\mathbf{\Lambda}$ back into the constraint:

$$\mathbf{y}^\top (\mathbf{z} + \mu \mathbf{y}) = 0 \implies \mathbf{y}^\top \mathbf{z} + \mu \mathbf{y}^\top \mathbf{y} = 0.$$

Since $\mathbf{y} \in \{-1, 1\}^N$, we have $\mathbf{y}^\top \mathbf{y} = N$. Therefore, $\mu = -\frac{\mathrm{y}^\top \mathrm{z}}{N}$. The projection is then given by:

$$\lambda_i = z_i - \frac{1}{N}\left(\sum_{j=1}^{N} y_j z_j\right) y_i.$$

To account for the non-negativity constraint, we set $\lambda_i = \max(0, \lambda_i)$. This has been implemented in the python script as solve_dual_svm. The value of the dual objective at optimality is 2.673735 and $\sum \lambda_i y_i = 7.073640e - 03$. In the script we get the reported values:

$$\sum_{i:y_i=1} \lambda_i = 2.673022$$
$$\sum_{i:y_i=-1} \lambda_i = 2.665949$$

5. The function solve_dual_svm returns the indices of the active constraints as well:

| **Index** | $\lambda_i$ | $y_i$ |
|---|---|---|
| 0 | 1.823139 | +1 |
| 1 | 0.849884 | +1 |
| 6 | 2.456700 | -1 |
| 9 | 0.209249 | -1 |

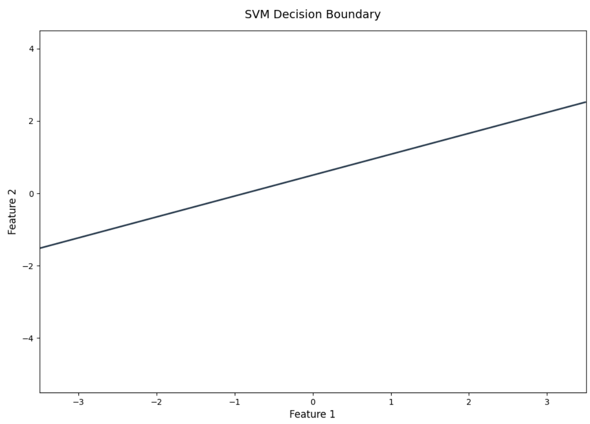
The script works by checking the indices where the value of $\lambda_i$ is more than tolerance which is $1e - 9$ in our case. These are reported as the active constraints.

6. The plots has been shown in the figure 2.

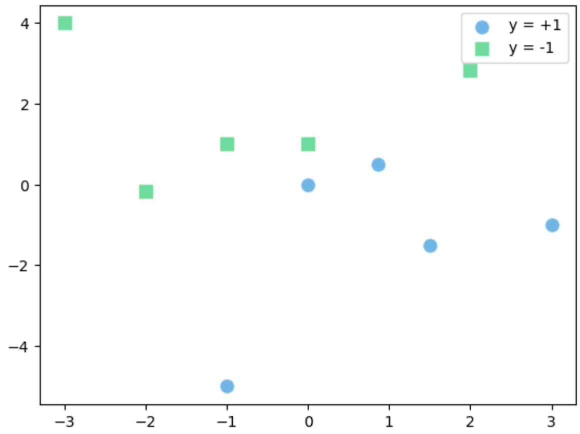
The complete picture of the SVM is provided by the figure 3.

The value of $\mathbf{w}^*$ and $b^*$ are:

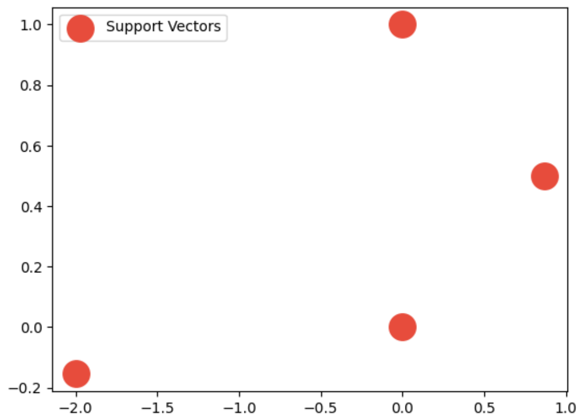
$$\mathbf{w}^* = (1.15452345 \ \ -1.99938719)^T$$
$$b^* = 0.9997426357897754$$



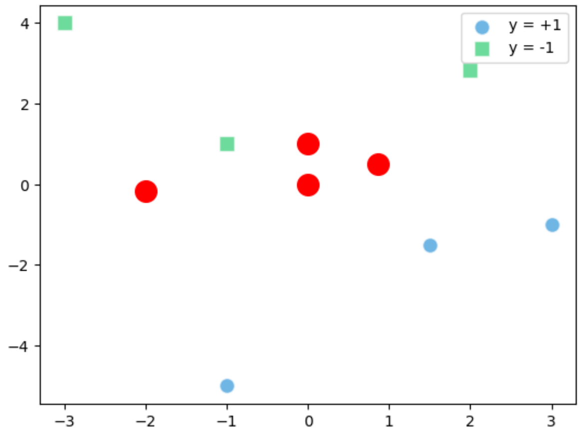

(a) Line $\mathbf{w}^\top \mathbf{x} + b = 0$




(b) Data Points




(c) Active constraints




(d) Active constraints with Data Points

Figure 2: Support Vector Machine Analysis

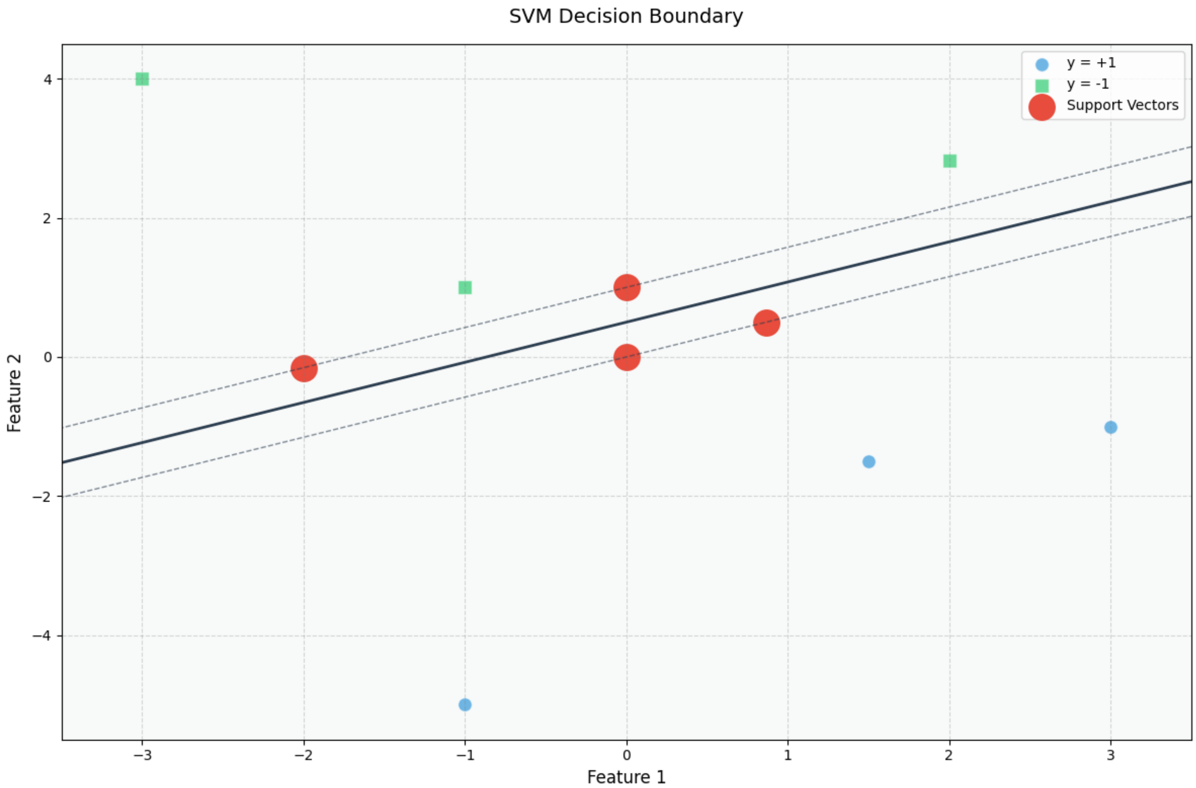



Figure 3: Support Vector Machine