Name:	Aneesh Panchal	**Assignment No:**	Assignment 3
SR No:	06-18-01-10-12-24-1-25223	**Course Code:**	E0230
Email ID:	aneeshp@iisc.ac.in	**Course Name:**	Computational Methods of Optimization
Date:	November 19, 2024	**Term:**	AUG 2024

Systems of Linear Equations

Solution 1

Given matrices are as follows,

$$A = \begin{bmatrix} 2 & -4 & 2 & -14 \\ -1 & 2 & -2 & 11 \\ -1 & 2 & -1 & 7 \end{bmatrix}_{3\times 4} \qquad \text{and,} \quad b = \begin{bmatrix} 10 \\ -6 \\ -5 \end{bmatrix}_{3\times 1}$$

Hence, the solution to system $Ax = b$ will be of dimension 4×1 i.e. Number of variables, $n = 4$.

$$Rank(A) = Rank\left(\begin{bmatrix} 2 & -4 & 2 & -14 \\ -1 & 2 & -2 & 11 \\ -1 & 2 & -1 & 7 \end{bmatrix}\right) = Rank\left(\begin{bmatrix} 1 & -2 & 1 & -7 \\ 0 & 0 & -1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}\right) = 2$$

$$Rank([A|b]) = Rank\left(\left[\begin{array}{cccc|c} 2 & -4 & 2 & -14 & 10 \\ -1 & 2 & -2 & 11 & -6 \\ -1 & 2 & -1 & 7 & -5 \end{array}\right]\right) = Rank\left(\left[\begin{array}{cccc|c} 1 & -2 & 1 & -7 & 5 \\ 0 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]\right) = 2$$

As, it is clear from above that $Rank(A) = Rank([A|b]) < n$.
Hence, the given system of equations has an infinite number of solutions.
Hence, proved.

Solution 2

ConvProb can be written as follows,

$$\text{ConvProb: } \min_x \frac{1}{2}\|x\|_2^2 \qquad \text{s.t. } Ax = b$$

Constraint $Ax = b$ is linear function of x. As we know all the linear functions are convex in nature.
Hence, **constraint is convex**.

Objective function is,

$$\frac{1}{2}\|x\|_2^2 = \frac{1}{2}x^T x$$

Hessian of objective function is,

$$H(x) = I > 0, \qquad \text{for all values of } x$$

. Hence, **objective function is strongly convex** in nature.

Solution 3

Lagrangian for ConvProb is given by,

$$\mathcal{L}(x, \lambda) = \frac{1}{2}x^T x - \lambda^T(Ax - b)$$

Karush-Kuhn-Tucker (KKT) conditions are given by,

1. Primal Feasibility: $Ax^* = b$
2. Dual Feasibility: $\lambda \geq 0$
3. Stationary: $\nabla_x \mathcal{L}(x^*, \lambda^*) = x^* - A^T\lambda^* = 0$
4. Complementary Slackness: every element of $\lambda^T(Ax^* - b) = 0$ (obvious from Primal Feasibility)

From Stationary condition and Primal feasibility condition we get,

$$x^* = A^T\lambda^*$$
$$Ax^* = b$$

Substitute value of x^* we get,

$$(AA^T)\lambda^* = b$$

$$\lambda^* = (AA^T)^\dagger b$$

where, $(AA^T)^\dagger$ is the pseudo-inverse of AA^T because AA^T is rank deficient matrix, hence we can't calculate its inverse directly. Substituting value of λ^* in stationary condition, we get,

$$x^* = A^T(AA^T)^\dagger b$$

Value of x^* obtained using python code is,

$$x^* = [0.59574468 \quad -1.19148936 \quad -0.36170213 \quad -0.34042553]$$

Solution 4

Let projection of z on constraint set ($Ax = b$) be given by x. Therefore we have,

$$P_c(z) = \arg\min_x \frac{1}{2}\|x - z\|_2^2 \qquad \text{s.t. } Ax = b$$

Lagrangian of the problem is given by,

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

Now, applying KKT conditions (stationary condition) we get,

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = (x^* - z) - A^T\lambda^* = 0$$

$$x^* = z + A^T\lambda^*$$

Hence, now applying Primal Feasibility condition we get,

$$Ax^* = b$$

$$A(z + A^T\lambda^*) = b$$

$$\lambda^* = (AA^T)^\dagger(b - Az)$$

Hence, substituting value of λ^* in stationary condition. We finally get,

$$P_c(z) = x^* = (I - A^T(AA^T)^\dagger A)z + A^T(AA^T)^\dagger b$$

which is required projection.

Solution 5

Projected Gradient Descent is given by,

$$x^{(k+1)} = P_c(x^{(k)} - \alpha\nabla f(x^{(k)})) = P_c(x^{(k)} - \alpha x^{(k)})$$

Assumed Parameters are as follows,

1. $\alpha = [0.5, 0.25, 0.1, 0.075, 0.05, 0.025]$
2. $x^{(0)} = [1, 1, 1, 1]^T$
3. Thresold on $\|.\|_2 = 1e - 10$
4. Max Iterations $= 1000$

Step Size	**Final Solution**	**Final Error** $\|x_f - x^*\|_2$	**Number of Iterations**
0.5	$[0.59574, -1.19149, -0.3617, -0.34043]$	1.95834×10^{-10}	34
0.25	$[0.59574, -1.19149, -0.3617, -0.34043]$	1.27597×10^{-10}	82
0.1	$[0.59574, -1.19149, -0.3617, -0.34043]$	1.05186×10^{-10}	224
0.075	$[0.59574, -1.19149, -0.3617, -0.34043]$	1.00157×10^{-10}	303
0.05	$[0.59574, -1.19149, -0.3617, -0.34043]$	1.00234×10^{-10}	460
0.025	$[0.59574, -1.19149, -0.3617, -0.34043]$	1.00040×10^{-10}	931

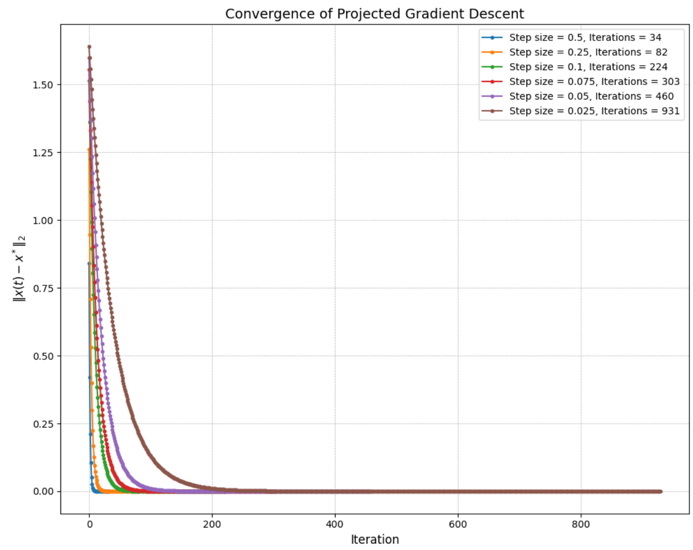

Figure 1: Convergence of Projected Gradient Descent using $\|.\|_2$

Support Vector Machines

Solution 1

Primal Function is given by,

$$w^*, b^* = \arg\min \frac{1}{2}\|w\|_2^2 \qquad \text{s.t. } y_i\left(w^T x_i + b\right) \geq 1 \quad \forall i = 1, 2, \ldots, N$$

For "Data.csv" and "Labels.csv", using CVXPY library in python, the primal solution turns out to be,

$$w^* = [1.1547, -2.0000] \qquad b^* = 1.0000$$

And the primal objective function value is,

$$p^* = 2.666666052505867$$

Solution 2

Given primal problem is,

$$w^*, b^* = \arg\min \frac{1}{2}\|w\|_2^2 \qquad \text{s.t. } y_i\left(w^T x_i + b\right) \geq 1 \quad \forall i = 1, 2, \ldots, N$$

Hence, the Lagrangian of the primal is given by,

$$\mathcal{L}(w, b, \lambda) = \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{N} \lambda_i \left(y_i\left(w^T x_i + b\right) - 1\right)$$

Then the KKT conditions for the primal problem are,

1. **Stationary:**

$$\nabla_w \mathcal{L}(w, b, \lambda) = w - \sum_{i=1}^{N} \lambda_i y_i x_i = 0 \qquad \Longrightarrow \quad w = \sum_{i=1}^{N} \lambda_i y_i x_i$$

$$\nabla_b \mathcal{L}(w, b, \lambda) = \sum_{i=1}^{N} \lambda_i y_i = 0$$

2. **Complementary Slackness:**
$$\lambda_i \left(y_i \left(w^T x_i + b \right) - 1 \right) = 0 \quad \forall i = 1, 2, \dots, N$$

3. **Primal Feasibility:**
$$y_i \left(w^T x_i + b \right) \geq 1 \quad \forall i = 1, 2, \dots, N$$

4. **Dual Feasibility:**
$$\lambda_i \geq 0 \quad \forall i = 1, 2, \dots, N$$

Substituting values into Lagrangian of the primal problem we get,

$$\mathcal{L}(w, b, \lambda) = \frac{1}{2} \left(\sum_{i=1}^{N} \lambda_i y_i x_i \right)^T \left(\sum_{j=1}^{N} \lambda_j y_j x_j \right) - \sum_{i=1}^{N} \lambda_i y_i \left(\left(\sum_{j=1}^{N} \lambda_j y_j x_j \right)^T x_i + b \right) + \sum_{i=1}^{N} \lambda_i$$

$$\mathcal{L}(w, b, \lambda) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_i \left(x_i^T x_j \right) - \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_i \left(x_i^T x_j \right) - \sum_{i=1}^{N} \lambda_i y_i b + \sum_{i=1}^{N} \lambda_i$$

$$\mathcal{L}(w, b, \lambda) = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_i \left(x_i^T x_j \right) + \sum_{i=1}^{N} \lambda_i$$

Hence, Dual problem is defined as,
$$\max_{\Lambda \geq 0} \quad g(\Lambda) = \Lambda^T b + \frac{1}{2} \Lambda^T A \Lambda \qquad \text{s.t. } \Lambda^T Y = 0$$

where notations are as follows,
$$\Lambda = [\lambda_1, \lambda_2, \dots, \lambda_N]^T \qquad \text{here, } k = N$$
$$Y = [y_1, y_2, \dots, y_N]^T$$
$$A_{i,j} = -y_i y_j \left(x_i^T x_j \right)$$
$$b_i = 1$$

Hence, final required values are as follows,

1. $A_{i,j} = -y_i y_j \left(x_i^T x_j \right)$
2. $b_i = 1$
3. $k = N$

Solution 3

From Stationary condition we get,
$$\sum_{i=1}^{N} \lambda_i y_i = 0$$

As we are given y_i can be $+1$ or, -1. Hence, we get,
$$\sum_{i:y_i=+1} \lambda_i y_i + \sum_{i:y_i=-1} \lambda_i y_i = 0 \qquad \implies \qquad \sum_{i:y_i=+1} \lambda_i - \sum_{i:y_i=-1} \lambda_i = 0$$

$$\sum_{i:y_i=+1} \lambda_i = \sum_{i:y_i=-1} \lambda_i = \gamma \text{ (say)}$$

For the given problem we have,
$$\sum_{i:y_i=+1} \lambda_i = 2.64776300458798$$
$$\sum_{i:y_i=-1} \lambda_i = 2.64705986447346$$

Hence, approximately, $\gamma = 2.6474114345307243$

Solution 4

Dual problem is defined as,

$$\min_{\Lambda \geq 0} \quad g(\Lambda) = \Lambda^T b + \frac{1}{2}\Lambda^T A \Lambda \qquad \text{s.t. } \Lambda^T Y = 0$$

where notations are as follows,

$$\Lambda = [\lambda_1, \lambda_2, \ldots, \lambda_N]^T \qquad \text{here, } k = N$$
$$Y = [y_1, y_2, \ldots, y_N]^T$$
$$A_{i,j} = y_i y_j \left(x_i^T x_j\right)$$
$$b_i = -1$$

Solving the problem using Projected Gradient Descent we get,
Assumptions,

1. Tolerance, $\|\lambda^{(t+1)} - \lambda^{(t)}\|_2 = 1e - 16$
2. Max Iterations $= 100000$
3. $\alpha = 1e - 4$
4. $\lambda_0 = [1, 1, \ldots, 1]^T$

Dual Objective Function Value at optimal point is,

$$d^* = 2.6672119866655404$$

which is equal to Primal objective function value (value may vary a little due to choice of α in Projected Gradient Descent method). This is due to the reason that Primal objective function and constraints are convex in nature and hence, **Strong Duality holds TRUE** here due to which $p^* = d^*$.

Optimal values of Dual variables are,

$$\Lambda = [\lambda_1, \lambda_2, \ldots, \lambda_{10}]^T = [1.80590286, 0.84186015, 0, 0, 0, 0, 2.43665503, 0, 0, 0.21040484]^T$$

And using these values of Λ we get, $w^* = [1.14988582, -1.98317533]^T$ and $b^* = 1$

Solution 5

It is clear from Solution 4 that, $\lambda_1, \lambda_2, \lambda_7$ and λ_{10} are not 0 and their values are,

$$\lambda_1 = 1.80590286 \neq 0$$
$$\lambda_2 = 0.84186015 \neq 0$$
$$\lambda_7 = 2.43665503 \neq 0$$
$$\lambda_{10} = 0.21040484 \neq 0$$

Hence, from Complementary Slackness of primal problem we get to know that for $i = 1, 2, 7$ and 10, we have,

$$y_i \left(w^T x_i + b\right) - 1 = 0$$

Therefore, **Primal Constraints $1, 2, 7$ and 10 are Active** and rest are Inactive, i.e. simply we can say that,

$$y_1 \left(w^T x_1 + b\right) = 1$$
$$y_2 \left(w^T x_2 + b\right) = 1$$
$$y_7 \left(w^T x_7 + b\right) = 1$$
$$y_{10} \left(w^T x_{10} + b\right) = 1$$

Solution 6

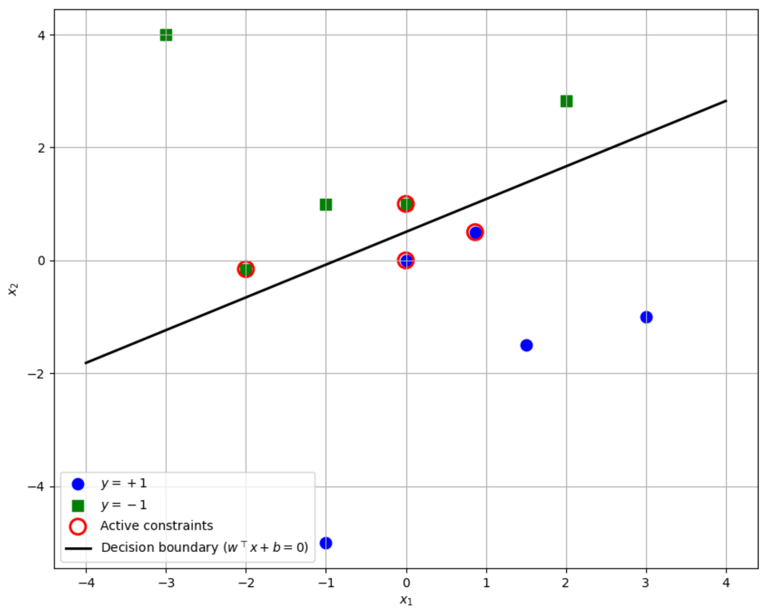

Figure 2: SVM for classification of data "Data.csv" and "Labels.csv"

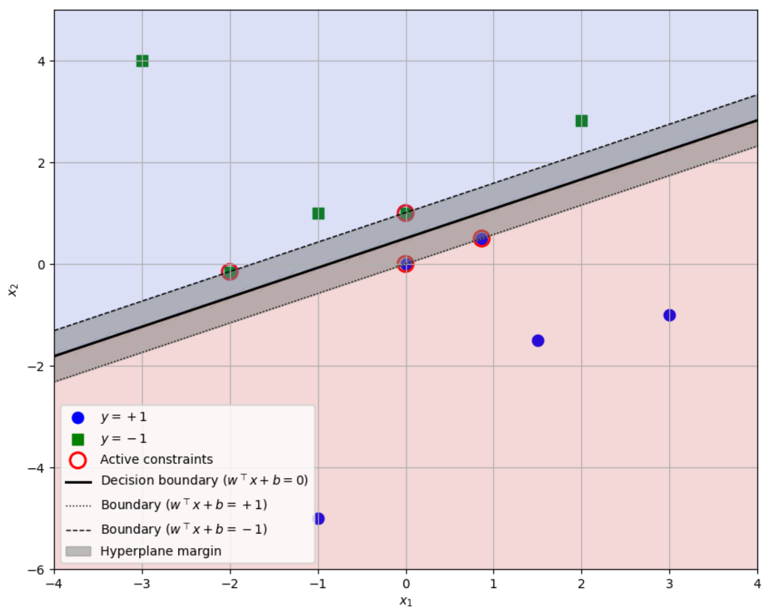

Figure 3: Explained SVM representation with separating hyperplane and decision space