CMO - Assignment - 3

Nirbhay sharma
SR: 24806
MTech - AI

November 14, 2024

1 Que-1

1.1 part-1

$$A = \begin{pmatrix} 2 & -4 & 2 & -14 \\ -1 & 2 & -2 & 11 \\ -1 & 2 & -1 & 7 \end{pmatrix} \tag{1}$$

$$b = \begin{pmatrix} 10 \\ -6 \\ -5 \end{pmatrix} \tag{2}$$

we write augmented matrix $(A|b)$ as follows and converting it to row echlon form as

$$(A|b) = \begin{pmatrix} 2 & -4 & 2 & -14 & 10 \\ -1 & 2 & -2 & 11 & -6 \\ -1 & 2 & -1 & 7 & -5 \end{pmatrix} \tag{3}$$

now apply operations as follows

$$R_1 \rightarrow R_1 + 2R_3 \tag{4}$$
$$R_2 \rightarrow R_2 - R_3 \tag{5}$$
$$R_3 \rightarrow -1R_3 \tag{6}$$
$$R_2 \rightarrow -1R_2 \tag{7}$$
$$R_1 \leftrightarrow R_3 \tag{8}$$

$$(A|b) = \begin{pmatrix} 1 & -2 & 1 & -7 & 5 \\ 0 & 0 & 1 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{9}$$

now applying same row operations to A as well, we get

$$A = \begin{pmatrix} 1 & -2 & 1 & -7 \\ 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{10}$$

we notice that rank(A) = rank$(A|b)$ = 2 (not full rank) and hence $Ax = b$ has infinitely many solutions

1.2 part-2

we pose the optimization problem (ConvProb) as follows

$$\min_{x \in R^4, Ax=b} \frac{1}{2}||x||^2 \tag{11}$$

$$\tag{12}$$

$Ax = b$ are the set of affine equations, which are convex in nature

to prove $\frac{1}{2}||x||^2$ as strongly convex, we know that a function is strongly convex if its hessian can be written as $\nabla^2 f(x) \succ mI$

we know for $f(x) = \frac{1}{2}||x||^2$, $\nabla^2 f(x) = 2I$ and hence it is strongly convex

1.3 part-3

(ConvProb) is written as

$$\min_{x \in R^4, Ax=b} \frac{1}{2}||x||^2 \tag{13}$$

$$L(x, \mu) = \frac{1}{2}||x||^2 - \mu^T(Ax - b) \tag{14}$$

$$\tag{15}$$

using KKT conditions we know $\nabla_x L(x, \mu) = 0$ and $Ax = b$, we get

$$\nabla_x L(x, \mu) = x - A^T\mu = 0 \implies x = A^T\mu \tag{16}$$

$$Ax = b \tag{17}$$

$$A(A^T\mu) = b \tag{18}$$

$$\mu = (AA^T)^{-1}b \tag{19}$$

$$x = A^T(AA^T)^{-1}b \tag{20}$$

for implementation purpose, since AA^T was not invertible matrix for

$$A = \begin{pmatrix} 2 & -4 & 2 & -14 \\ -1 & 2 & -2 & 11 \\ -1 & 2 & -1 & 7 \end{pmatrix} \tag{21}$$

Therefore we use non zero rows of echlon form of A which give us new set of constraints as follows

$$A = \begin{pmatrix} 1 & -2 & 1 & -7 \\ 0 & 0 & 1 & -4 \end{pmatrix}, b = \begin{pmatrix} 5 \\ 1 \end{pmatrix} \tag{22}$$

As shown in Fig 2, least norm solution is found by implementing the above algorithm

1.4 part-4

projection for (ConvProb) can be computed by solving the following problem

$$P_s(z) = \min_{x \in R^4, Ax=b} \frac{1}{2}||x - z||^2 \tag{23}$$

$$L(x, \lambda) = \frac{1}{2}||x - z||^2 - \lambda^T(Ax - b) \tag{24}$$

$$\nabla_x L(x, \lambda) = x - z - A^T\lambda = 0 \implies x = z + A^T\lambda \tag{25}$$

$$Ax = b \implies A(z + A^T\lambda) = b \implies \lambda = (AA^T)^{-1}(b - Az) \tag{26}$$

$$P_s(z) = x = z + A^T(AA^T)^{-1}(b - Az) \tag{27}$$

1.5 part-5

The gradient projection algorithm updates $x^{(k+1)} = P_s(x^k - \alpha \nabla f(x^k))$. Therefore using equation 27 for projection, we can apply projection gradient algorithm. Fig 2 shows the optimal x^*. The convergence based on varying α is shown in Fig 1

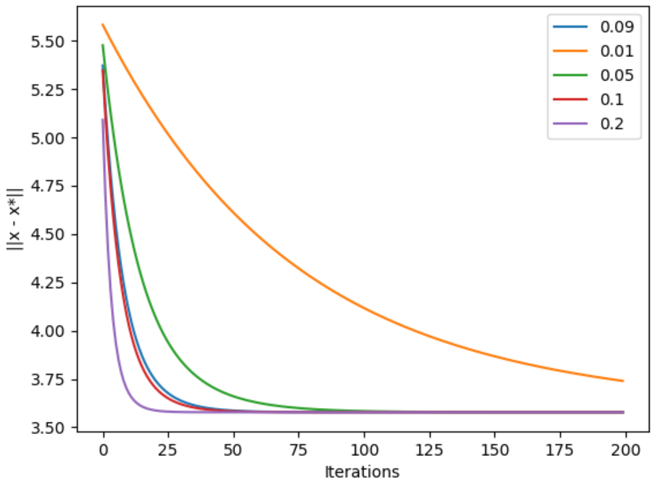
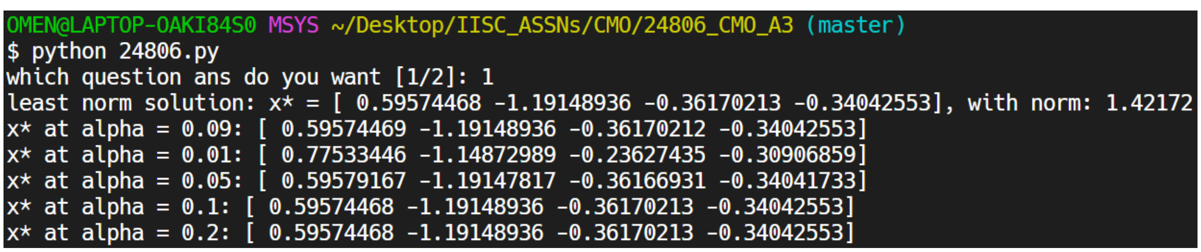


Figure 1: convergence with varying α

1.6 que-1 Code output



```
OMEN@LAPTOP-OAKI84S0 MSYS ~/Desktop/IISC_ASSNs/CMO/24806_CMO_A3 (master)
$ python 24806.py
which question ans do you want [1/2]: 1
least norm solution: x* = [ 0.59574468 -1.19148936 -0.36170213 -0.34042553], with norm: 1.42172
x* at alpha = 0.09: [ 0.59574469 -1.19148936 -0.36170212 -0.34042553]
x* at alpha = 0.01: [ 0.77533446 -1.14872989 -0.23627435 -0.30906859]
x* at alpha = 0.05: [ 0.59579167 -1.19147817 -0.36166931 -0.34041733]
x* at alpha = 0.1: [ 0.59574468 -1.19148936 -0.36170213 -0.34042553]
x* at alpha = 0.2: [ 0.59574468 -1.19148936 -0.36170213 -0.34042553]
```

Figure 2: code output for que1

2 Que-2

$$\min_{w,b} \frac{1}{2}||w||^2 \tag{28}$$

$$y_i(x_i^T w + b) \geq 1; i = 1, ...N \tag{29}$$

$$L(w, b, \lambda) = \frac{1}{2}||w||^2 - \sum_{i=1}^{N} \lambda_i (y_i(x_i^T w + b) - 1) \tag{30}$$

$$\nabla_w L(w, b, \lambda) = w - \sum_i \lambda_i y_i x_i \tag{31}$$

$$\nabla_b L(w, b, \lambda) = \sum_i \lambda_i y_i \tag{32}$$

$$\tag{33}$$

by KKT conditions we get $\nabla_w L(w,b,\lambda) = 0$, $\nabla_b L(w,b,\lambda) = 0$ and $\lambda_i(y_i(x_i^T w + b) - 1) = 0$

$$\nabla_w L(w,b,\lambda) = 0 \implies w = \sum_i \lambda_i y_i x_i \tag{34}$$

$$\nabla_b L(w,b,\lambda) = \sum_i \lambda_i y_i = 0 \tag{35}$$

$$\tag{36}$$

using both the conditions we get

$$\frac{1}{2}w^T w - \sum_i \lambda_i y_i w^T x_i - \sum_i \lambda_i y_i b + \sum_i \lambda_i \tag{37}$$

$$\frac{1}{2}\sum_i \lambda_i y_i x_i^T \sum_j \lambda_j y_j x_j - \sum_i \lambda_i y_i \sum_j \lambda_j y_j x_j^T x_i + \sum_i \lambda_i \tag{38}$$

$$g(\lambda) = -\frac{1}{2}\sum_i \sum_j \lambda_i \lambda_j y_i y_j x_i^T x_j + \sum_i \lambda_i \tag{39}$$

2.1 part-1

we solve primal problem using CVXPY and the objective function value is shown in Fig 4 which turns out to be 2.665

2.2 part-2

we know that $g(\lambda) = -\frac{1}{2}\sum_i \sum_j \lambda_i \lambda_j y_i y_j x_i^T x_j + \sum_i \lambda_i$ hence it can be written as

$$g(\Lambda) = \Lambda^T b + \frac{1}{2}\Lambda^T A \Lambda \tag{40}$$

where $b_i = 1 \forall i$ and $A_{ij} = y_i y_j x_i^T x_j$ and $k = N$ which are number of constraints

2.3 part-3

from KKT conditions we derived above, we get $\sum_i y_i \lambda_i = 0$, since $y_i = 1$ or $y_i = -1$, we can divide the sum as follows

$$\sum_i y_i \lambda_i = \sum_{i:y_i=1} \lambda_i - \sum_{i:y_i=-1} \lambda_i = 0 \implies \sum_{i:y_i=1} \lambda_i = \sum_{i:y_i=-1} \lambda_i = \gamma \tag{41}$$

as shown in Fig 4, γ_i turns out to be 2.668

2.4 part-4

We solve dual problem which is stated as follows using CVXPY and we get λ_i as output

$$\max_{\lambda_i \geq 0, \sum_i \lambda_i y_i = 0} g(\lambda)(= -\frac{1}{2}\sum_i \sum_j \lambda_i \lambda_j y_i y_j x_i^T x_j + \sum_i \lambda_i) \tag{42}$$

The dual objective function value at optimality is shown in Fig 4

2.5 part-5

The KKT conditions for primal problem involves the following condition as well $\forall i$

$$\lambda_i(y_i(w^T x_i + b) - 1) = 0 \tag{43}$$

when we solve dual problem we get optimal λ_i, hence for those constraints whose $\lambda_i > 0$ are active, The active constraints list is shown in shown in Fig 4

2.6 part-6

The hyperplane is shown in Fig 3

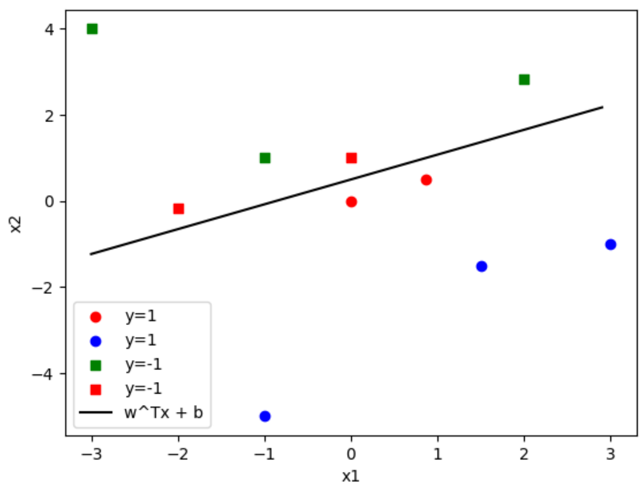


Figure 3: hyperplane for SVM

2.7 Que-2 code output

```
OMEN@LAPTOP-OAKI84S0 MSYS ~/Desktop/IISC_ASSNs/CMO/24806_CMO_A3 (master)
$ python 24806.py
which question ans do you want [1/2]: 2
PRIMAL: w = [ 1.15467408 -1.99991345], b = [0.99996244], objective value = 2.6665
DUAL: w = [ 1.15474913 -2.00010341], b = 1.0, objective value = 2.6667
sum_lambda (y = 1): 2.6668, (y=-1): 2.6668
active constraints are: [ 1  2  7 10]
```

Figure 4: code output for que2