IS5152 Data-driven Decision Making
SEMESTER 2, 2022-2023
Assignment 2 - Suggested solution

1. (10 points) A firm wants to know if the total promotional cost and the number of sales people in each of its shops can be used to predict whether the shops achieve their monthly profit targets. The data below have been obtained from the firm's internal records.

Shop #	Promotional cost (hundreds of dollars)	Number of sales people	Target achieved?
1	5	2	No
2	6	2	No
3	6	3	No
4	3	3	No
5	8	2	Yes
6	8	4	Yes
7	5	6	Yes
8	6	4	Yes

Assign $d_i = +1$ for class label *Yes*, and $d_i = -1$ for class label *No*.

(a) (2 points) State the primal quadratic programming (QP) problem for finding the hyperplane with the largest margin of separation.

$$\min \frac{1}{2}\left(w_1^2 + w_2^2\right)$$

subject to

$$\begin{aligned}
-5w_1 - 2w_2 - b &\geq 1 \\
-6w_1 - 2w_2 - b &\geq 1 \\
-6w_1 - 3w_2 - b &\geq 1 \\
-3w_1 - 3w_2 - b &\geq 1 \\
8w_1 + 2w_2 + b &\geq 1 \\
8w_1 + 4w_2 + b &\geq 1 \\
5w_1 + 6w_2 + b &\geq 1 \\
6w_1 + 4w_2 + b &\geq 1
\end{aligned}$$

(b) (2 points) Write the optimality conditions for your QP.

Feasibility and

$$
\begin{aligned}
w_1 + 5\alpha_1 + 6\alpha_2 + 6\alpha_3 + 3\alpha_4 - 8\alpha_5 - 8\alpha_6 - 5\alpha_7 - 6\alpha_8 &= 0 \\
w_2 + 2\alpha_1 + 2\alpha_2 + 3\alpha_3 + 3\alpha_4 - 2\alpha_5 - 4\alpha_6 - 6\alpha_7 - 4\alpha_8 &= 0 \\
\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \alpha_5 - \alpha_6 - \alpha_7 - \alpha_8 &= 0 \\
\alpha_1(-5w_1 - 2w_2 - b - 1) &= 0 \\
\alpha_2(-6w_1 - 2w_2 - b - 1) &= 0 \\
\alpha_3(-6w_1 - 3w_2 - b - 1) &= 0 \\
\alpha_5(-3w_1 - 3w_2 - b - 1) &= 0 \\
\alpha_5(8w_1 + 2w_2 + b - 1) &= 0 \\
\alpha_6(8w_1 + 4w_2 + b - 1) &= 0 \\
\alpha_7(5w_1 + 6w_2 + b - 1) &= 0 \\
\alpha_8(6w_1 + 4w_2 + b - 1) &= 0 \\
\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8 &\geq 0
\end{aligned}
$$

(c) (2 points) You are given the information that data points corresponding to Shop #3 (6,3), Shop #5 (8,2), and Shop # 8 (6,4) are support vectors. Find the optimal hyperplane.

$$
\begin{aligned}
-6w_1 - 3w_2 - b &= 1 \\
8w_1 + 2w_2 + b &= 1 \\
6w_1 + 4w_2 + b &= 1 \\
\text{Solution } w_1 = w_2 &= 2 \\
b &= -19
\end{aligned}
$$

(d) (4 points) Show that the hyperplane you find in (c) above is indeed optimal by finding Lagrange multipliers $\alpha_1, \alpha_2, \ldots, \alpha_8$ that satisfy all the conditions for optimality.

$$
\begin{aligned}
\alpha_1 = \alpha_2 = \alpha_4 = \alpha_6 = \alpha_7 &= 0 \\
w = \begin{pmatrix} 2 \\ 2 \end{pmatrix} &= -\alpha_3 \begin{pmatrix} 6 \\ 3 \end{pmatrix} + \alpha_5 \begin{pmatrix} 8 \\ 2 \end{pmatrix} + \alpha_8 \begin{pmatrix} 6 \\ 4 \end{pmatrix} \\
0 &= -\alpha_3 + \alpha_5 + \alpha_8 \\
\text{Solution } \alpha_3 &= 4 \\
\alpha_5 &= 1 \\
\alpha_8 &= 3
\end{aligned}
$$

- Complementarity: $\alpha_i\,(d_i(\mathbf{w}^t\mathbf{x}_i + b)) = 0$ for all $i = 1, 2, \ldots 8$.
- Feasibility: $d_i(\mathbf{w}^t\mathbf{x}_i + b) \geq 0$ for all $i = 1, 2, \ldots 8$.

$$
-5w_1 - 2w_2 - b = -14 - (-19) = 5 \quad \geq \quad 1
$$

$$
\begin{aligned}
-6w_1 - 2w_2 - b = -16 - (-19) = 3 &\quad \geq \quad 1 \\
-6w_1 - 3w_2 - b = -18 - (-19) = 1 &\quad \geq \quad 1 \\
-3w_1 - 3w_2 - b = -12 - (-19) = 7 &\quad \geq \quad 1 \\
8w_1 + 2w_2 + b = 20 + (-19) = 1 &\quad \geq \quad 1 \\
8w_1 + 4w_2 + b = 24 + (-19) = 5 &\quad \geq \quad 1 \\
5w_1 + 6w_2 + b = 22 + (-19) = 3 &\quad \geq \quad 1 \\
6w_1 + 4w_2 + b = 20 + (-19) = 1 &\quad \geq \quad 1
\end{aligned}
$$

- Multipliers: $\alpha_i \geq 0$ for all $i = 1, 2, \ldots 8$.

2. (10 points) Consider the data points in $\mathbb{R}^2$ as follows:

$$
\begin{aligned}
\mathbf{x_1} : \quad (0,1) &\qquad d_1 = +1 \ \ (Class\ \ 1) \\
\mathbf{x_2} : \quad (1,0) &\qquad d_2 = -1 \ \ (Class\ \ 2) \\
\mathbf{x_3} : (0,-1) &\qquad d_3 = +1 \ \ (Class\ \ 1) \\
\mathbf{x_4} : (-1,0) &\qquad d_4 = -1 \ \ (Class\ \ 2)
\end{aligned}
$$

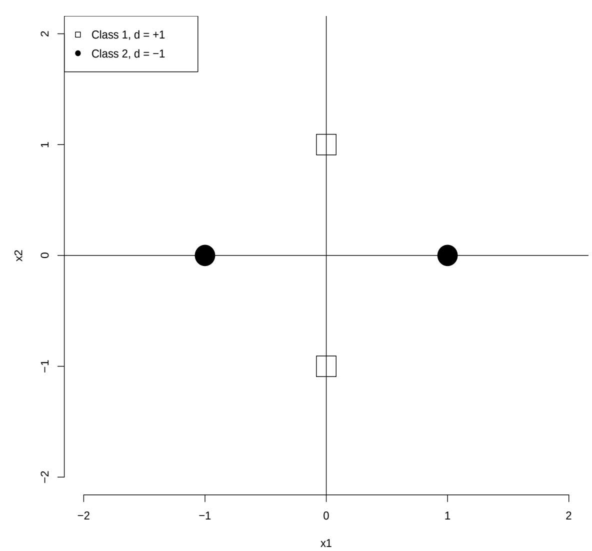

Figure 1: A binary classification problem with 4 data samples.

Show how a support vector machine with polynomial kernel ($p = 2$) would transform the input space to a higher dimensional feature space such that a *linear* decision surface can be constructed.

Hint: You *do not need* to find the optimal weights, instead find values for $(w_0, w_1, ..., \dots w_5)$ such that $\mathbf{w}^T\Phi(\mathbf{x}) > 0$ for Class 1 samples, and $\mathbf{w}^T\Phi(\mathbf{x}) < 0$ for Class 2 samples.

- $\Phi(\mathbf{x}) = \left(1, x_1^2, \sqrt{2}x_1x_2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2\right)^T$
- $\Phi(\mathbf{x_1}) = \left(1,0,0,1,0,\sqrt{2}\right)^T \quad \Phi(\mathbf{x_2}) = \left(1,1,0,0,\sqrt{2},0\right)^T$
- $\Phi(\mathbf{x_3}) = \left(1,0,0,1,0,-\sqrt{2}\right)^T \quad \Phi(\mathbf{x_4}) = \left(1,1,0,0,-\sqrt{2},0\right)^T$

Let $w_0 = 0, w_1 = -1, w_2 = 0, w_3 = 1, w_4 = 0, w_5 = 0$, then

- $\Phi(\mathbf{x})^T\mathbf{w} = -x_1^2 + x_2^2$
- $\Phi(\mathbf{x_1})^T\mathbf{w} = 1 \quad \Phi(\mathbf{x_2})^T\mathbf{w} = -1$
- $\Phi(\mathbf{x_3})^T\mathbf{w} = 1 \quad \Phi(\mathbf{x_4})^T\mathbf{w} = -1$

3. (10 points) We are investigating the efficiency of 3 local tuition centers using the input oriented CCR approach. The following monthly information has been collected.

Input 1 = number of tutors,

Input 2 = rental cost(hundred \$),

Output 1 = number of students who are in primary school,

Output 2 = number of students who are in secondary school.

	Inputs		Outputs	
	1	2	1	2
Center 1	10	20	40	4
Center 2	20	24.5	35	20
Center 3	10	12	20	36

(a) (4 points) State the primal and dual linear programs for Center 2.

```
Primal LP: max z = 35 t1 + 20 t2
   s.t. -40 t1 -  4 t2 + 10 w1 + 20   w2 >= 0
        -35 t1 - 20 t2 + 20 w1 + 24.5 w2 >= 0
        -20 t1 - 36 t2 + 10 w1 + 12   w2 >= 0
                         20 w1 + 24.5 w2  = 1
                       t1, t2, w1, w2    >= 0


Dual LP: min e
  s.t  10 L1 + 20 L2 + 10 L3 - 20   e   <= 0
       20 L1 + 24.5 L2 + 12 L3 - 24.5 e <= 0
```

```
40 L1 + 35 L2 + 20 L3 - 35        >= 0
 4 L1 + 20 L2 + 36 L3 - 20        >= 0
              L1, L2, L3, e       >= 0
```

(b) (2 points) When the primal problem is solved, it is found that the values for the outputs are 0.02 and 0.0025 for output 1 and output 2, respectively. The corresponding cost for the inputs are 0.001 and 0.04 for input 1 and input 2, respectively. Explain how we can conclude that this Center 2 is not efficient.

$Z = 35t_1 + 20t_2 = 35(0.02) + 20(0.0025) = 0.70 + 0.05 = 0.75 < 1.$

(c) (4 points) Suggest two ways to improve the efficiency of Center 2.

- Increase its output by a factor $1/0.75 = 1.3333$ (increase by 33.33 %), new output:

$$\frac{4}{3}\begin{pmatrix} 35 \\ 20 \end{pmatrix} = \begin{pmatrix} 46.6667 \\ 26.6667 \end{pmatrix}$$

- or decrease its input by 25% to 75% current level, new input:

$$\frac{3}{4}\begin{pmatrix} 20 \\ 24.5 \end{pmatrix} = \begin{pmatrix} 15 \\ 18.375 \end{pmatrix}$$

4. (10 points) A local bank makes 7 different types of loans available to its customers. These loans and their associated interest rates are listed below.

Type of loan		Interest rates per annum (%)
1.	Small & Medium Enterprises	1.2
2.	Home mortgage	1.4
3.	Industrial	1.5
4.	Commercial	1.7
5.	Automobile	1.8
6.	Home renovation	2.0
7.	Unsecured	2.3

A total amount of $20 million is available for all loans this year at a new branch of the bank. Let $x_1, x_2, \ldots x_7$ be the amount (in million dollars) allocated to the seven types of loans in the above table. The following requirements have been established for the allocation of the money at this new branch:

- the total amount of money given out as loans is actually $20 million.
- no more than 10% of the total dollar amount of loans may be Unsecured.
- at least 50% of the total dollar amount of loans must be Small & Medium Enterprise, Home mortgage, or Commercial loans.

- the total Small & Medium Enterprise, Automobile, and Unsecured loans cannot exceed 30% of the total dollar amount of loans.
- the total Industrial and Home-renovation loans must be less than or equal to the total of Small & Medium Enterprise and Commercial loans.
- no loan should exceed \$8 million.

(a) (2 points) Formulate a linear program (LP) model to maximize total annual interest.

$$\max Z = 0.012x_1 + 0.014x_2 + 0.015x_3 + 0.017x_4 + 0.018x_5 + 0.020x_6 + 0.023x_7$$

subject to

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 &= 20 \\ x_7 &\leq 2 \\ x_1 + x_2 + x_4 &\geq 10 \\ x_1 + x_5 + x_7 &\leq 6 \\ x_3 + x_6 - x_1 - x_4 &\leq 0 \\ 0 \leq x_1, x_2, x_3, x_4, x_5, x_6, x_7 &\leq 8 \end{aligned}$$

(b) (6 points) The maximum total interest obtained from the optimal mix of loans is \$360,000, or an effective interest rate of 1.8%. Suppose now we still have the same requirements as before (as in part (a)), but the following goals are also to be achieved:

Goal 1: Effective interest rate of at least 1.8%.

Goal 2: At least \$2 million for Home mortgage.

Goal 3: At least \$1 million for Small & Medium Enterprise.

Goal 4: The combined total of Small & Medium Enterprise and Industrial loans is at least \$4 million.

It is decided that goal 1 is twice as important as goal 4 and that goal 2 is twice as important as goal 1, while goal 3 and goal 4 are equally important.

i. (3 points) Formulate the goal programming LP model for this problem.

$$\min W = 2s_1^- + 4s_2^- + s_3^- + s_4^-$$

subject to

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 &= 20 \\ x_7 &\leq 2 \\ x_1 + x_2 + x_4 &\geq 10 \\ x_1 + x_5 + x_7 &\leq 6 \end{aligned}$$

$$\begin{aligned}
x_3 + x_6 - x_1 - x_4 &\leq 0 \\
0 \leq x_1, x_2, x_3, x_4, x_5, x_6, x_7 &\leq 8 \\
0.012x_1 + 0.014x_2 + 0.015x_3 + 0.017x_4 + 0.018x_5 + 0.020x_6 + 0.023x_7 + s_1^- - s_1^+ &= 0.36 \\
x_2 + s_2^- - s_2^+ &= 2 \\
x_1 + s_3^- - s_3^+ &= 1 \\
x_1 + x_3 + s_4^- - s_4^+ &= 4
\end{aligned}$$

ii. (3 points) The optimal values are as follows: $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 7, x_5 = 0, x_6 = 5, x_7 = 2$. Which of the four goals (if any) are not achieved? Explain briefly.

- Goal 1: $0.012x_1 + 0.014x_2 + 0.015x_3 + 0.017x_4 + 0.018x_5 + 0.020x_6 + 0.023x_7 = 0.012 + 0.028 + 0.045 + 0.119 + 0 + 0.100 + 0.046 = 0.35$. This goal is not achieved.
- Goal 2: $x_7 = 2$. This goal is achieved.
- Goal 3: $x_1 = 1$. This goal is achieved.
- Goal 4: $x_1 + x_3 = 4$. This goal is achieved.

(c) (2 points) Suppose now Goal 1 has the highest priority, Goal 2 the second highest priority, goal 3 the third highest priority, and Goal 4 priority has the lowest priority. Goal with higher priority must be achieved, before any lower priority goals are considered.

- How would you formulate this requirement?
  Let the objective function be:
  $$\min \; W = P_1 s_1^- + P_2 s_2^- + P_3 s_3^- + P_4 s_4^-$$
  where $P_1 >>> P_2 >>> P_3 >>> P_4$.
- Do you expect the solution to be different from the optimal values in part (b-ii)? Explain briefly.

Yes, since currently Goal 1 is not achieved. When a pre-emptive goal programming approached is applied with this goal as a top priority, the loan mix would be such this goal is first satisfied before considering the other 3 goals of lower priority. This first goal is achievable if we ignore the other 3 goals, as the solution of the original LP gives a total return of 0.36 (million).