Optimization - Practical Report

Robin Khatri, M1 MLDM

March 24, 2019

1 Unconstrained Optimization

1.1 Problem 1

Consider the following problem:

$$\min_{x_1,x_2} (x_1 - 4)^2 + 7(x_2 - 4)^2 + 4x_2$$

- Formulate the problem in AMPL (using a model file only) and solve it. Check that you get the correct solution by solving on your own.

1.2 Solution

AMPL Code:

```
reset;
options solver loqo;
var x1; # Define decision variables
var x2;

minimize f: (x1-4)^2 + 7*(x2-4)^2 + 4*x2; # Objective function

solve; # Display optimized value for objective function
display x1; # Display value of x1 for optimized value of objective function
display x2; # Display value of x2 for optimized value of objective function
```

We get following output on running the above model in AMPL.

```
ampl: model problem1_model.mod
LOQO 7.03: optimal solution (2 QP iterations, 3 evaluations)
primal objective 15.42857143
```

```
  dual objective 15.42857143
x1 = 4

x2 = 3.71429
```

So, we get `Minimized value of the objective function = 15.42857143` `x1 = 4` and `x2 = 3.71429`

1.3 Verification

Partial differentials:

$$f = (x_1 - 4)^2 + 7(x_2 - 4)^2 + 4x_2$$

$$\frac{\partial f}{\partial x_1} = 2 * (x_1 - 4)$$

$$\frac{\partial f}{\partial x_2} = 14 * (x_2 - 4) + 4$$

For minimum value of f,

$$\frac{\partial f}{\partial x_1} = 0 => 2 * (x_1 - 4) = 0 => x_1 = 4$$

$$\frac{\partial f}{\partial x_2} = 0 => 14 * (x_2 - 4) + 4 = 0 => x_2 = \frac{52}{14} => x_2 = 3.17429$$

H is the Hessian.

$$H = \begin{bmatrix} 2 & 0 \\ 0 & 14 \end{bmatrix}$$

For point $4, 3.17429$,

$$det(H) = 28 > 0, H[1, 1] = 2 > 0$$

So, we have a minima at $(4, 3.17429)$

For point $(4, 3.17429)$,

$$f(4, 3.17429) = 15.4285$$

It is the same solution that we got from our AMPL model.

1.4 Problem 2

Consider the following problem:

$$\min_{x_1,x_2,x_3} \quad 2x_1^3 + 2(3x_2 - x_3)^2 - 4x_3 + 2$$

- Is this problem convex? Why?
- Find the analytical solution to the problem (if needed do not hesitate to add some constraints).
- Formulate the problem in AMPL (using a model file only) and solve it. Check that you get the correct solution by solving the problem on your own.

2 Solution

$$f = 2x_1^3 + 2(3x_2 - x_3)^2 - 4x_3 + 2$$

Here as x_3 increases in value, function f decreases. It is essentially unbounded below for x_3. We put a constraint,

$$\begin{aligned} -4x_3 + 2 \leq 0 \\ \implies x_3 \leq 0.5 \end{aligned}$$

We now have a constrained problem, Further, for $det(H) \geq 0$ ($det(H)$ is the product of eigen values of Hessian and so it should not be negative for a convex function. For convexity, we must have $x_1 \geq 0$. Our transformed problem is,

$$\begin{aligned} \min_{x_1,x_2,x_3} \quad & 2x_1^3 + 2(3x_2 - x_3)^2 - 4x_3 + 2 \\ \text{subject to} \quad & x_1 \geq 0, x_3 \leq 0.5 \end{aligned}$$

Clearly, x_3 at which f achieves minimum values is 0.5 in our constrained problem. Setting x_3 to 0.5, we observe following graph for function f.

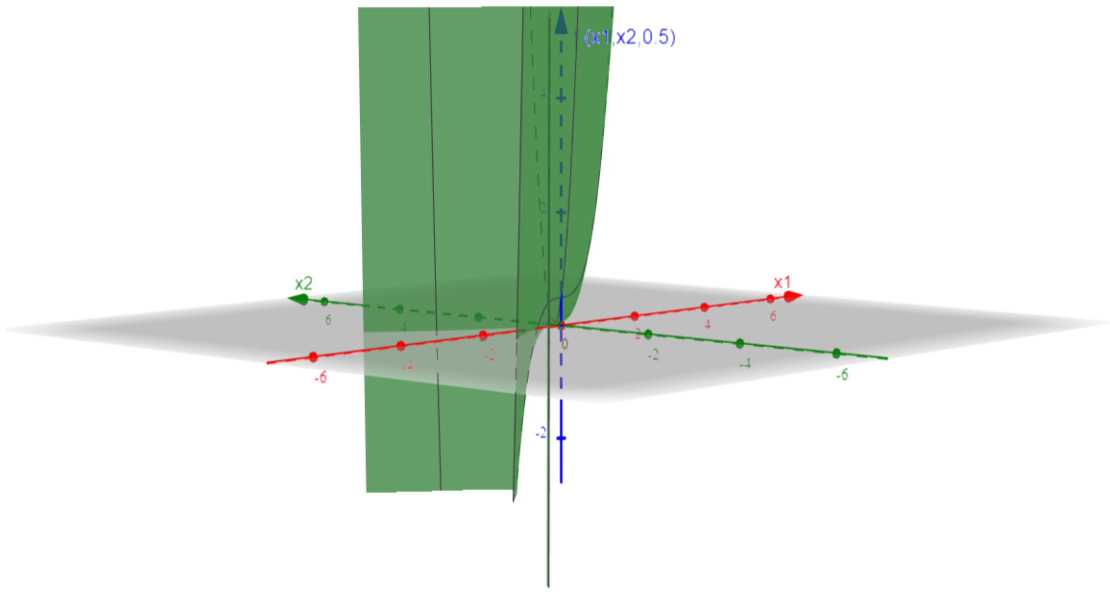

Figure 1: $f(x_1, x_2, 0.5)$

$x_1 = 0, x_2 = 0.16667$ is an inflexion point. Function $f(x_1, x_2, 0.5)$ becomes convex from concave at this point. This point therefore is not in our convex set. We can choose a convex set as $(x_1, x_2, 0.5) : x_1, x_2 \in R^+$ - (0,0.16667).

Following AMPL code was used for optimization of this transformed problem:

AMPL code:

```
reset;
option solver loqo;

var x1 >= 0;
var x2;
var x3 <= 0.5;

minimize f: 2*x1^3 + 2*(3*x2 - x3)^2 - 4*x3 + 2

solve;

display x1;
display x2;
display x3;
```

Output of the AMPL code is:

```
ampl: model problem2.mod;
LOQO 7.03: optimal solution (20 iterations, 20 evaluations)
primal objective 5.650402368e-09
  dual objective -5.050275982e-09
x1 = 0.00127898

x2 = 0.166667

x3 = 0.5
```

The solution belongs to our convex set and achieves value close to 0 (which is close to saddle point and inflexion point for the projection of plot on the xz plane as per figure 1.) AMPL could not find the minima at $x_1 = 0$ and $x_2 = 0.16667$ for $x_3 = 0.5$ because at this point, this is a point where the curve changes its behavior.

3 Problem 3

Consider the following constrained problem:

$$\min_{x_1, x_2 \in R} \quad (x_1 - 2)^2 + 3 * x_2, s.t. - x_1 - x2 + 4 \geq 0$$

- Reformulate this problem using the log-barrier function in AMPL;
- Try different values of the coefficient associated to the log-barrier function. What do you observe?

3.1 Solution

Here variable x_2 can be $-\infty$ making the minimum value of the objective function $(x_1 - 2)^2 + 3 * x_2 \ -\infty$. We can however set a constraint that $x_2 \geq 0$.

To solve this problem using log barrier function, we can make add $-log(x_1 - x_2 + 4)$ to our objective function. So, our updated obtimization problem would be:

$$\min_{x_1, x_2 \in R} \quad (x_1 - 2)^2 + 3 * x_2 - log(-x_1 - x2 + 4)$$

s.t. $x_1 \geq 0$, $x_2 \geq 0$ and $t \geq 0$ is a large value.

and we reformulated our objective function into AMPL code as follows:

AMPL Code:

```
var x>=0;
var y>=0;
```

```
param t = 10; # t can be changed
minimize objective:
(x-2)^2+3*y - (1/t)*log(-x-y+4);
```

For the above code, we obtained following output:

```
ampl: model model3.mod;
LOQO 7.03: iterlim 30
dual
LOQO 7.03: optimal solution (10 iterations, 10 evaluations)
primal objective -0.06993206144
  dual objective -0.06993206864
x = 1.9753

y = 8.96067e-10
```

On changing the value of t to 8,10,12,15,200,300, we obtained following results:

- t = 8: Objective = -0.08760508342, x = 1.96922, y = 8.69176e-10
- t = 10: Objective = -0.06993206144, x = 1.9753, y = 8.96067e-10
- t = 12: Objective = -0.05819184574, x = 1.97938, y = 9.14371e-10
- t = 15: Objective = -0.04648530377, x = 1.98347, y = 9.32982e-10
- t=200: Objective = -0.003467294411, x = 1.99875, y = 1.00457e-09
- t = 300 : Objective = -0.00277358520, x = 1.91988, y = 1.00577e-09

We tested with other values of t as well. As t tends to infinity, the approximation became more and more accurate and the function becomes closer to an indicator function. However, we could not observe a very significant approximation with increasing t because the model was approximated pretty well at a low level of t.

4 Modeling constrained problems

4.1 Water resources

A city needs 500,000 liters of water per day, which can be drawn from either a reservoir or a stream. Characteristics of these two sources are found in Table 1. No more than 100,000 liters per day can be drawn from the stream, and the concentration of pollutants in the water served to the city cannot exceed 100 ppm. The problem is to find how much water the city should draw from each source (subject to the constraints) so as to minimize the cost.

	Reservoir	Stream
Cost (€ per 1,000L)	100	50
Upper limit ($\times$ 1,000L)	∞	100
Pollution (ppm)	50	250

4.1.1 Solution

In order to solve the water resources problem, we need to introduce two variables. Those variables are r for the reservoir and s for the stream.

Mathematical Model:

The decision variables are:

r = Amount of water drawn from reservoir(In 1,000's L)
s = Amount of water drawn from stream (In 1,000's L)

Objective function:

$$C = 100r + 50s$$

subject to:

$$s \leq 100$$

$$50r + 250s \leq 100(r + s)$$

$$r + s \geq 500$$

AMPL code (personal41.mod):

```
reset;
option solver loqo;

var r >= 0; # Define decision variables (in 1000 L's)
var s >= 0;

minimize objective: 100*r + 50*s;

subject to constraint1: s <= 100; # Upper limit
subject to constraint2: 50*r + 250*s <= 100*(r+s); # Pollution (ppm)
subject to constraint3: r+s >= 500; # Need of the city (In 1,000's L)
```

```
solve;
display r;
display s;
```

The output of the AMPL code is:

```
ampl: model problem41.mod;
LOQO 7.03: iterlim 30
dual
LOQO 7.03: optimal solution (8 QP iterations, 8 evaluations)
primal objective 45000.00065
  dual objective 44999.99978

"option abs_boundtol 6.078933125763797e-07;"
or "option rel_boundtol 6.078933125763797e-09;"
will change deduced dual values.

r = 400

s = 100
```

The resulting solution is `r = 400` and `s = 100` which means that we should get 400,000 liters from the reservoir and 100,000 liters from the stream. Minimized cost is $100 \times 400 + 50 \times 100 = 45000$ €.

4.2 Good-smelling perfume design

A perfume company located in the south of France is trying to develop new perfumes based on mixes of 4 blends of essential oils developed in their lab. Each of these 4 blends embeds some essences of real flowers (rose, bergamot orange, lily of the valley, thymus). The table below describes the composition of each blend and the cost to produce each of them.

	blend 1	blend 2	blend 3	blend 4
rose	30	20	40	20
bergamot orange	35	60	35	40
lily of the valley	20	15	5	30
thymus	15	5	20	10
Cost(€/liter)	55	65	35	85

The company wants to develop its own trademark and imposes some contraints on the possible mixes to design a new perfume:

- the percentage of blend 2 in the perfume must be at least 5
- the percentage of blend 3 has to be at least 30

- the percentage of blend 1 has to be between 10
- the final percentage of bergamot orange content in the perfume must be at most 50
- the final percentage of thymus content has to be between 8
- the final percentage of rose content must be at most 35
- finally, the percentage of lily of the valley content has to be at least 19

We are looking for the least costly way of mixing the 4 blends of essential oils to produce a new perfume subject to the constraints given above. Questions:

1. Model this problem as an constrained optimization problem.
2. Formulate it in AMPL (using a model file only) and solve it. Check that the optimal solution makes sense. What is the optimal cost?

4.2.1 Solution

To solve the prefume design problem, we introduce four variables for the proportions of the four blends to be "b1", "b2", "b3" and "b4" respectively. Since these are proportions of the perfume, the sum of these proportions should be 1.

Mathematical Model:

The decision variables are:

b1: Proportion of blend 1 in the new perfume
b2: Proportion of blend 1 in the new perfume
b3: Proportion of blend 1 in the new perfume
b4: Proportion of blend 1 in the new perfume

Minimize objective function:

$$C = 55b1 + 65b2 + 35b3 + 85b4$$

Subject to:

$$\text{b1 + b2 + b3 + b4} = 1$$
$$0.05 \leq b2 \leq 0.2$$
$$\text{b3} \geq 0.3$$
$$0.1 \leq b1 \leq 0.25$$

$$0.35\text{b1} + 0.6\text{b2} + 0.35\text{b3} + 0.4\text{b4} \leq 0.5$$
$$0.08 \leq 0.15b1 + 0.05b2 + 0.2b3 + 0.1b4 \leq 0.13$$
$$0.3\text{b1} + 0.2\text{b2} + 0.4\text{b3} + 0.2\text{b4} \leq 0.35$$
$$0.2\text{b1} + 0.15\text{b2} + 0.05\text{b3} + 0.3\text{b4} \geq 0.19$$
$$\text{b1} \geq 0, b2 \geq 0, b3 \geq 0, b4 \geq 0$$

Non-negative constraints on $b1$, $b2$, $b3$ are also satisfied by by first three constraints.

AMPL Code (model_problem3.2.mod):

```
reset;
options solver loqo;

var b1 >= 0;
var b2 >= 0;
var b3 >= 0;
var b4 >= 0;

minimize Cost: 55*b1 + 65*b2 + 35*b3 + 85*b4;

subject to constraint1: b1+b2+b3+b4 = 1;
subject to constraint2: 0.05 <= b2 <= 0.2;
subject to constraint3: b3 >= 0.3;
subject to constraint4: 0.1 <= b1 <= 0.25;
subject to constraint5: 0.35*b1 + 0.6*b2 + 0.35*b3 + 0.4*b4 <= 0.5;
subject to constraint6: 0.08 <= 0.15*b1 + 0.05*b2 + 0.2*b3 + 0.1*b4 <= 0.13;
subject to constraint7: 0.3*b1 + 0.2*b2 + 0.4*b3 + 0.2*b4 <= 0.35;
subject to constraint8: 0.2*b1 + 0.15*b2 + 0.05*b3 + 0.3*b4 >= 0.19;

solve;
display cost;
display b1;
display b2;
display b3;
display b4;
```

The output of the AMPL code is:

```
ampl: model model_problem3.2.mod
LOQO 7.03: optimal solution (12 QP iterations, 12 evaluations)
primal objective 63.00000007
  dual objective 62.9999998
```

```
"option abs_boundtol 7.960464509793042e-10;"
or "option rel_boundtol 2.653488169931014e-09;"
will change deduced dual values.

b1 = 0.14

b2 = 0.14

b3 = 0.3

b4 = 0.42
```

There are eight constraints used for the model. The first one basic as it indicates that all the proportions of the blends are percentages so they must add to 100. Next, the following three constraints limit those proportions of the blends. Then, the four constraints after that are limiting the usages of essences of real flowers in the blends that are forming our new wanted blend.

The solution is `b1 = 0.14`, `b2 = 0.14`, `b3 = 0.3` and `b4 = 0.42`. This means to prepare the product with 14%, 14%, 30% and 42% in order for the four blends.

4.3 Roadway expenses

4.3.1 Rural/Urban case

France has 200M€ to spend on roadway improvements this coming year, and the government has to decide how much to spend on rural projects, and how much to spend on urban projects. Let x_{rural} and x_{urban} represent the amount of money spent on these two categories, in millions of euros. The benefit from spending x_rural on rural projects is:

$$B_{rural} = 7000log(1 + x_{rural}),$$

and the benefit from spending xurban on urban projects is

$$B_{urban} = 5000log(1 + x_{urban}).$$

We want to maximize the net benefit to the state, that is, $B_{rural} + B_{urban} - x_{rural} - x_{urban}$. Questions:

1. Model this problem as a constrained optimization problem.

2. Formulate it in AMPL (using a model file only) and solve it. Check that the optimal solution makes sense. What is the optimal benefit?

4.3.2 Solution

In order to solve the Roadway expenses problem, we need to introduce two variables that will be provide the solution later which is the amount of cost for rural areas and the cost for urban areas. Those variables are “a” for amount spent on rural projects and “b” for amount spent on urban projects.

Mathematical Model:

Decision variables:

a: Amount to spend on rural projects (In Million Euros)
b: Amount to spend on urban projects (In Million Euros)

Maximize Benefit:

$$B = 7000log(1 + a) + 5000log(1 + b) - a - b$$

Subject to:

$$a + b \leq 200$$
$$a \geq 0, b \geq 0$$

AMPL Code(model_problem3.3.mod):

```
reset;
option solver loqo;

var a >= 0;
var b >= 0;

maximize Benefit: (7000*log(1 + a)) + (5000*log(1 + b)) - a - b;

subject to constraint: a + b <= 200;
```

The output of the AMPL code is:

```
ampl: model model_problem3.3.mod
LOQO 7.03: optimal solution (20 iterations, 20 evaluations)
primal objective 55348.89305
  dual objective 55348.89343
```

```
a = 116.833

b = 83.1667
```

The resulting solution is `a = 116.833` and `b = 83.1667` which means that we should spend 116.8 millions on rural areas and 83 million euros on urban areas. Maximized benefit is 55348.893 €.

4.3.3 General case

Now suppose we have a set I of categories of roads in which we can invest. The benefit from spending x_i on projects of category i is:

$$B_i = C_i log(1 + x_i),$$

where C_i's are parameters of the model (not variables). The available budget (in millions of euros) to be spent is given by parameter budget. The net benefit to the state is now: $\sum_{i \in I} B_i - \sum_{i \in I} x_i$. In the previous case, we had budget $= 200$, $I = rural, urban$, $C_{rural} = 7000$ and $C_{urban} = 5000$.

Questions:

1. Model this general case using an AMPL model file where no numbers appear (except for the nonnegativity constraints). You must use parameters for C_i 's and budget, and a set for the road types. You can define a set myset and a parameter myparam for each element of the set in the following way:

   ```
   set myset;
   param myparam{i in my_set};
   ```

2. Write an AMPL data file that assigns the value of the previous section. Solve the problem and make sure that you get the same result as before.

3. Write a new AMPL data file that adds a third road category to the problem (suburban) with $C_{suburban} = 7000$ and solve the problem.

4.3.4 Solution

Mathematical Model:

Decision Variables:

$$X_i = \text{Amount spend on projects of category } i,\ I \text{ is the set of categories}$$

Maximize Benefit:

$$Z = \sum_{i \in I} (C_i log(1 + x_i)) - \sum_{i \in I} x_i$$

Subject to:

$$\sum_{i \in I} x_i \leq Budget$$

Where Budget is the available budget.

$$C_i \geq 0$$

AMPL code for model in general case:

```
reset;
option solver loqo;

set items;
var x{i in items}; # Define decision variables
param C{i in items};
param budget;

maximize Benefit: sum{i in items} (C[i]*log(1 + x[i]) - x[i]);
subject to constraint: sum{i in items} (x[i]) <= budget;
```

We can solve problems for this general model by supplying a data file with values for budget and C's.
To solve the previous problem (Section 3.3.2) with this code, we can supply following data file (problem3.3.2.dat).

```
set items := rural urban;
param C := rural 7000 urban 5000;
param budget := 200;
```

The output of the AMPL code is:

```
ampl: model model_problem3.3.3.mod;
ampl: data problem3.3.2.dat;
ampl: solve;
LOQO 7.03: optimal solution (27 iterations, 27 evaluations)
primal objective 55348.89318
  dual objective 55348.89318
ampl: display Benefit;
Benefit = 55348.9

ampl: display x;
x [*] :=
rural  116.833
urban   83.1667
```

After executing the model and the data file, it provides the same answer as before. Now we'll add a new category suburban in the data file with C = 7000. A new file `problem3.3.3.dat`) is defined as following:

```
set items := rural urban suburban;
param C := rural 7000 urban 5000 suburban 7000;
param budget := 200;
```

The output of the AMPL code is:

```
ampl: model model_problem3.3.3.mod;
ampl: data problem3.3.3.dat;
ampl: solve;
LOQO 7.03: optimal solution (25 iterations, 25 evaluations)
primal objective 80096.50464
  dual objective 80096.50465
ampl: display Benefit;
Benefit = 80096.5

ampl: display x;
x [*] :=
   rural  73.7895
suburban  73.7895
   urban  52.4211
```

The result of such data is `rural 73.7895`, `suburban 73.7895` and `urban 52.4211`. Hence, France should spend 73.7 million euros on roadway improvements on rural areas, equally on suburban areas and 52.4 millions on urban areas.

4.4 Design you own optimization problem

In this section, we defined our own optimization problem. I am fond of travelling and therefore I chose a simple problem which I can solve by optimization. I want to visit Tallinn (Estonia), Warsaw (Poland), Pompei (Italy), Lviv (Ukraine), Stavagner (Norway), and Utrecht (Netherlands). For this purspose, I want to know how many days I should spend in which city with a budget of maximum 2000 Euros. Since that will be an integer problem. For the purpose of this exercise, therefore I have considered days to be a real number.

The condition is also that I should spend at least 2 day in each of the city.

Mathematical Model:

If d_i is the number of days I should spend in one city, then

$$\begin{gathered}\text{Maximize} \sum d_i \\ \text{subject to } d_i \geq 2, \text{ for all } i\text{'s} \\ \text{and} \sum (flight_i + d_i * living_i) <= 2000\end{gathered}$$

Where $flight_i$ and $living_i$ represent flight and living costs in a city respectively.

The AMPL code was defined as follows(personal.mod):

```
reset;
option solver loqo;

set cities;
var d{i in cities} >= 2; # Define decision variables
param flightCost{i in cities};
param livingCost{i in cities};

maximize f: sum{i in cities} (d[i]);
subject to constraint: sum{i in cities} (flightCost[i] + d[i]*livingCost[i]) <= 2000;

data personal.dat;
solve;
display p;
```

The data file was defined as follows:

```
  set cities := Talinn Warsaw Pompei Stavagner Lviv Utrecht;
param flightCost := Talinn 140 Warsaw 180 Pompei 220 Stavagner 400 Lviv 250
Utrecht 240;
param livingCost := Talinn 20 Warsaw 23 Pompei 20 Stavagner 40 Lviv 20
Utrecht 25;
```

Here we have taken approximate flight cost and living cost values. Since I shall be visiting next year, so these costs will need to be updated. But since the model file will be general, the data can easily be changed. Living cost shows the cost of living per day. And also the cities are written in desired visiting order.

We obtained following output:

```
ampl: model personal.mod;
LOQO 7.03: optimal solution (11 QP iterations, 11 evaluations)
primal objective 25.69999983
  dual objective 25.70000004
d [*] :=
     Lviv  6.56667
   Pompei  6.56667
Stavagner  2
   Talinn  6.56667
  Utrecht  2
   Warsaw  2
;
```

So, I should spend 6.57 days in Lviv, Pompei and Talinn and 2 days each in Stavagner, Utrecht and Warsaw to maximize the number of days spent. The problem can be better solved by interger programming because we can be interested in days as integers. But regardless the model provides a specific solution to our problem. However, the data file can easily be updated for more cities and more information. Another way to solve this problem could be implementing Travelling Salesman Problem with some constraints on tours.

5 Data Analysis

We'll try to optimize parameters for linear regression using Iris data. Iris data was obtained from UCI Machine Learning Repository `https://archive.ics.uci.edu/ml/datasets/Iris`. The data has four features and a class. In our problem, we used first feature (Sepal length) and class (Species) to predict another feature (Sepal width). The names of species were converted into three numeric categories: 1 for iris-setosa, 2 for iris-versicolor and 3 for iris-virginica. First column is sepal length, second column is species and third column is sepal width.

We used regularization techniques for our optimization: Optimization without any regularization, with L1 norm and with L2 norm. Lambda chosen has impact for L1 and L2 regularization. The mathematical model and results are present below.

The dataset is stored in a csv file that is read from .dat file.

AMPL Code is defined as follows and written in file personalmodel.mod:

AMPL Code:

```
reset;
option solver loqo;
option loqo_options 'iterlim 30 dual', solver loqo;

param n>=1;
param x1{1..n};
param x2{1..n};
param y{1..n};
var theta1;
var theta2;
var intercept;
param lambda=3;

minimize f: sum{i in 1..n} (theta1*x1[i] + theta2*x2[i] + intercept - y[i])^2;

data personaldata.dat;
solve;
display theta1;
display theta2;
display intercept;
```

The data file was defined as follows:

```
data;
param n:=140;
read {i in 1..n} (x1[i], y[i], x2[i]) < irisdata.csv;
```

Without regularization we obtained following output:

```
LOQO 7.03: iterlim 30
```

```
dual
LOQO 7.03: optimal solution (2 QP iterations, 3 evaluations)
primal objective 18.3192542
  dual objective 18.3192542
theta1 = 0.310614

theta2 = -0.498327

intercept = 2.21521
```

We also performed the task using L1 and L2 regularization by adding a dual $\lambda * (\mid \sum(x_i) \mid)$ and $\lambda * \sum(x_i^2)$.

The AMPL code for L2 regularization is defined as follows:

```
reset;
option solver loqo;
option loqo_options 'iterlim 30 dual', solver loqo;

param n>=1;
param x1{1..n};
param x2{1..n};
param y{1..n};
var theta1;
var theta2;
var intercept;
param lambda=3; # Can be changed

minimize f1: sum{i in 1..n} (theta1*x1[i] + theta2*x2[i] + intercept - y[i])^2
+ lambda*(theta1^2+theta2^2);

data personaldata.dat;
solve;
display theta1;
display theta2;
display intercept;
```

Output was as follows:

```
ampl: model personalmodell1.mod;
LOQO 7.03: iterlim 30
dual
LOQO 7.03: optimal solution (2 QP iterations, 3 evaluations)
```

```
primal objective 19.22561855
  dual objective 19.22561855
theta1 = 0.261348

theta2 = -0.44337

intercept = 2.39497
```

For L1 regularization, the model can be updated by replacing objective function in above codes to -

```
minimize f2: sum{i in 1..n} (theta1*x1[i] + theta2*x2[i] + intercept - y[i])^2
+ lambda*abs(theta1 + theta2);
```

The selection of parameters were tested in R and provided results as expected from this data.

The model can be applied to any linear regression problem by updating the personal.dat file and updating features in model file accordingly if there are more than two features (*e.g.* adding *var* $theta3$ and param $x3$ if there is a third feature. So this implementation can be general.