Linear regression of the outputs of a device and optimizing the linear model using linear programming

Dristy Parveg
Mohsin Khan
Finland

March 22, 2017

1 Linear Modeling and optimization:

It has been observed that the device has two positive outputs $y_1, y_2 \in \mathbb{R}$. We have learnt that y_1 is a function of 5 positive variables $x_1, x_2, x_3, x_4, x_5 \in \mathbb{R}$ and we name this function F. We also have learnt that y_2 is a function of 3 variables x_1, x_2, x_3 and we name this function G. However, we don't exactly know the definitions of F and G. Nonetheless we can sample these two functions at different input points through experiments as discussed in the previous sections. We have results of 18 experiments. We are interested to optimize the functions F and G. We are interested to optimize the functions in few alternative ways. The candidates are as follows:

1. $Maximize\ F - G$: Finding values o_1, o_2, o_3, o_4, o_5 so that $F(o_1, o_2, o_3, o_4, o_5) - G(o_1, o_2, o_3)$ becomes as small as possible.

2. $Minimize\ |\frac{F}{G} - d|$: Finding the values o_1, o_2, o_3, o_4, o_5 so that $\frac{F(o_1,o_2,o_3,o_4,o_5)}{G(o_1,o_2,o_3)}$ becomes as close as possible to a certain constant d

3. $Minimize\ |F - d_1| + |G - d_2|$: Finding the values o_1, o_2, o_3, o_4, o_5 so that $F(o_1, o_2, o_3, o_4, o_5)$ and $G(o_1, o_2, o_3)$ goes as close as possible to constant values d_1 and d_2 respectively.

Let $f_i : \mathbb{R} \rightarrow \mathbb{R}$ be a function that takes the variable x as input while the other input variables x_j for $j \in \{1, ...5\} \backslash i$ of F are kept at constant values i.e., $f_i(x|a_1, ..., a_{i-1}, x, a_{i+1}, ..., a_5) = F(a_1, ..., a_{i-1}, x, a_{i+1}, ..., a_5)$. We define the function $g_i : \mathbb{R} \rightarrow \mathbb{R}$ in the same way so that $g_i(x|a_1, ..., a_{i-1}, x, a_{i+1}, ..., a_5) = G(a_1, ..., a_{i-1}, x, a_{i+1}, ..., a_5)$. From the result of the experiments, presented in previous section, it is evident that f_i and g_i are almost, if not exactly, linear functions.

We are interested to find models $f : \mathbb{R}^5 \rightarrow \mathbb{R}$, $g : \mathbb{R}^3 \rightarrow \mathbb{R}$ where $f(x_1, x_2, x_3, x_4, x_5) = G(x_1, x_2, x_3, x_4, x_5)$ and $g(x_1, x_2, x_3) = G(x_1, x_2, x_3)$. We make an educated guess that both F and G are linear. We do not provide an analytical argument to justify the linearity of F and G. Instead, we use linear regression on the 18 experimental results and obtain the linear models

$$\begin{aligned} f(x_1, x_2, x_3, x_4, x_5) &= -7.2667 \cdot x_1 + 0.3724 \cdot x_2 + 29.1759 \cdot x_3 - 3.2784 \cdot x_4 \\ &\quad +1.8231 \cdot x_5 + 78.2264 \\ g(x_1, x_2, x_3) &= -5.477 \cdot x_1 + 4.640 \cdot x_2 + 4.290 \cdot x_3 + 24.343 \end{aligned}$$

For an input $(x_1, x_2, x_3, x_4, x_5)$, the error percentages of the models f and g are $\frac{|f(x_1,x_2,x_3,x_4,x_5) - F(x_1,x_2,x_3,x_4,x_5)|}{F(x_1,x_2,x_3,x_4,x_5)} \times 100$ and $\frac{|g(x_1,x_2,x_3) - G(x_1,x_2,x_3)|}{G(x_1,x_2,x_3)} \times 100$ respectively. We compute the mean, variance, minimum and maximum error percentages of the models f and g over all the 18 inputs and present in the below table

		Error percentage		
Model	Mean	Standard Deviation	Min	Max
f	2.209766	2.480924	0.5398369	9.525519
g	1.002286	1.329765	0.008706377	5.115356

We reckon, these error percentages indicate that the underlying functions F and G are fairly linear and the obtained models f and g are fairly good linear approximations of F and G. Consequently, we optimize F and G by optimzing f and g respectively. As F and G are not not exactly linear and our obtained models f and g are linear approximations, the optimal values obtained from the models f and g will have some errors comparing to the optimal values of F and G. We expect these errors to be similar to the errors of the models.

As both f and g are linear and all the inputs of both of the functions are positive, the optimization problems are most likely to be representable as a linear programming problem. The linear programming problem would involve 5 variables. A linear programming problem with 5 variables is a small instance of a linear programming problem. There exist efficient algorithms, namely, simplex, ellipsoid, etc. for solving linear programming problems. In practice, a linear programming problem can easily be solved by any standard linear programming solver like lp_solve.

Now we present the optimization problems discussed above as s linear programming problems in the following subsections:

1.1 *Maximize* $f - g$:

We need to find values o_1, o_2, o_3, o_4, o_5 so that $f(o_1, o_2, o_3, o_4, o_5) - g(o_1, o_2, o_3)$ becomes as small as possible, i.e., we need to maximize

$$\begin{aligned}
& f(x_1, x_2, x_3, x_4, x_5) - g(x_1, x_2, x_3) \\
= \;& (-7.2667 \cdot x_1 + 0.3724 \cdot x_2 + 29.1759 \cdot x_3 - 3.2784 \cdot x_4 \\
& +1.8231 \cdot x_5 + 78.2264) \\
& -(-5.477 \cdot x_1 + 4.640 \cdot x_2 + 4.290 \cdot x_3 + 24.343) \\
= \;& -1.7897 \cdot x_1 - 3.9176 \cdot x_2 + 24.8859 \cdot x_3 \\
& -3.2784 \cdot x_4 + 1.8231 \cdot x_5 + 53.8834
\end{aligned}$$

Consequently the linear programming problem looks as follows

$$\text{max:} \quad -1.7897 \cdot x_1 - 3.9176 \cdot x_2 + 24.8859 \cdot x_3 - 3.2784 \cdot x_4 + 1.8231 \cdot x_5 + 53.8834;$$

$$\begin{aligned}
&x_1 <= 5;\\
&x_2 <= 5;\\
&x_3 <= 3;\\
&x_4 <= 15;\\
&x_5 <= 21;\\
&x_1 >= 1;\\
&x_2 >= 1;\\
&x_3 >= 1;\\
&x_4 >= 2;\\
&x_5 >= 5;
\end{aligned}$$

The above linear programming problem can be solved by saving the above problem in a file named *model.lp* and run the command *lp_solve model.lp*. We get the maximum value of the objective function to be: 154.562 at the input $x_1 = 1, x_2 = 1, x_3 = 3, x_4 = 2, x_5 = 21$. Feeding this input to f and g we find

$$\begin{aligned}
f(1,1,3,2,21) &= -7.2667 + 0.3724 + 29.1759 \times 3 - 3.2784 \times 2\\
&\quad +1.8231 \cdot 21 + 78.2264\\
&= 190.5881\\
g(1,1,3) &= -5.477 + 4.640 + 4.290 * 3 + 24.343 = 36.376
\end{aligned}$$

We can find the value of $F(1,1,3,2,21)$ and $G(1,1,3)$ by running an experiment to see how good the optimization is.

1.2 *Minimize* $|\frac{f}{g} - d|$:

We need to find the values o_1, o_2, o_3, o_4, o_5 so that $\frac{f(o_1,o_2,o_3,o_4,o_5)}{g(o_1,o_2,o_3)}$ becomes as close as possible to a certain constant d. Let us set $d = 6$. Then we need to minimize $\frac{f(x_1,x_2,x_3,x_4,x_5)}{g(o_1,o_2,o_3)} - 6$. In other words we need to minimize

$$\begin{aligned}
& |f(x_1, x_2, x_3, x_4, x_5) - 6g(x_1, x_2, x_3)| \\
= & |(-7.2667 \cdot x_1 + 0.3724 \cdot x_2 + 29.1759 \cdot x_3 - 3.2784 \cdot x_4 \\
& +1.8231 \cdot x_5 + 78.2264) \\
& -6(-5.477 \cdot x_1 + 4.640 \cdot x_2 + 4.290 \cdot x_3 + 24.343)| \\
= & |25.5953 \cdot x_1 - 27.4675 \cdot x_2 + 3.4358 \cdot x_3 - 3.2784 \cdot x_4 \\
& +1.8231 \cdot x_5 - 67.8316|
\end{aligned}$$

As a result the linear programming problem looks as follows: the linear program is written in model1.lp file the solution is: $x_1 = 3.4891, x_2 = 1, x_3 = 1, x_4 = 2, x_5 = 5$. The optimized value of the linear program is 0. And
$f(x_1 = 3.4891, x_2 = 1, x_3 = 1, x_4 = 2, x_5 = 5) = 84.97916$
$g(x_1 = 3.4891, x_2 = 1, x_3 = 1) = 14.1632$

We can find the value of $F(3.4891, 1, 1, 2, 5)$ and $G(3.4891, 1, 1)$ by running an experiment to see how good the optimization is.

1.3 *Minimize* $|f - d_1| + |g - d_2|$:

Finding the values o_1, o_2, o_3, o_4, o_5 so that $f(o_1, o_2, o_3, o_4, o_5)$ and $g(o_1, o_2, o_3)$ goes as close as possible to constant values d_1 and d_2 respectively. Let us set $d_1 = 120$ and $d_2 = 24$. We need to minimize

$$\begin{aligned}
& |f(x_1, x_2, x_3, x_4, x_5) - 120| + |g(x_1, x_2, x_3) - 24| \\
= & |-7.2667 \cdot x_1 + 0.3724 \cdot x_2 + 29.1759 \cdot x_3 - 3.2784 \cdot x_4 + 1.8231 \cdot x_5 + 78.2264 - 120| \\
& +|-5.477 \cdot x_1 + 4.640 \cdot x_2 + 4.290 \cdot x_3 + 24.343 - 24| \\
= & |-7.2667 \cdot x_1 + 0.3724 \cdot x_2 + 29.1759 \cdot x_3 - 3.2784 \cdot x_4 \\
& +1.8231 \cdot x_5 - 41.776| + |-5.477 \cdot x_1 + 4.640 \cdot x_2 + 4.290 \cdot x_3 + 0.343|
\end{aligned}$$

So, the linear program looks as follows: the linear program is in the file model2.lp The optimized value of the linear program is 0. The value of the variables are $X = 0, Y = 0, x_1 = 2.42593, x_2 = 1, x_3 = 1.93562, x_4 = 2, x_5 = 5$. And
$f(x_1 = 2.42593, x_2 = 1, x_3 = 1.93562, x_4 = 2, x_5 = 5) = 120.0025$
$g(x_1 = 2.42593, x_2 = 1, x_3 = 1.93562) = 23.99999$

We can find the value of $F(2.42593, 1, 1.93562, 2, 5)$ and $G(2.42593, 1, 1.93562)$ by running an experiment to see how good the optimization is.