The first objective is to find all features on which the cost depends from the train set. Then, we find which machine learning algorithm will apply depending of these features.

Let's define the features used in the training set. The description is given in the codebook.

Feature	Variable
x_1	tube_assembly_id
x_2	supplier
x_3	quote_date
x_4	annual_usage
x_5	min_order_quantity
x_6	bracket_pricing
x_7	quantity
x_8	cost

Let's denote $h_\beta(x)$ our cost heuristic function of a tube assembly given by a supplier where β is our learning parameters. Since the output of this analysis is known (the cost a supplier will quote for a given tube assembly), then we will use a supervised algorithm.

Per the codebook, x_6 determines on which features the cost depends. We will use 2 classes to base our cost estimation.

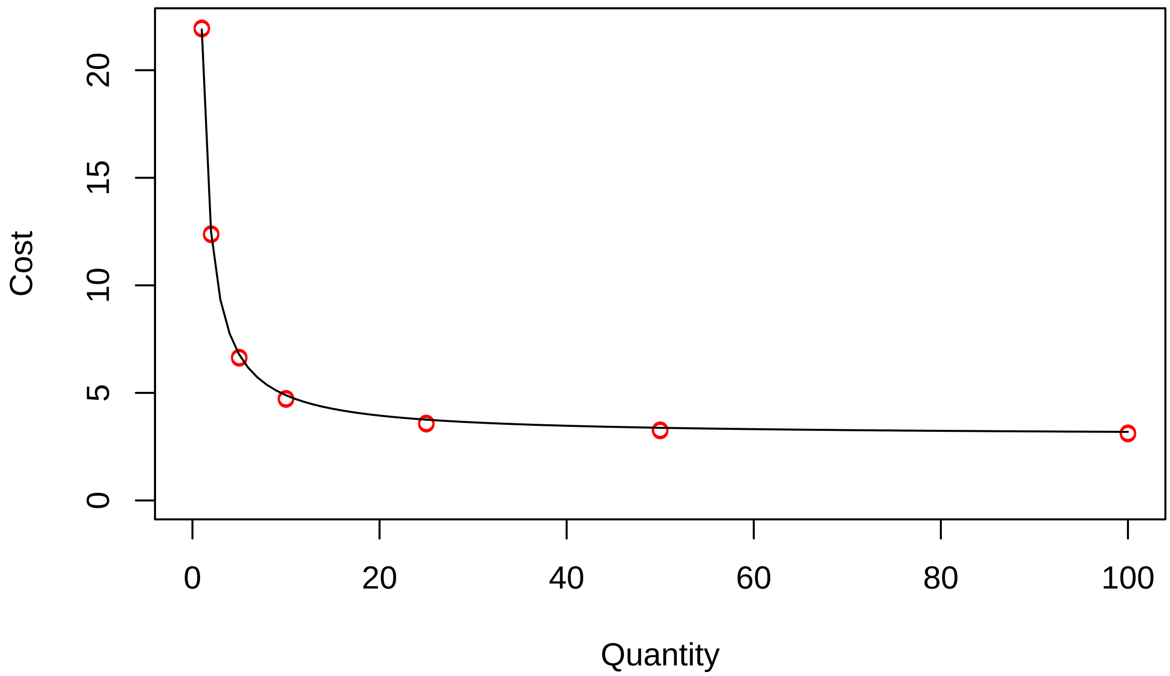
1. Bracket pricing ($x_6 = 1$ (Yes)) where the function $C(x)$ depends of x_7 amoung other features.
2. Non-bracket pricing ($x_6 = 0$ (No)) where the function $C(x)$ depends of x_5 and x_7.

Test with Supplier S-0066

In this section, we simplify the dataset where we use the bracket pricing by the supplier S-0066. Our goal is to modelize the costs of the first tube assemblies and to find patterns with these data.

We start with the two first tube assemblies (TA-00002 and TA-00004).

fkTubeAssembly	supplierID	diameter	wallThickness	length	volume	numberOfBends	quantity	cost
2	S-0066	6.35	0.71	137	1723.487	8	1	21.905933
2	S-0066	6.35	0.71	137	1723.487	8	250	2.999060
2	S-0066	6.35	0.71	137	1723.487	8	100	3.082521
2	S-0066	6.35	0.71	137	1723.487	8	50	3.224406
2	S-0066	6.35	0.71	137	1723.487	8	25	3.541561
2	S-0066	6.35	0.71	137	1723.487	8	10	4.687769
2	S-0066	6.35	0.71	137	1723.487	8	5	6.601826
2	S-0066	6.35	0.71	137	1723.487	8	2	12.341214
4	S-0066	6.35	0.71	137	1723.487	9	100	3.149291
4	S-0066	6.35	0.71	137	1723.487	9	50	3.291176
4	S-0066	6.35	0.71	137	1723.487	9	25	3.608331
4	S-0066	6.35	0.71	137	1723.487	9	10	4.754539
4	S-0066	6.35	0.71	137	1723.487	9	5	6.668596
4	S-0066	6.35	0.71	137	1723.487	9	2	12.407983
4	S-0066	6.35	0.71	137	1723.487	9	1	21.972702
4	S-0066	6.35	0.71	137	1723.487	9	250	3.065829




```
##          used (Mb) gc trigger (Mb) max used (Mb)
## Ncells 519294 27.8     940480 50.3   607767 32.5
## Vcells 708033  5.5    1308461 10.0   875906  6.7
```

From the plot, we see that the curve representing the points is clearly an hyperbola of equation $h_\beta(x_7) = \frac{\beta_0}{x_7} + \beta_1$ where $x_7 \geq 1$, β_1 is the cost at the last level of purchase based on quantity and $\beta_0 = h_\beta(1) - \beta_1$. This equation will avoid high bias or variance.

From the table, we can see that for a fixed quantity, the cost varies. This implies that the cost depends also on tube assembly features. By comparing the tube TA-00002 and TA-00004 from the table `TubeAssembly`, we see that the only feature that varies is the number of bends. The tube TA-00002 is made with 8 bends and the tube TA-00004 is made with 9 bends. Thus, we can find the variation of the cost per bend for the supplier S-0066. From the table, we have $21.9727024365 - 21.9059330191 = 0.066769417$ for the minimal quantity (which is 1). Therefore, the equation for the cost of bends is given by $C_B(n) = 0.066769417n$.

Let the total volume estimation of a tube assembly be denoted by V_T. The volume is function of the length, the wall thickness and the diameter of the tube and its formula is $V_T = \pi LW(D - W)$, where W is the wall thickness, D the outside diameter and L the developped length of the tube.