

Preprocessing real steps

1. Reduction: Attribute selection
 2. Cleaning: Missing values handling
 3. Cleaning: Noise data handling
 4. Cleaning: Data inconsistency handling
 5. Transformation: Datatype Conversion (to int)
 6. Transformation: Binning
 7. Transformation: Scaling
 8. Transformation: Sampling
-

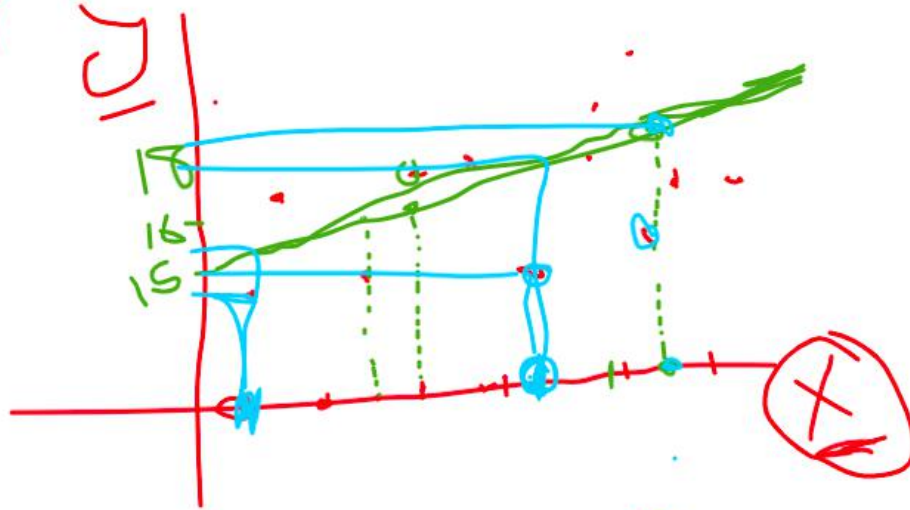
Check Scales

- 2 sex - student's sex (binary: 'F' - female or 'M' - male)
4 address - student's home address type (binary: 'U' - urban or 'R' - rural)
5 famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
6 Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
14 studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
15 failures - number of past class failures (numeric: n if $1 \leq n < 3$, else 4)
16 schoolsup - extra educational support (binary: yes or no)
17 famsup - family educational support (binary: yes or no)
18 paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19 activities - extra-curricular activities (binary: yes or no)
21 higher - wants to take higher education (binary: yes or no)
22 internet - Internet access at home (binary: yes or no)
23 romantic - with a romantic relationship (binary: yes or no)
24 famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25 freetime - free time after school (numeric: from 1 - very low to 5 - very high)
26 goout - going out with friends (numeric: from 1 - very low to 5 - very high)
27 Dalc - weekday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28 Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29 health - current health status (numeric: from 1 - very bad to 5 - very good)
30 absences - number of school absences (numeric: from 0 to 93)

- 2 sex - 0, 1
4 address - 0, 1
5 famsize - 0, 1
6 Pstatus - 0, 1
14 studytime - 1, 2, 3, 4
15 failures - 1, 2, 3, 4
16 schoolsup - 0, 1
17 famsup - 0, 1
18 paid - 0, 1
19 activities - 0, 1
21 higher - 0, 1
22 internet - 0, 1
23 romantic - 0, 1
24 famrel - 1, 2, 3, 4, 5
25 freetime - 1, 2, 3, 4, 5
26 goout - 1, 2, 3, 4, 5
27 Dalc - 1, 2, 3, 4, 5
28 Walc - 1, 2, 3, 4, 5
29 health - 1, 2, 3, 4, 5
30 absences - 0 - 93
-

Description of Regression (Math view)

G3



$$\begin{pmatrix} 15 - 18 \\ 15 - 16 \\ 18 - 16 \end{pmatrix} \sqrt{RMSE}$$

