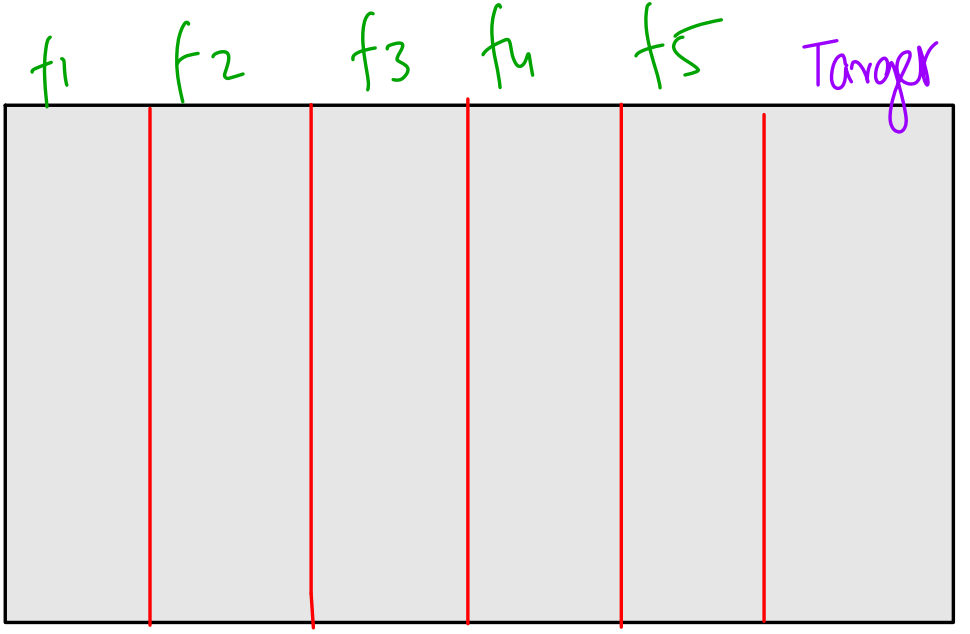
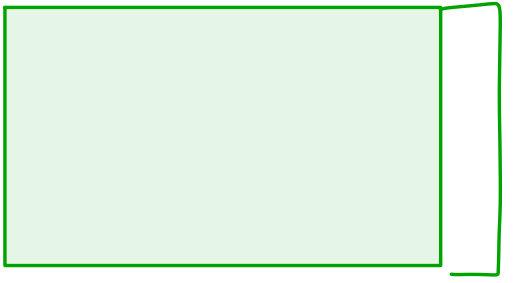
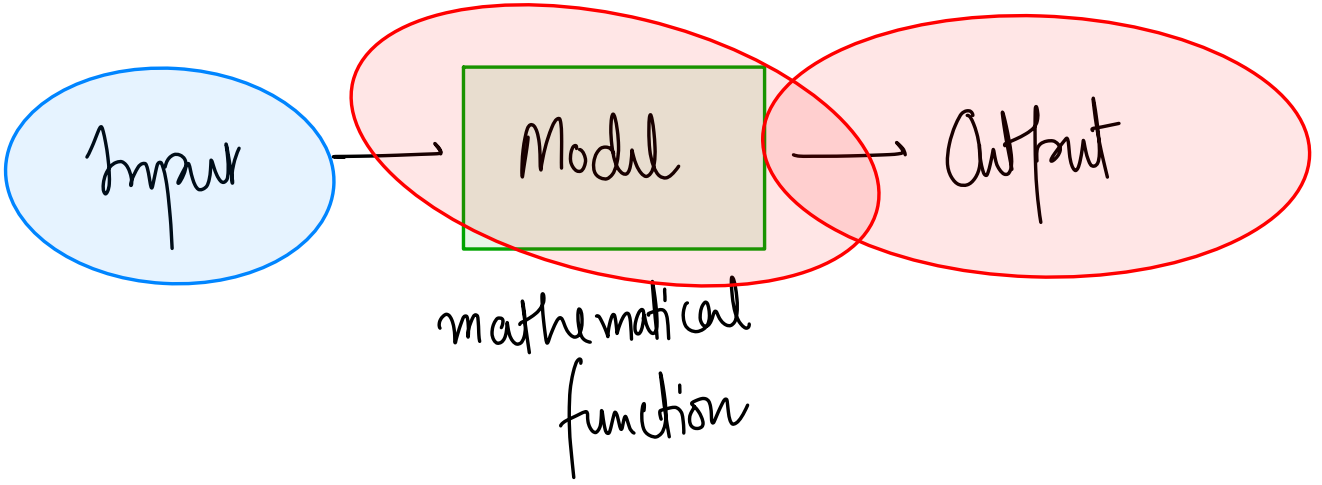
FEATURE ENGINEERING-3

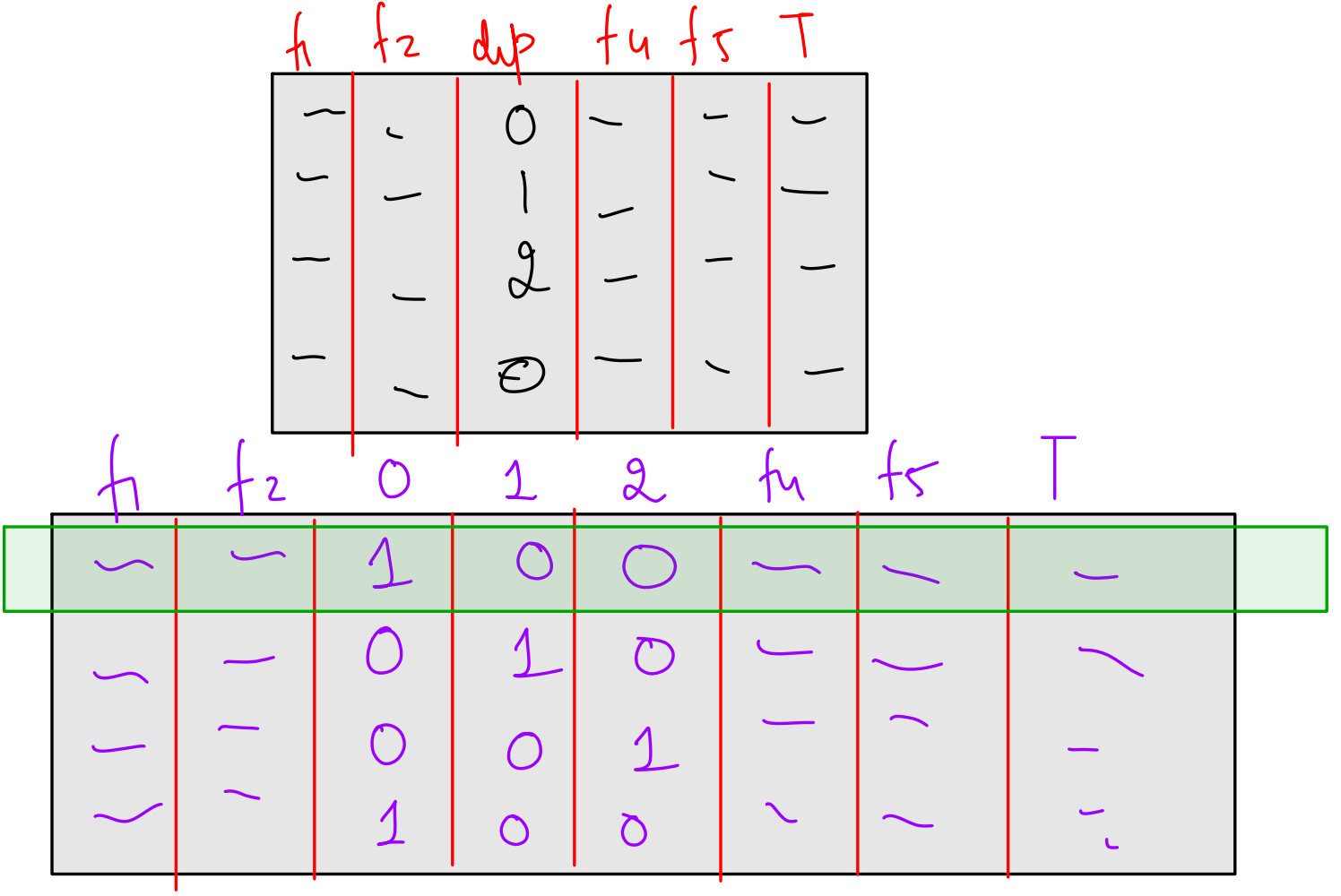

f1
f2
f3
f4
f5
Target
Applicant income
Total household income




Input
Model
Output
mathematical
function

① OHE → One hot Encoding

	Cat	DOG	Elephant	Rat
Cat →	1	0	0	0
dog	0	1	0	0
elephant	0	0	1	0
Cat	1	0	0	0
dog	0	1	0	0
Rat	0	0	0	1


f1 f2 dip f4 f5 T
0
1
2
3
f1 f2 0 1 2 f4 f5 T
1 0 0
0 1 0
0 0 1
1 0 0

② label Encoding

Nominal		ordinal
Cat	→	0
DOG	→	1
elephant	→	2
Cat	→	0
dog	→	1
Rat	→	3

ordinality → Order

Categorical

label Encoding works for features where no. of categories = 2

③ Target Encoding

HIT/FLOP
1 0

Drama	0
Rom	0
Thrill	1
Action	0
Rom	1
Action	1
Thrillu	1
Rom	1

column	target	P[Target/Category]
Drama	0	0
Rom	0	0.66
Rom	1	0.66
Rom	1	0.66
Thriller	1	1
Thriller	1	1
Action	0	0.5
Action	1	0.5