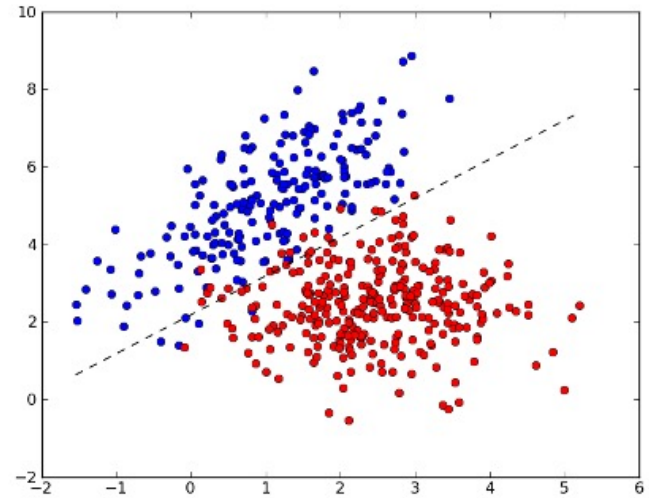
Classification model

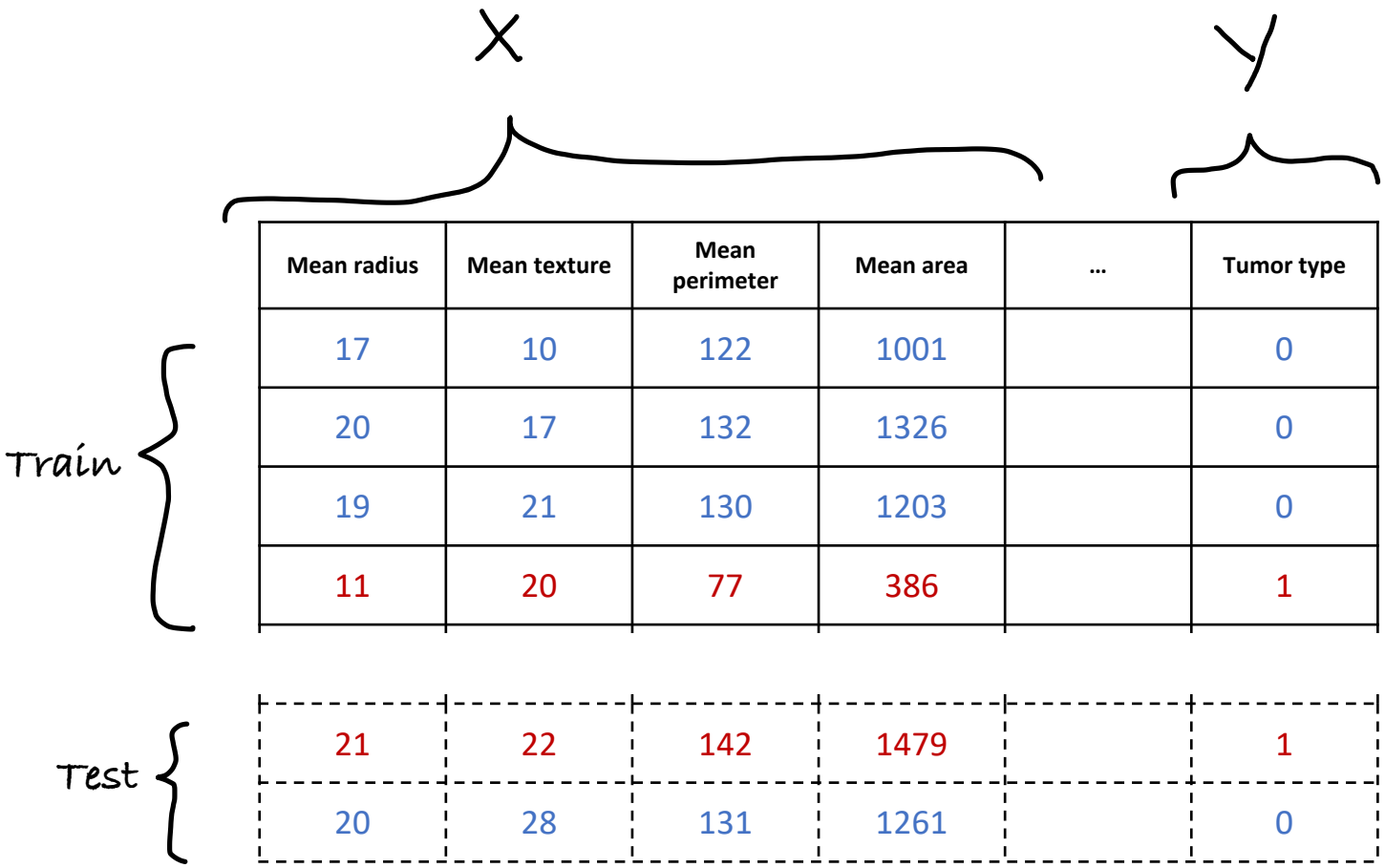
X (Features, Independent var, Input var)

Y (class label, Dependent var, Output var)

N (Rows, Samples)

Mean radius	Mean texture	Mean perimeter	Mean area	...	Tumor type
17	10	122	1001		0
20	17	132	1326		0
19	21	130	1203		0
11	20	77	386		1
...	...	...	...		...
21	22	142	1479		1
20	28	131	1261		0

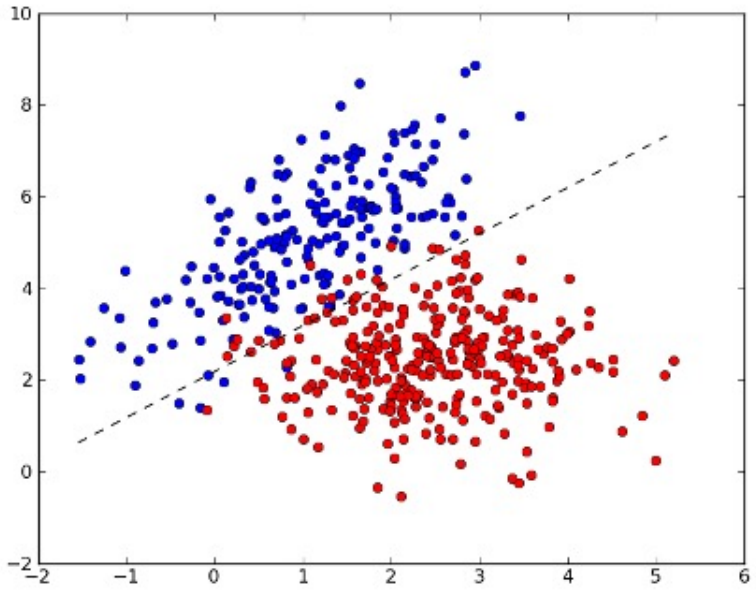




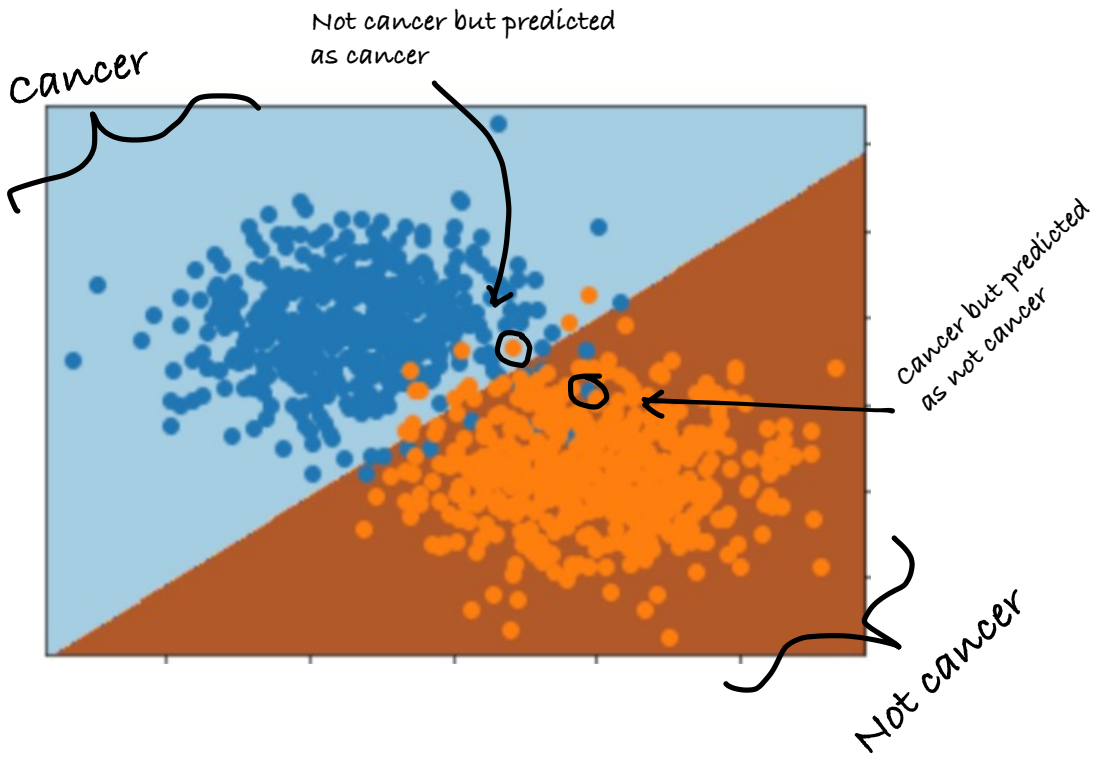
Mean radius	Mean texture	Mean perimeter	Mean area	...	Tumor type
17	10	122	1001		0
20	17	132	1326		0
19	21	130	1203		0
11	20	77	386		1
21	22	142	1479		1
20	28	131	1261		0

	X	y
Train	X_train	y_train
Test	X_test	y_test

Real data

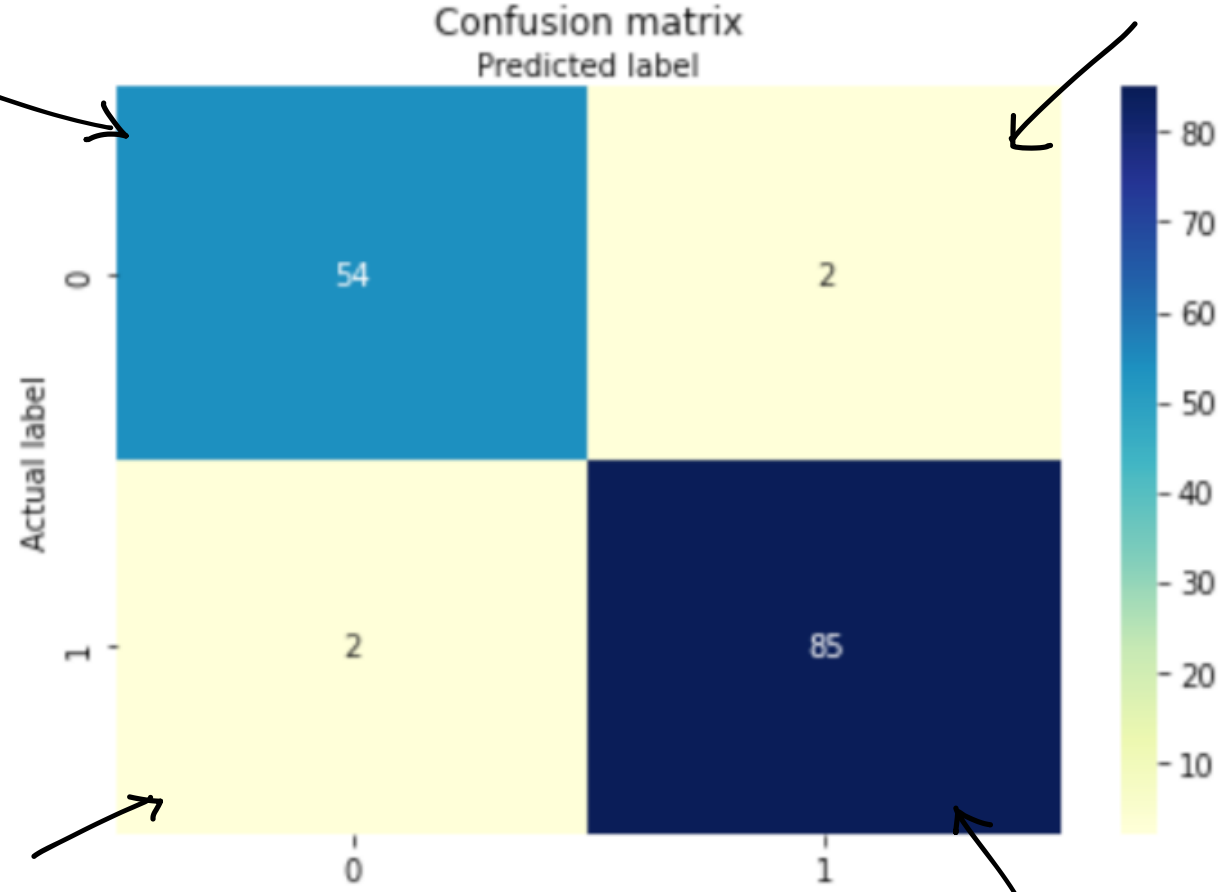


Predictions



Not cancer and predicted as not cancer

Not cancer but predicted as cancer

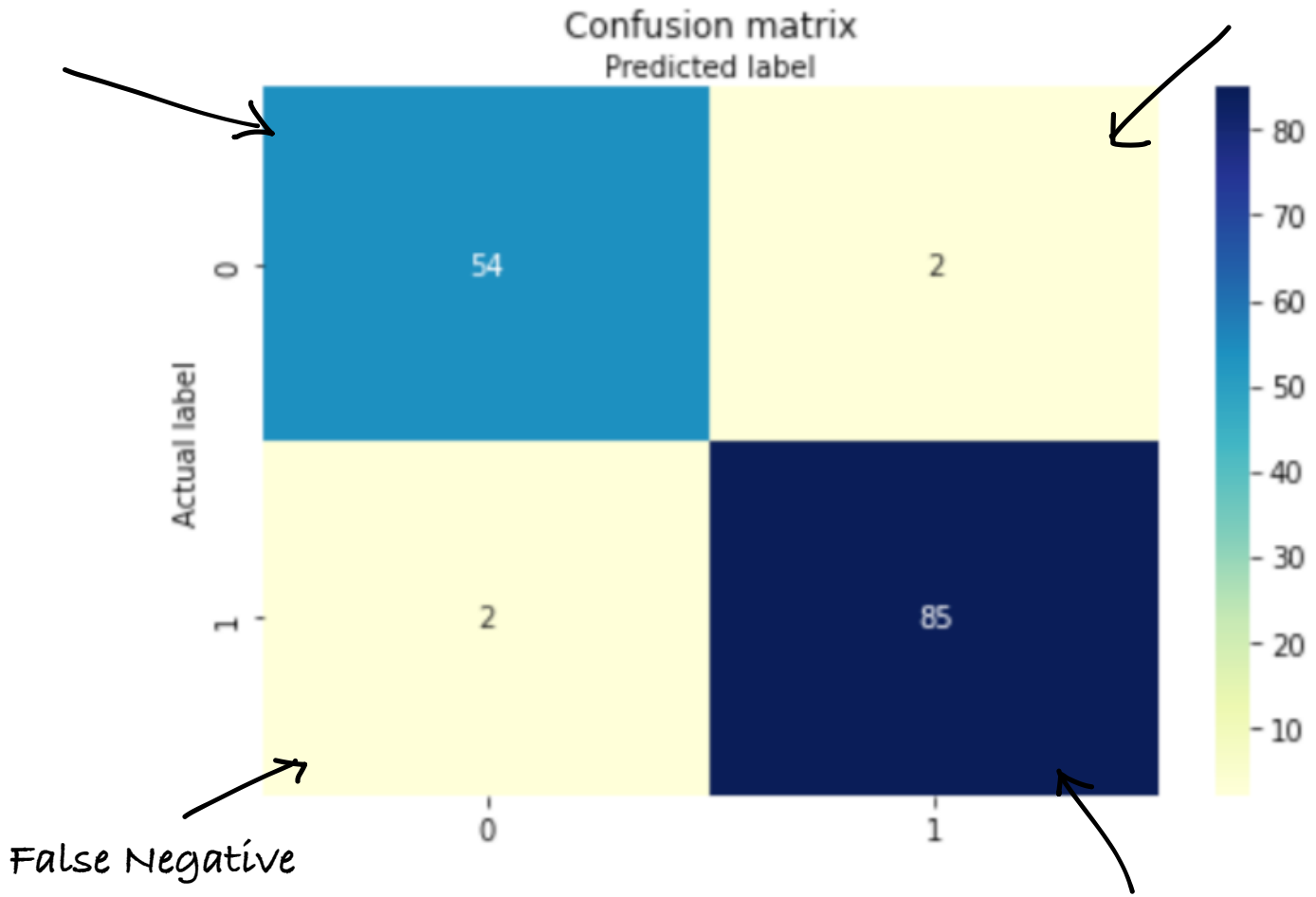


Cancer but predicted as not cancer

Cancer and predicted as cancer

True negative

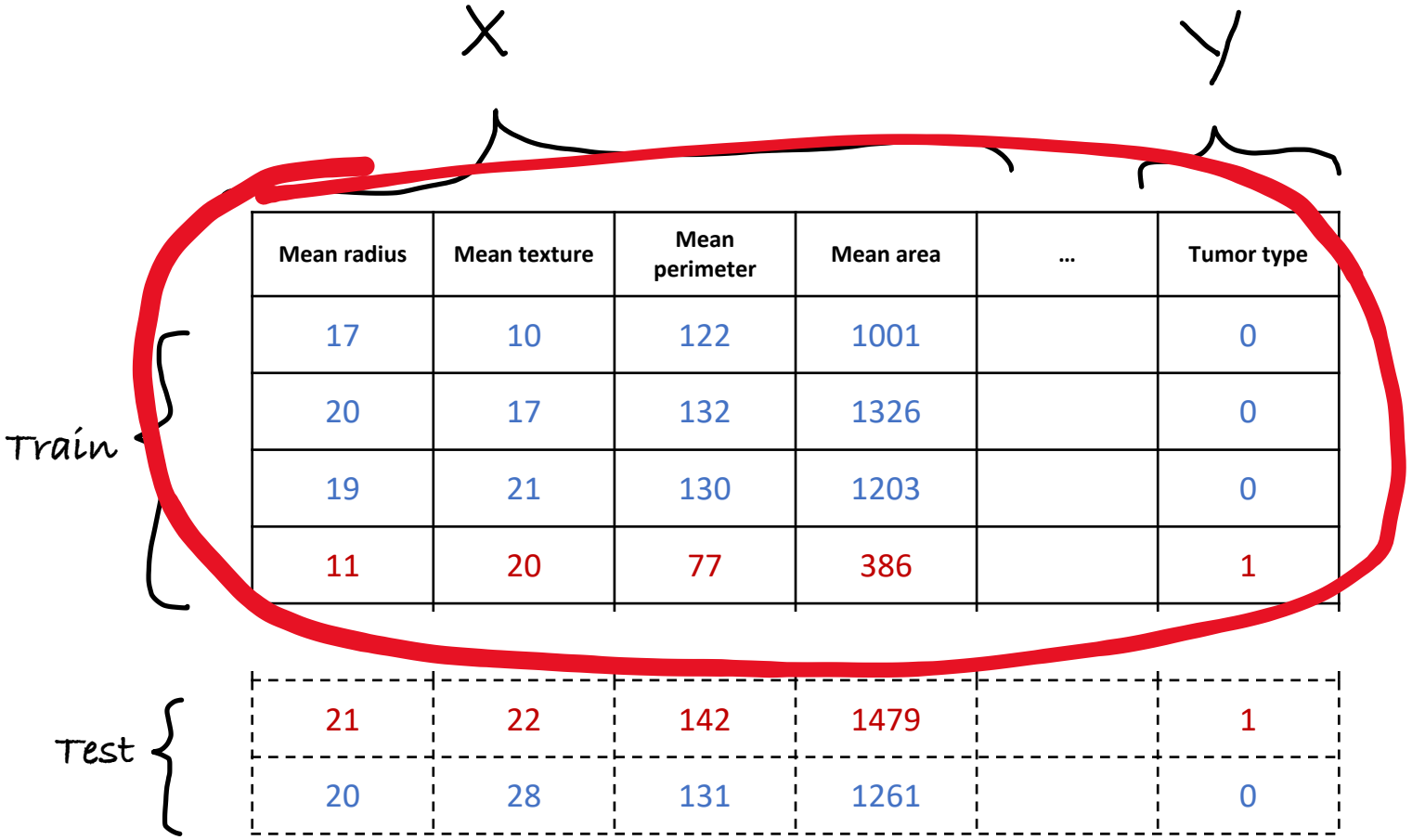
False positive



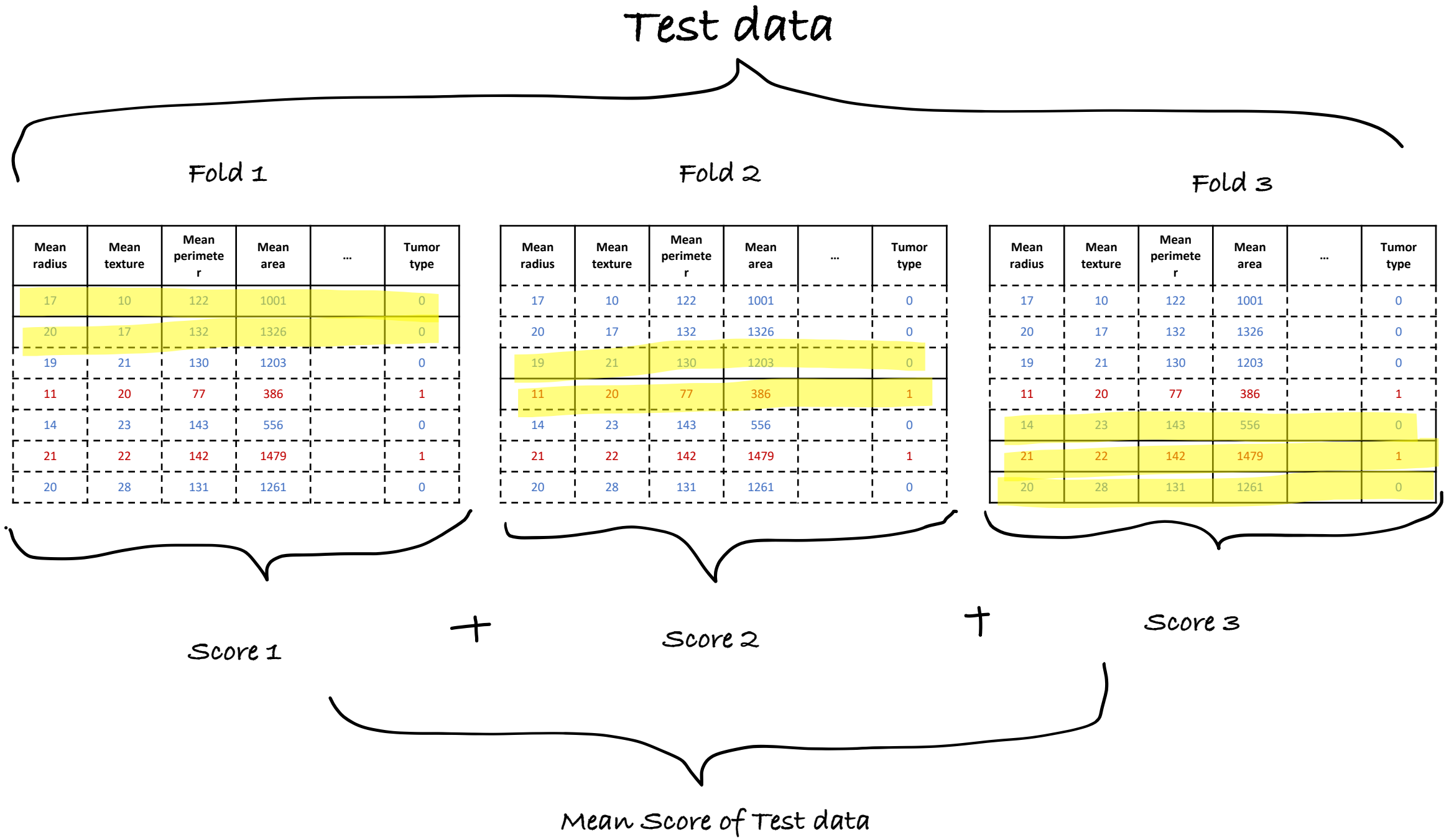
False Negative

True Positive

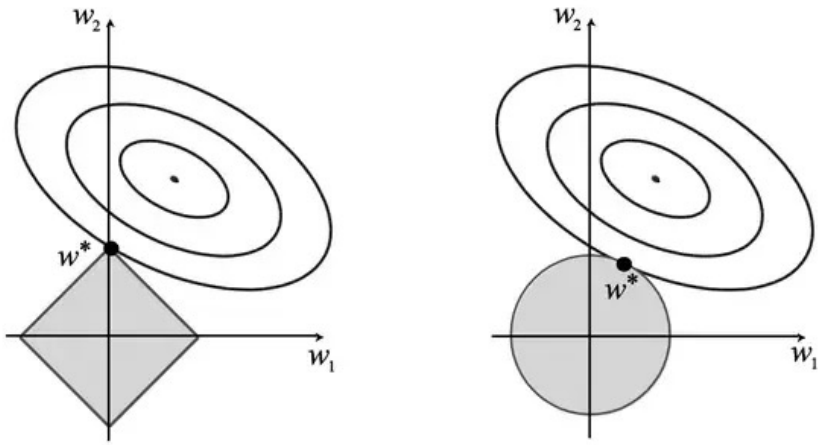
What is CV

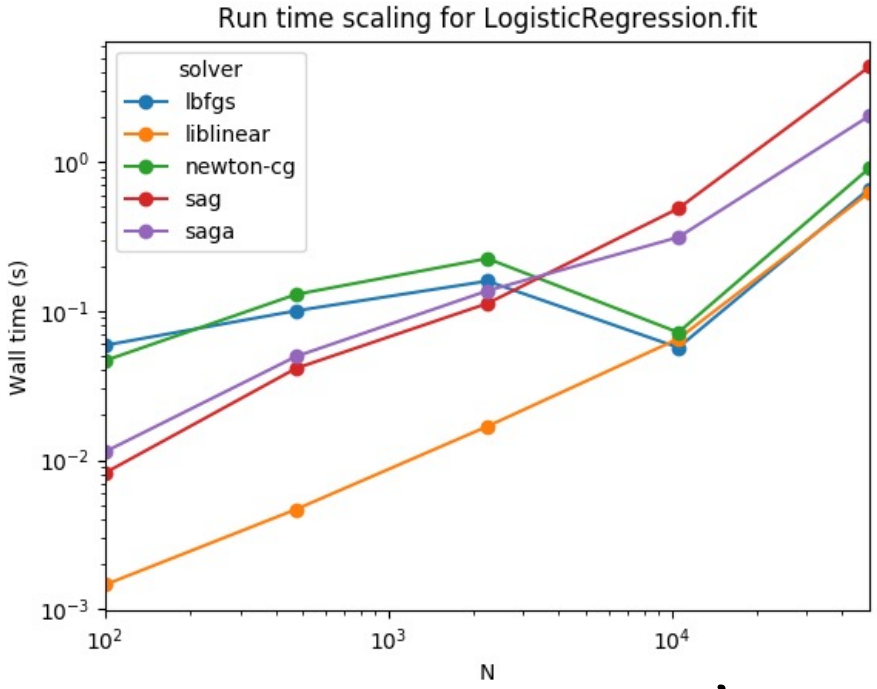


Mean radius	Mean texture	Mean perimeter	Mean area	...	Tumor type
17	10	122	1001		0
20	17	132	1326		0
19	21	130	1203		0
11	20	77	386		1
21	22	142	1479		1
20	28	131	1261		0


Test data
Fold 1
Fold 2
Fold 3
Mean radius
Mean texture
Mean perimeter
Mean area
...
Tumor type
17 10 122 1001 0
20 17 132 1326 0
19 21 130 1203 0
11 20 77 386 1
14 23 143 556 0
21 22 142 1479 1
20 28 131 1261 0
Score 1
+
Score 2
+
Score 3
Mean Score of Test data

What is grid search?





$\times$

$= 2^0$

penalty:

- 'none': no penalty is added;
- 'l2': add a L2 penalty term and it is the default choice;
- 'l1': add a L1 penalty term;
- 'elasticnet': both L1 and L2 penalty terms are added.

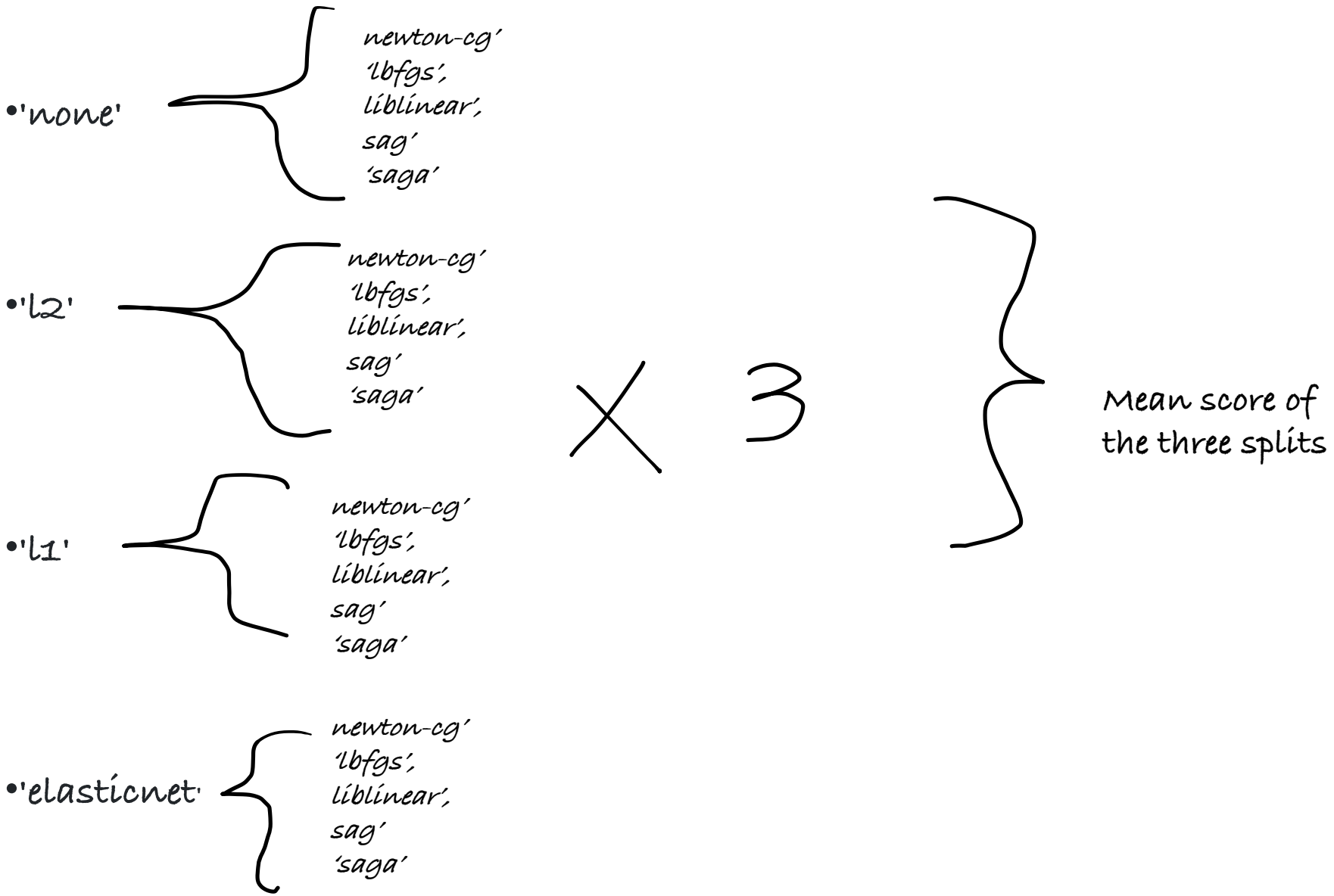
4

Solver

{'newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga'}, default='lbfgs'

- For small datasets, 'liblinear' is a good choice, whereas 'sag' and 'saga' are faster for large ones;
- For multiclass problems, only 'newton-cg', 'sag', 'saga' and 'lbfgs' handle multinomial loss;
- 'liblinear' is limited to one-versus-rest schemes.

5


•'none'
newton-cg'
'lbfgs',
liblinear',
sag'
'saga'
•'l2'
newton-cg'
'lbfgs',
liblinear',
sag'
'saga'
•'l1'
newton-cg'
'lbfgs',
liblinear',
sag'
'saga'
•'elasticnet'
newton-cg'
'lbfgs',
liblinear',
sag'
'saga'
X 3
Mean score of
the three splits