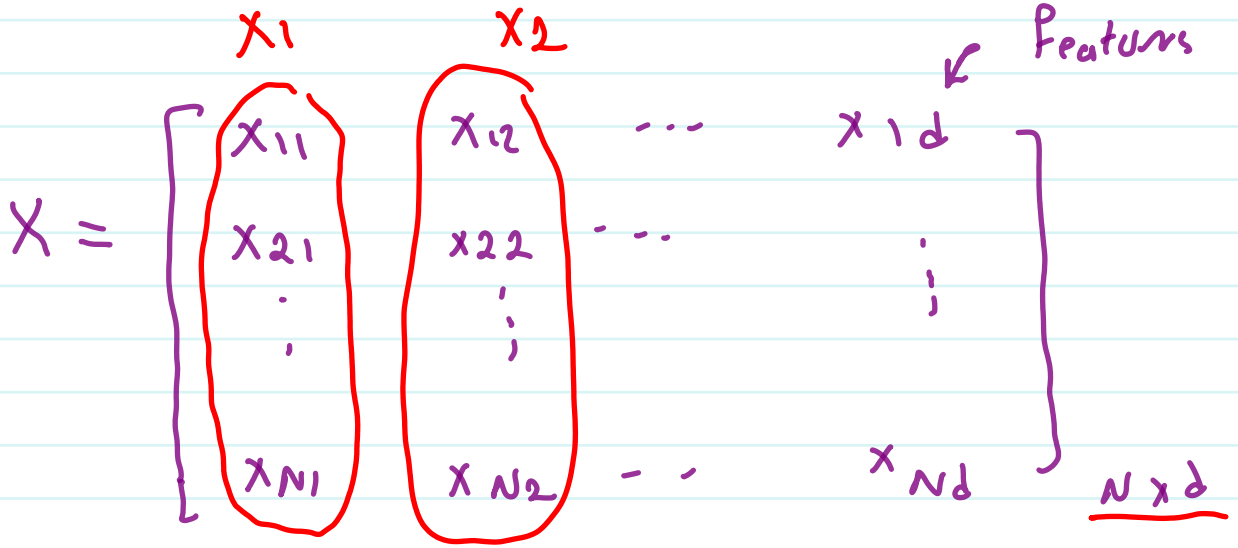
Data Matrix

N # observations

d # features

$X \in \mathbb{R}^{N \times d}$



$x_1 = \lambda x_2 + \mu$

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_j x_d + \epsilon$$

$$\beta^* = [x^T x]^{-1} x^T y$$

$(d \times N)$ $(N \times d)$

$d \times d$

Colinearity

$\frac{2}{0} \rightarrow \frac{2}{0.001} = 2{,}000$

original_raw

	A (salary)	B (age)	C
0	Non	11	Nan
1	1	5	5
2	2	Nan	10
3	3	Nan	11
4	Nan	Non	Nan
5	Nan	8	8

$\frac{4}{0.0001} = 4000$

row- b fill

original_raw

	A	B	C
0	1 ~~Non~~	11 ✓	~~Nan~~ 5
1	1 ✓	5 ✓	5 ✓
2	2 ✓	~~Nan~~ 8	10
3	3	~~Nan~~ 8	11
4	Nan	~~Non~~ 8	~~Nan~~ 8
5	Nan	8	8

col- bfill

original_raw

	A	B	C
0	11 ~~Non~~	11	Nan ✓
1	1	5	5
2	2	10 ~~Nan~~	10
3	3	11 ~~Nan~~	11
4	Nan ✓	Non ✓	Nan ✓
5	8 ~~Nan~~	8	8

row - ffill

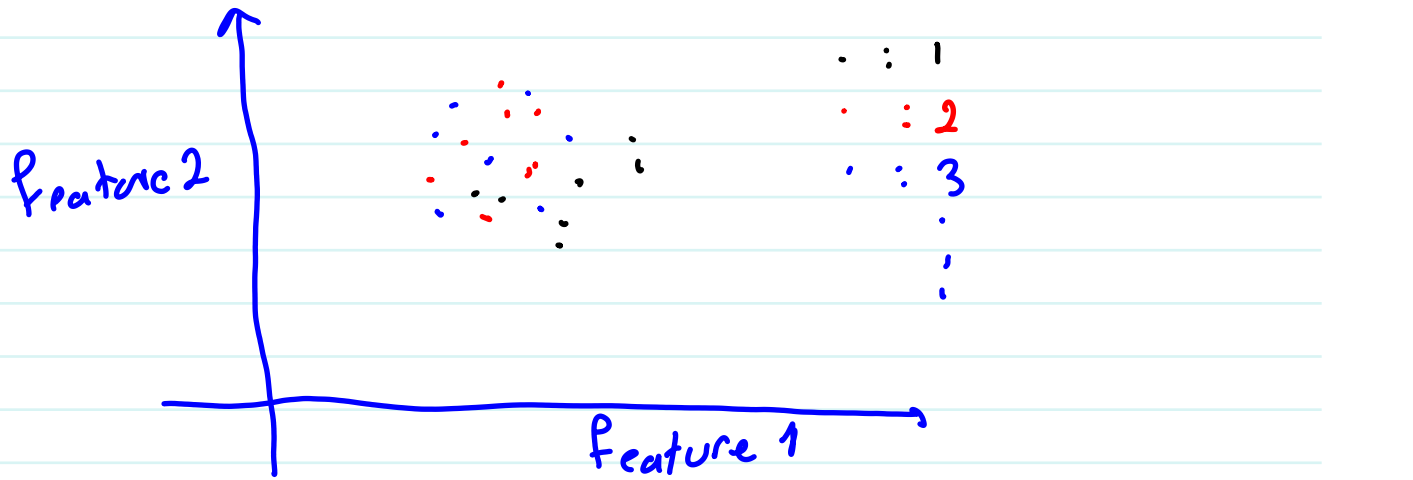
original_raw

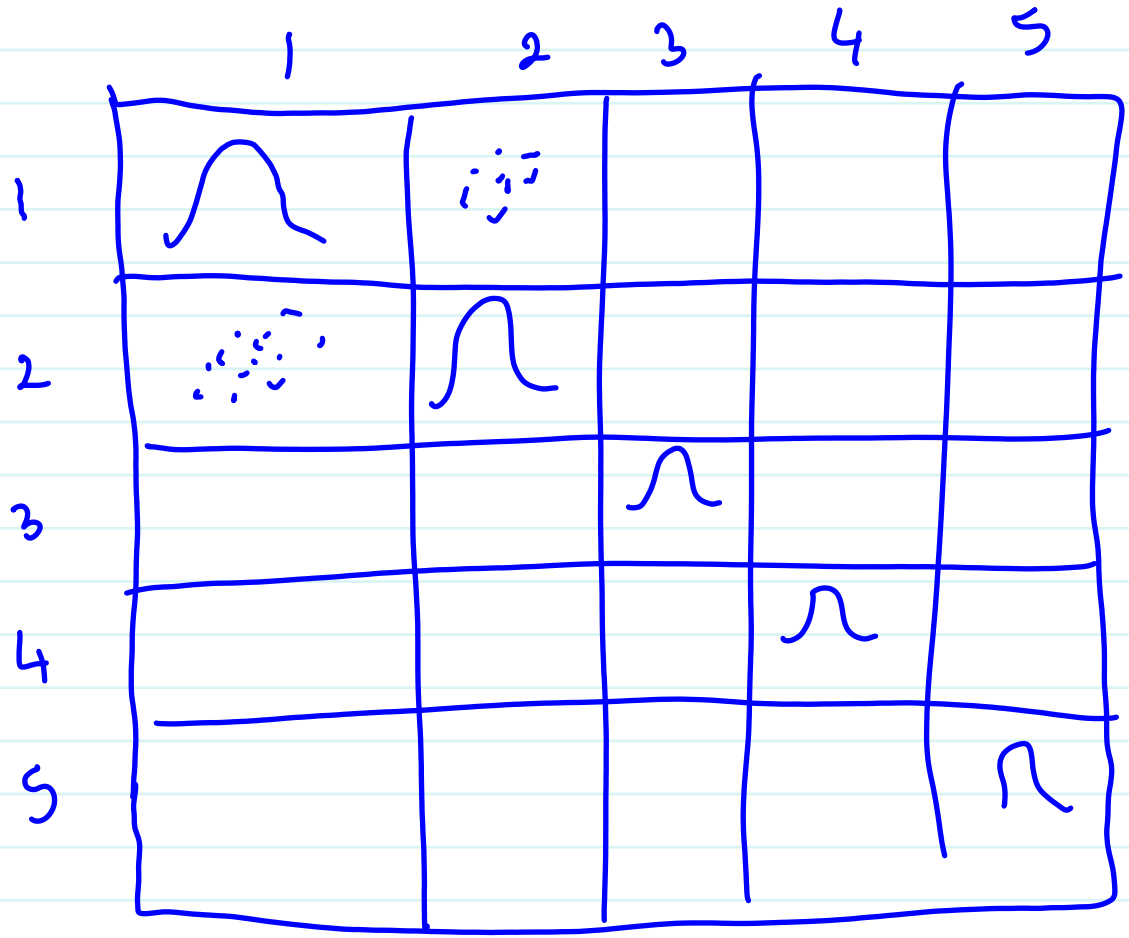
	A	B	C
0	Non	11	Nan
1	1	5	5
2	2	~~Nan~~ 5	10
3	3	~~Nan~~ 5	11
4	3 ~~Nan~~	~~Nan~~ 5	~~Nan~~ 11
5	3 ~~Nan~~	8	8

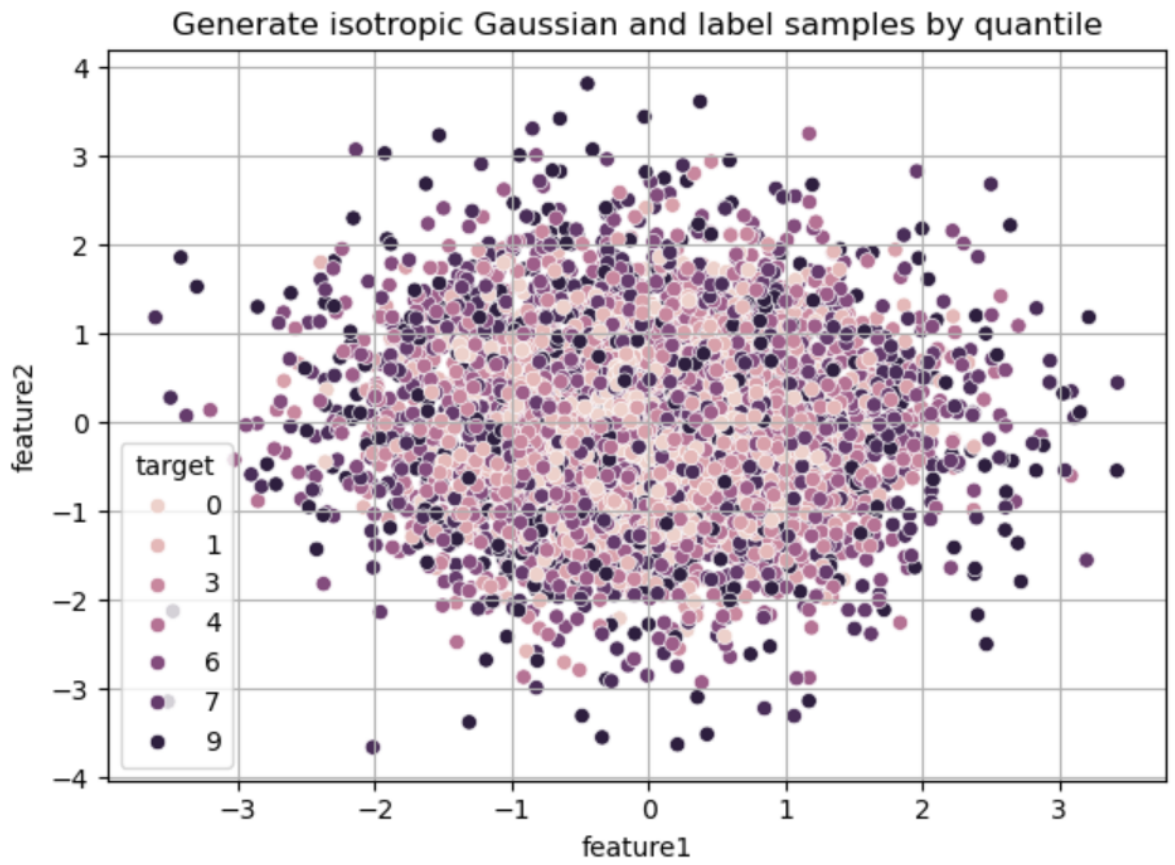
col - ffill

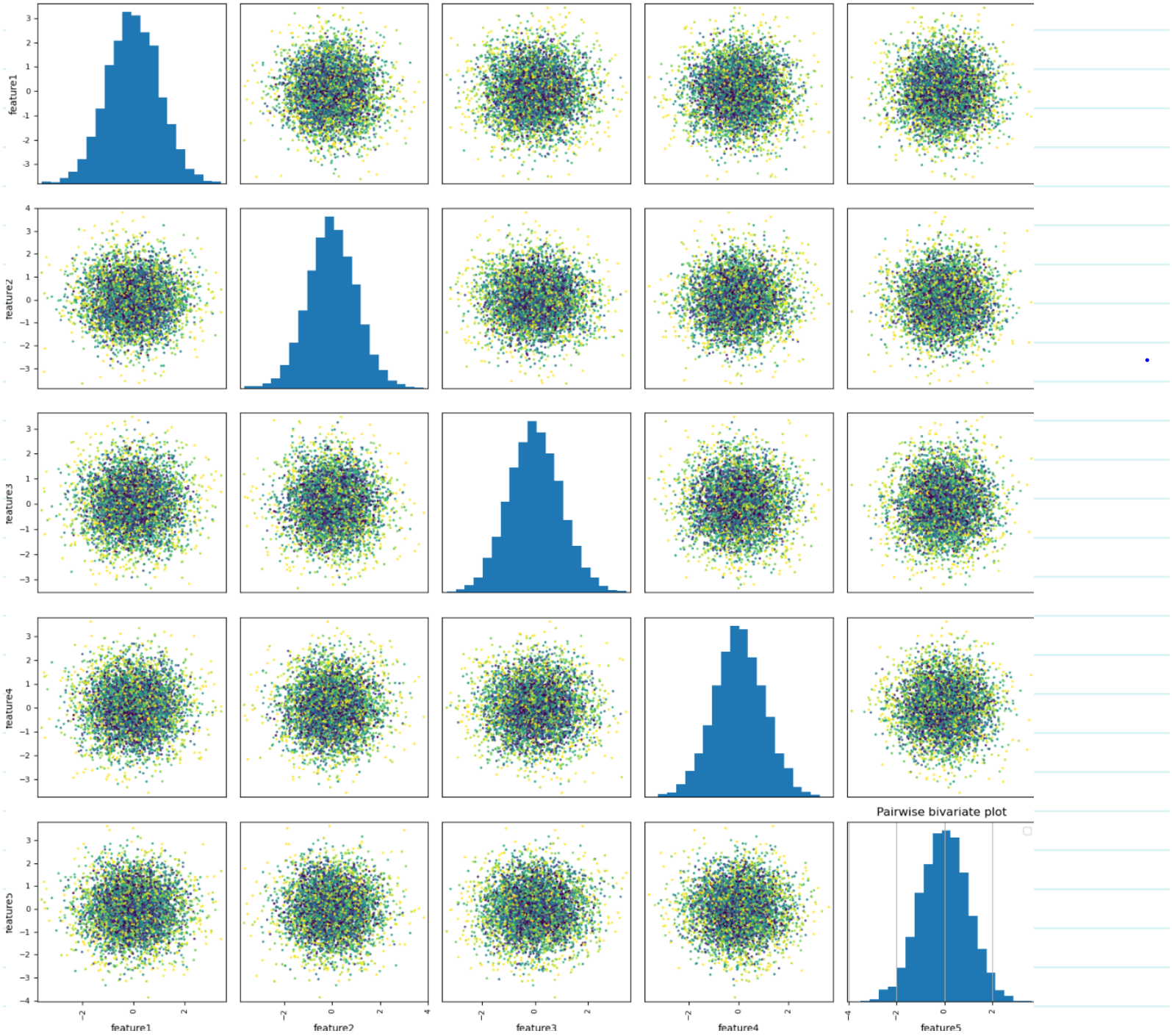
original_raw

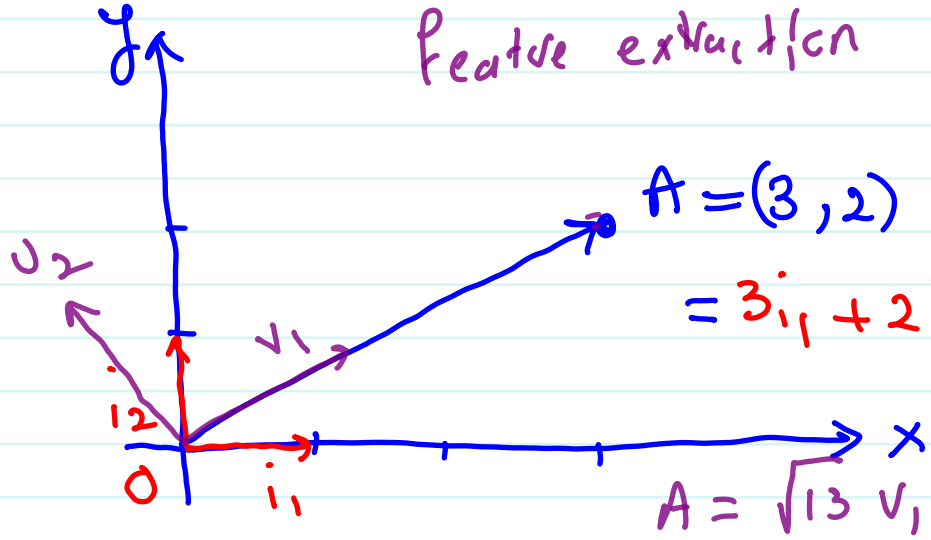
	A	B	C
0	Non	11	~~Nan~~ 11
1	1	5	5
2	2	2 ~~Nan~~	10
3	3	3 ~~Nan~~	11
4	Nan	Non	Nan
5	Nan	8	8




1
2
3
4
5
1
2
3
4
5


Generate isotropic Gaussian and label samples by quantile
feature2
feature1
target
0
1
3
4
6
7
9
4
3
2
1
0
−1
−2
−3
−4
−3
−2
−1
0
1
2
3


Pairwise bivariate plot
feature1
feature2
feature3
feature4
feature5



Feature extraction

S = basis set = {i1, i2}

i1 = (1, 0)

i2 = (0, 1)

S' : new basis = {v1, v2}