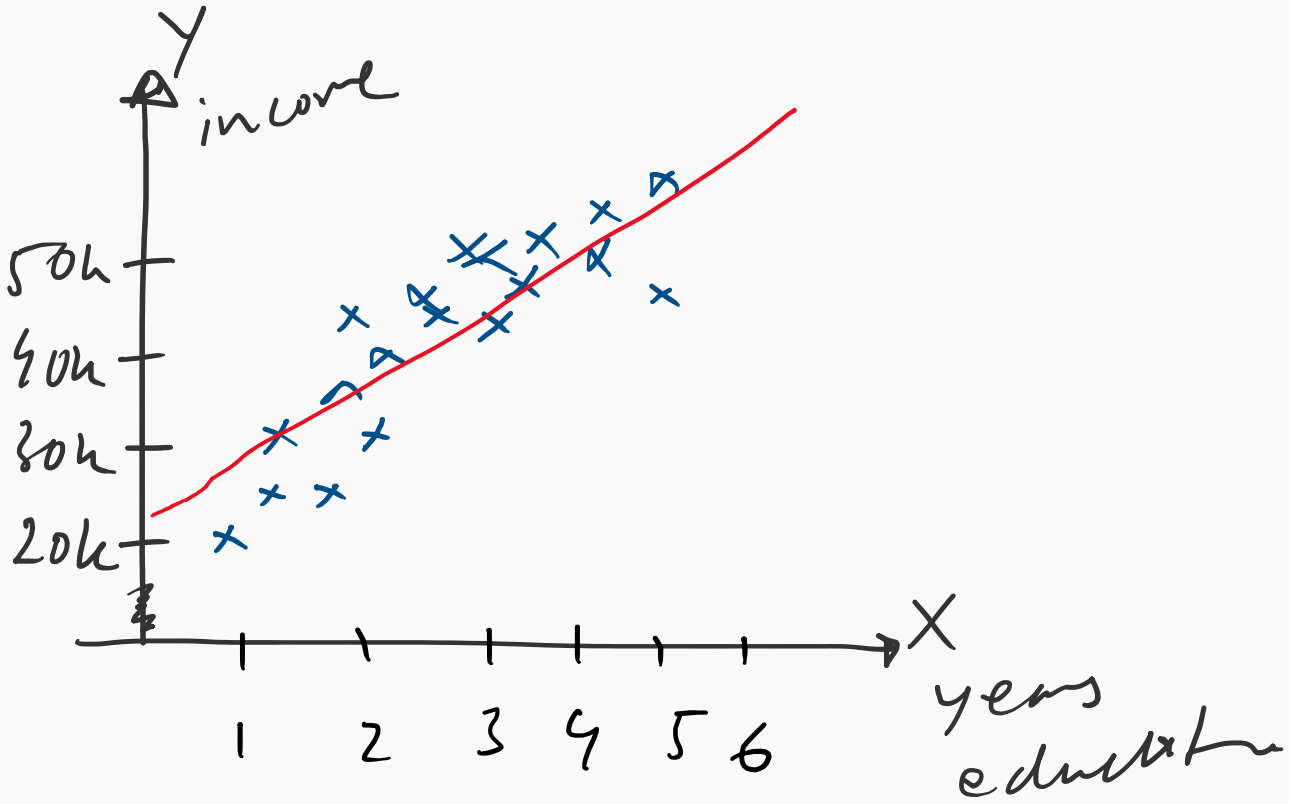
kokchun giang

linear regression



predict income example 1 – how to draw the line?

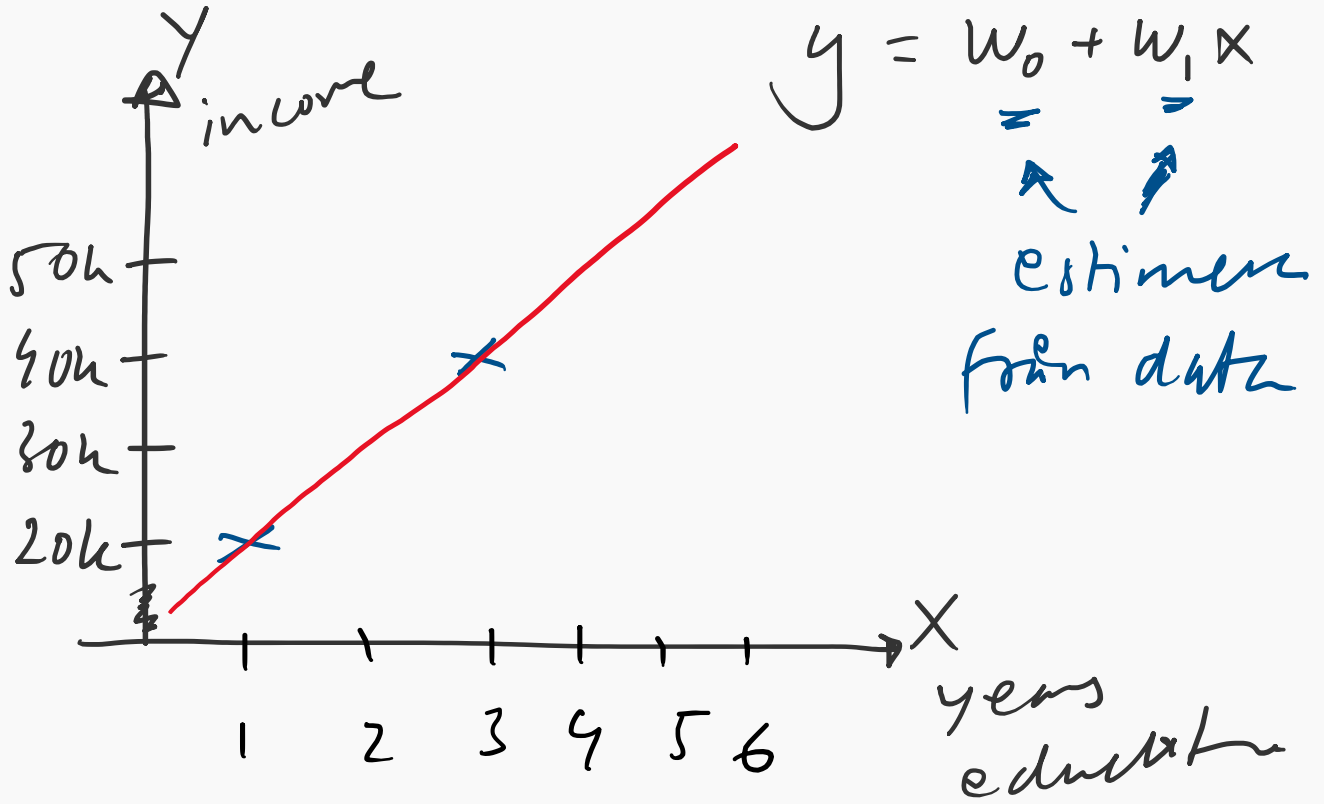
we make a survey to people to ask how long education they have and how much their income is.

collect data

years

income (SEK)

X	Y
1	20k
3	40k



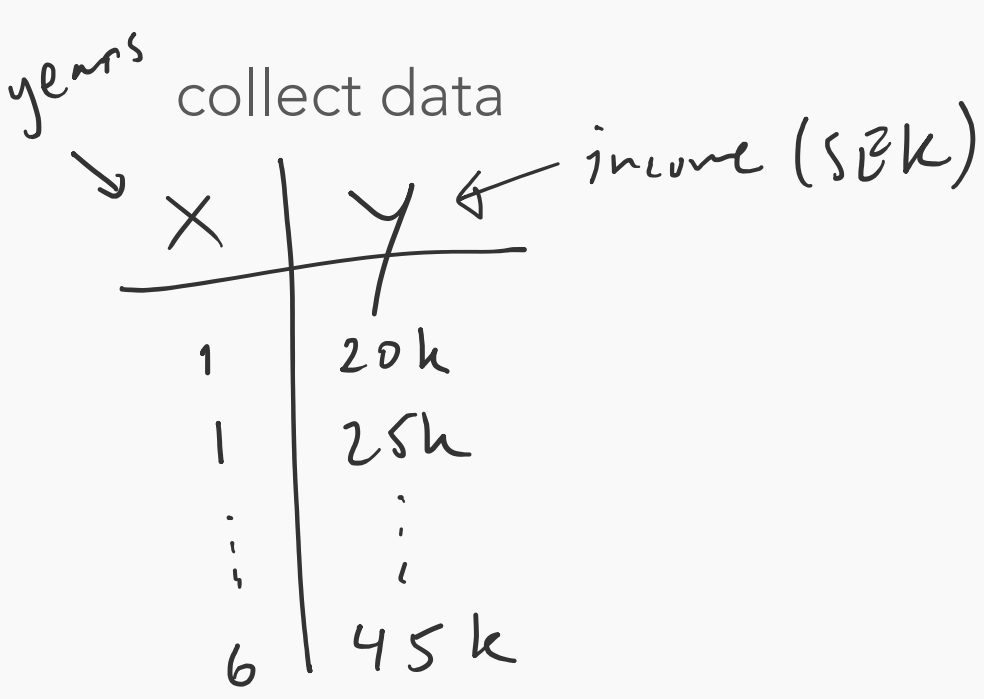
intercept y axis

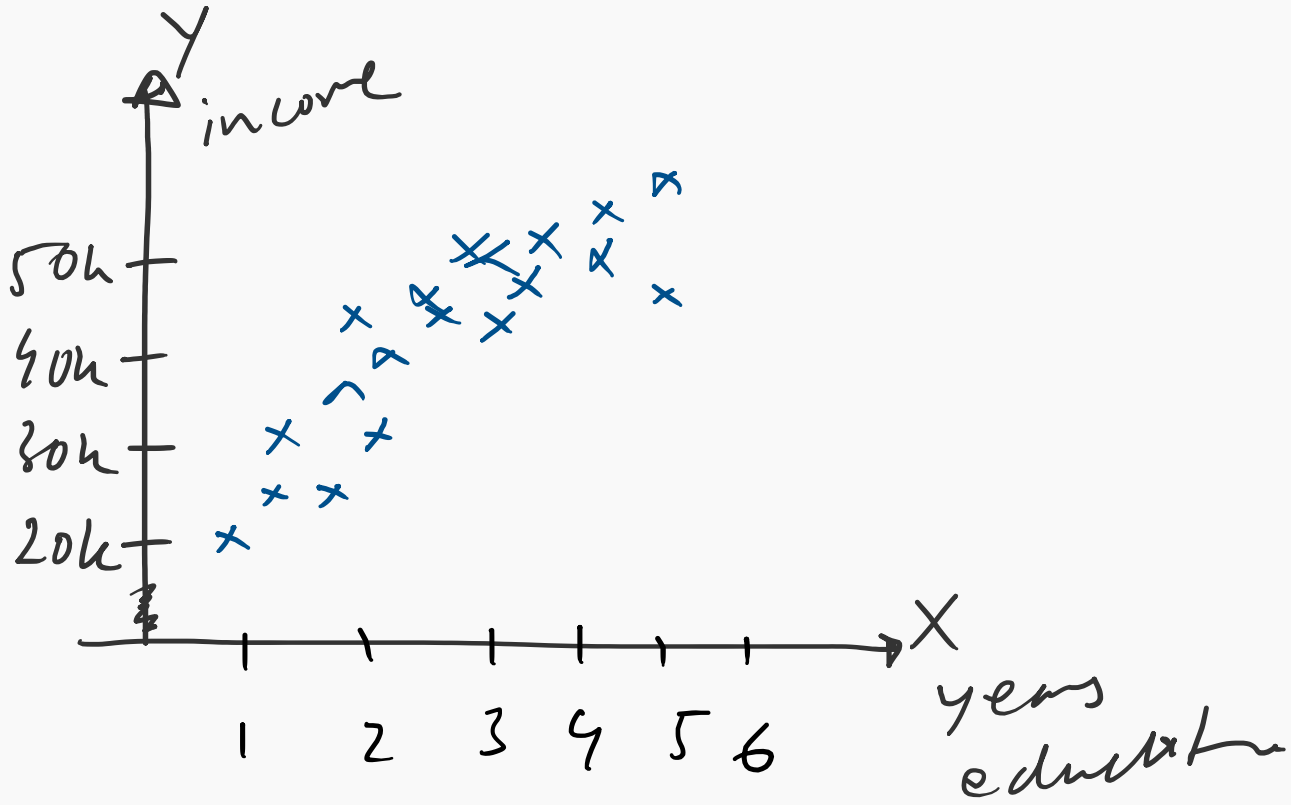
slope

$y = w_0 + w_1 x$

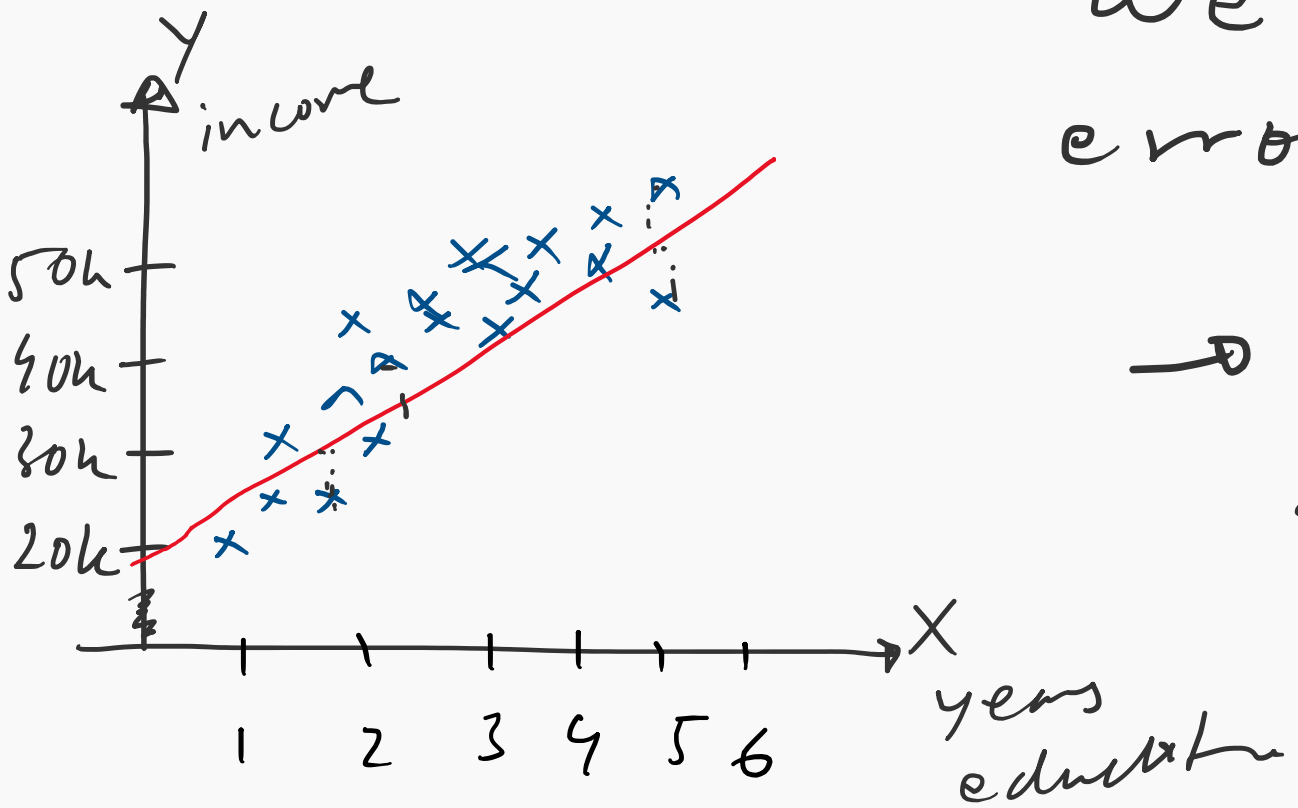
estimates from data

predict income example 2 – how to draw the line now for best fit of the points?





predict income example 3 – pick w_0 and w_1 to **minimize the error**



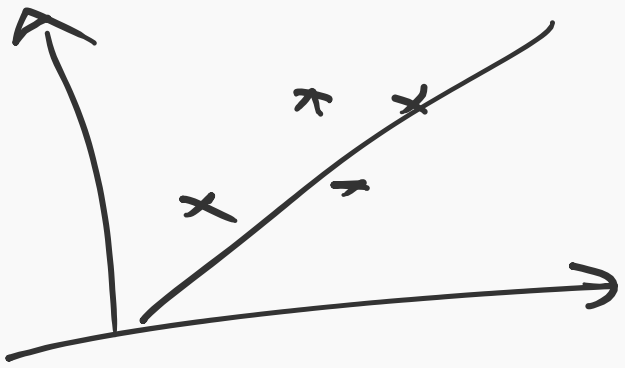
we minimize error in squares

→ residual sum of squares (RSS)

number of **features**

1 feature →

Simple linear regression

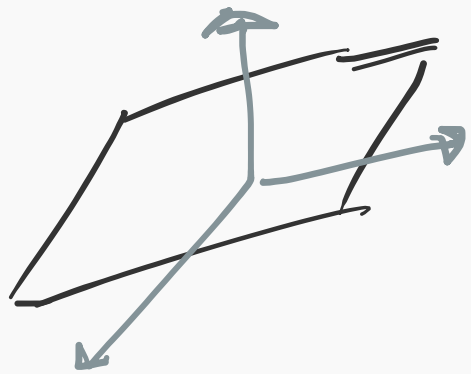


a line

n features

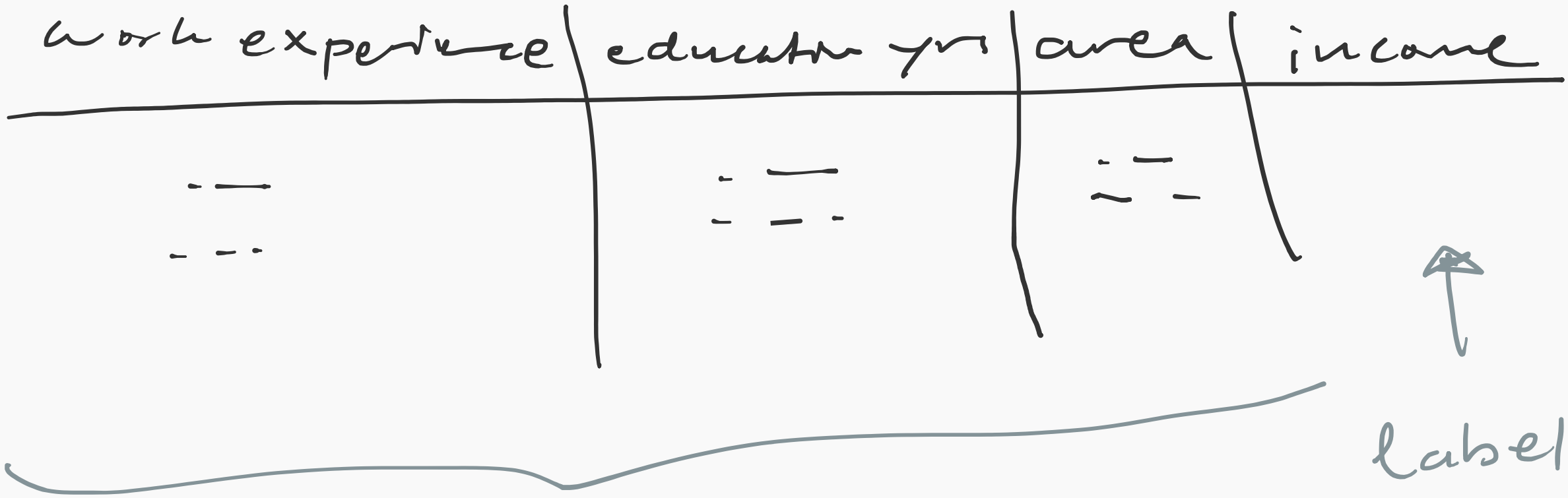
→ multiple linear regression

n dimensional plane

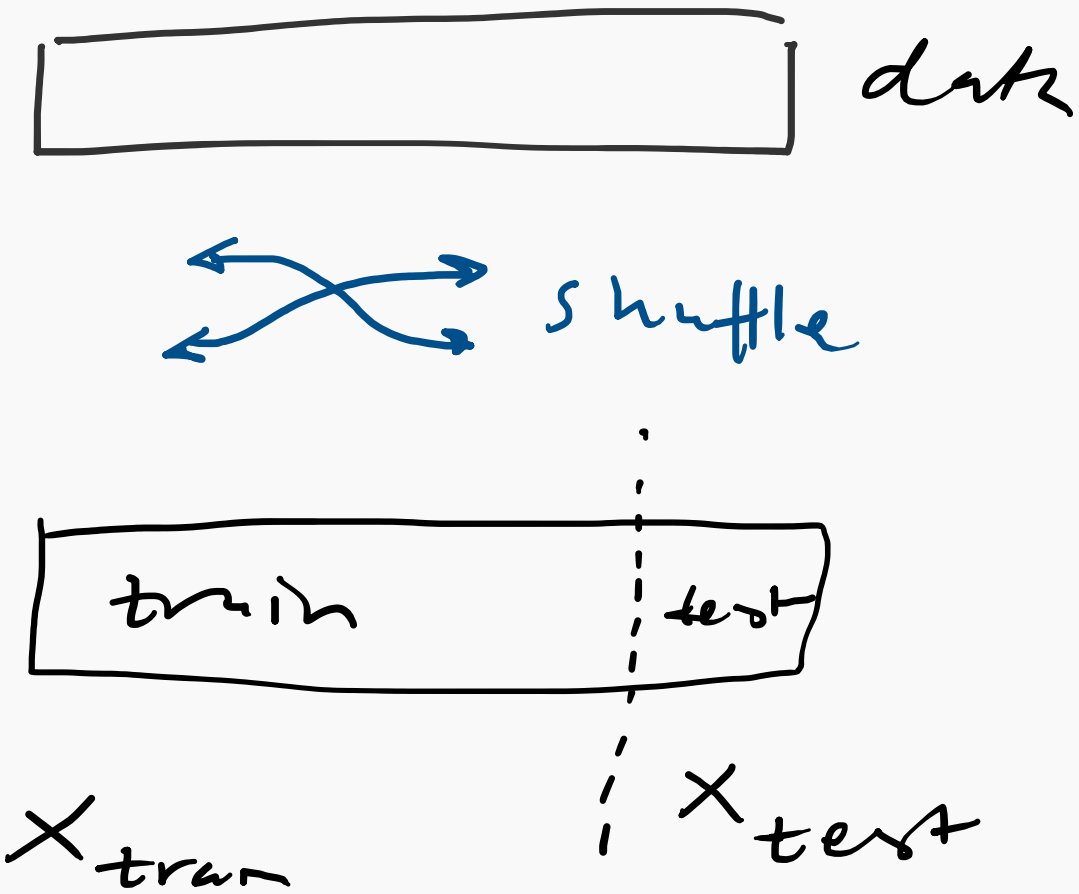


when n = 3

predict income example 4 – **more features**



train|test split



1. train model with X_{train} & y_{train}
2. predict on X_{test} => y_{pred}
3. evaluate → compare y_{pred} against y_{test}

data leakage

evaluate on

train data

=> data leakage

evaluate on unseen

test data

=> avoids

data leakage