AI = code + data

↓

Python

a) Model - based

b) data - based

1990s - Scripting Era

SWIG

2000s - Numpy

Scipy

MATLAB = Matplotlib

2010s - As an alternative R

Wes McKinney Pandas 2011

↳ dataframe

2007 Scikit - learn

from Scikit.linear_models import LR

from sk_metrics import R^2

Scikit
- Dataset (Boston)
- Imputer — cleaning & filling missing values
- preprocessing
  - ↳ feature Scaling
- Pipeline → Pipeline / FeatureUnion
- models → LR, Regressor, Classifier
- Metrics — R^2, MSE, Accuracy

Feature Transformation (or) Scaling

1. Min-Max Scaler = Normalization

$$= \frac{x - \min(x)}{\text{Range}}, \quad \text{Range} = \max(x) - \min(x)$$

2. StandardScaler = Z-score

$$\frac{x - \mu}{\sigma}$$

3. Robust Scaler = rescale outliers using IQR, $IQR = Q_1 - Q_3$

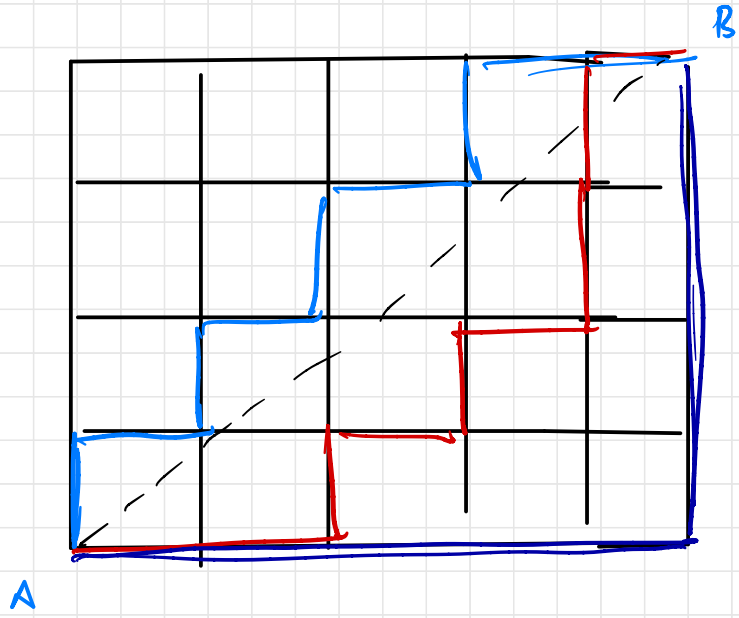
$$= \frac{x_i - Q_1(x)}{Q_3(x) - Q_1(x)} \leftarrow IQR$$

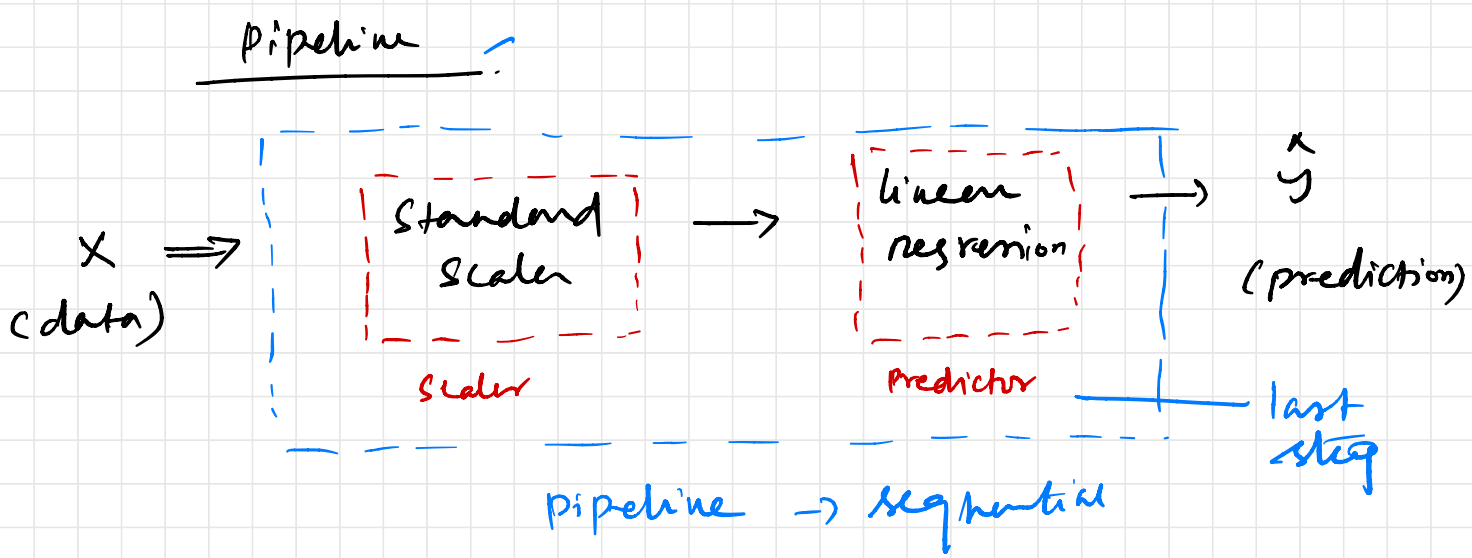
④ Normalizer, (1) ≠ (4)

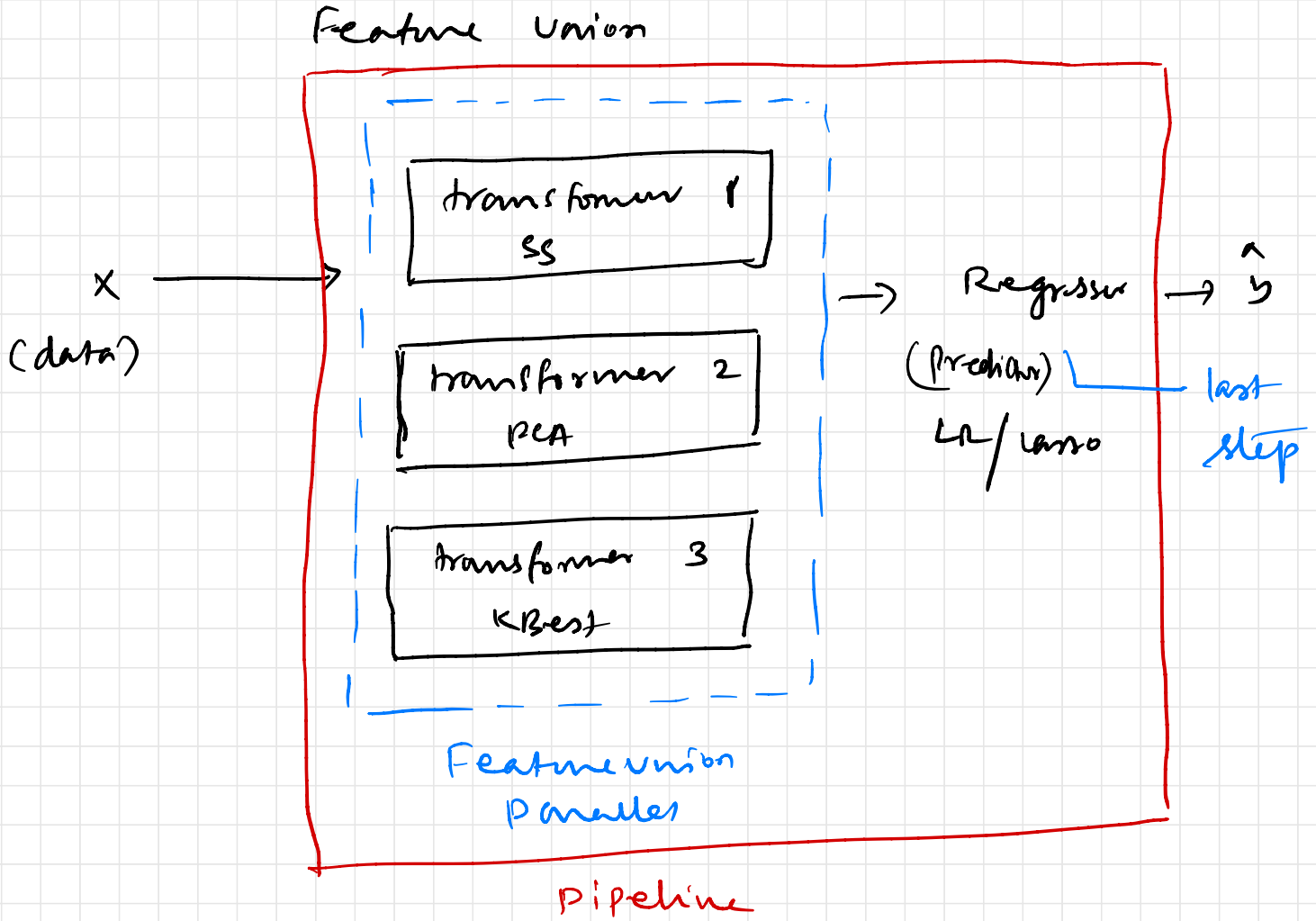
a) L_1 Norm — Taxicab (or) Manhattan Norm ✓

b) L_2 Norm — N4 Bird line, Euclidian

c) Max Norm




Pipeline
X
(data)
Standard
Scaler
Scaler
linear
regression
Predictor
ŷ
(prediction)
last
step
pipeline → sequential


Feature Union
X
(data)
transformer 1
SS
transformer 2
PCA
transformer 3
KBest
Featureunion
Parallel
Regressor
(Predictor)
LR / Lasso
last
step
ŷ
pipeline