

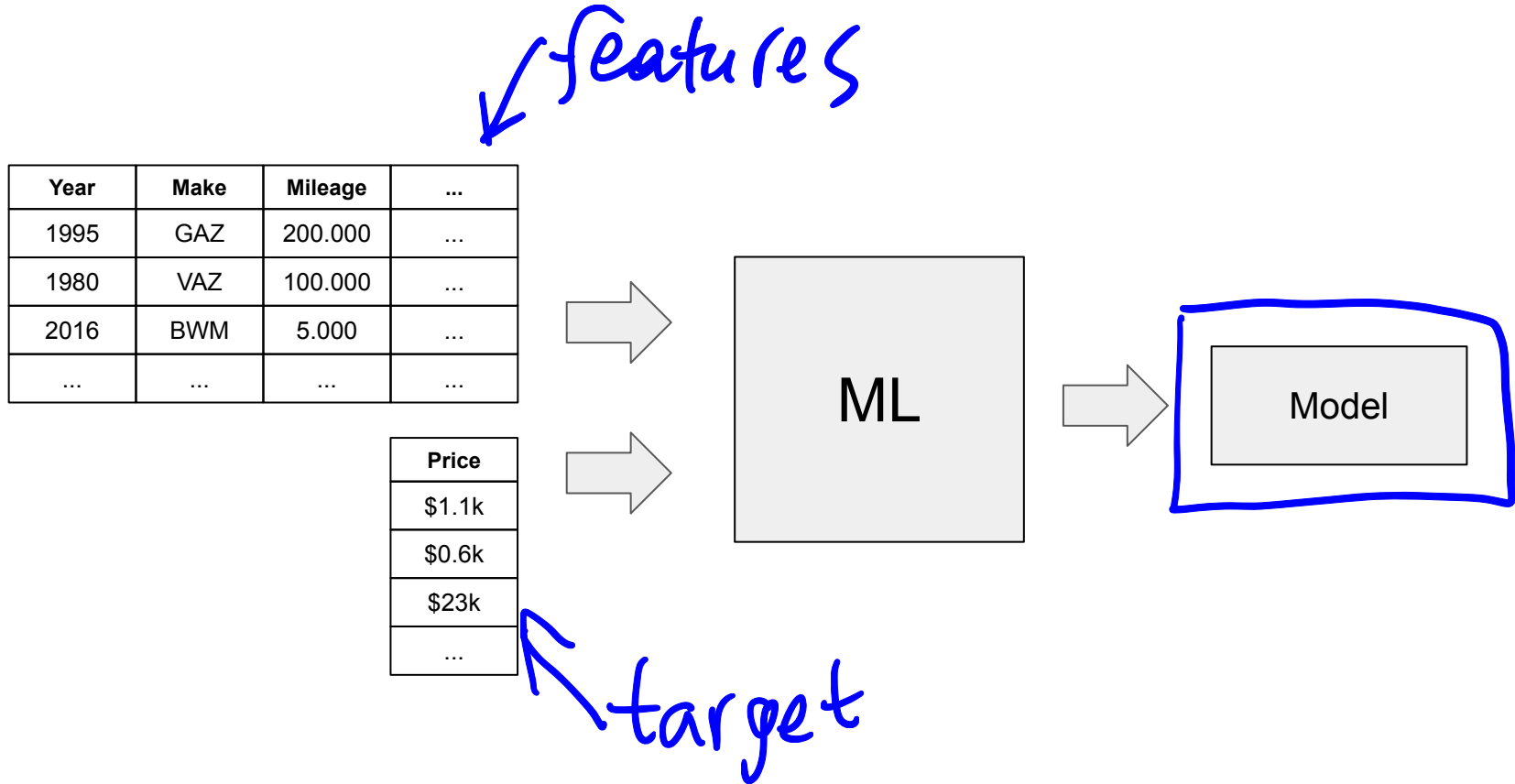
Machine Learning Zoomcamp

Session #1

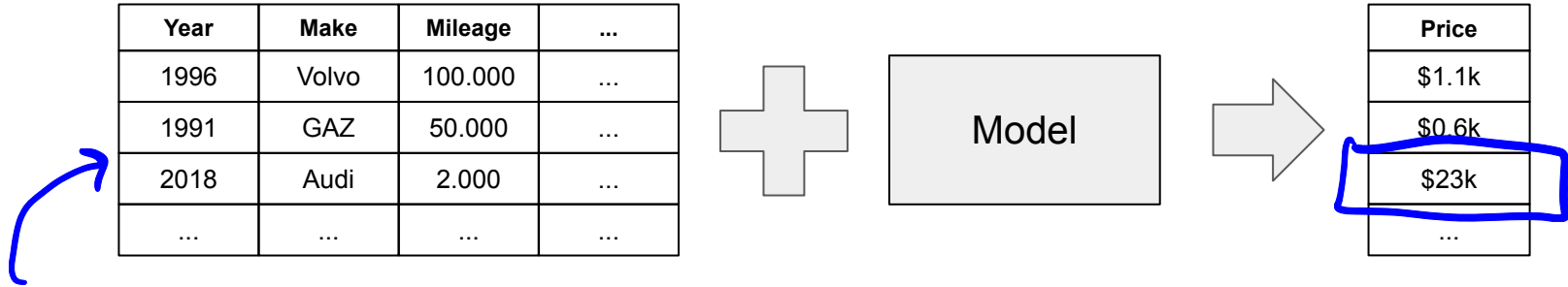
Summary

DataTalks.Club

1.1 Introduction to ML



1.1 Introduction to ML



1.2 Rules vs ML

- If sender = promotions@online.com then “spam”
- If title contains “tax review” and sender domain is “online.com” then “spam”
- If body contains a word “deposit”
 - If sender domain is “test.com” then “spam”
 - If body ≥ 100 words then spam
- Otherwise, “good email”

1.2 Rules vs ML

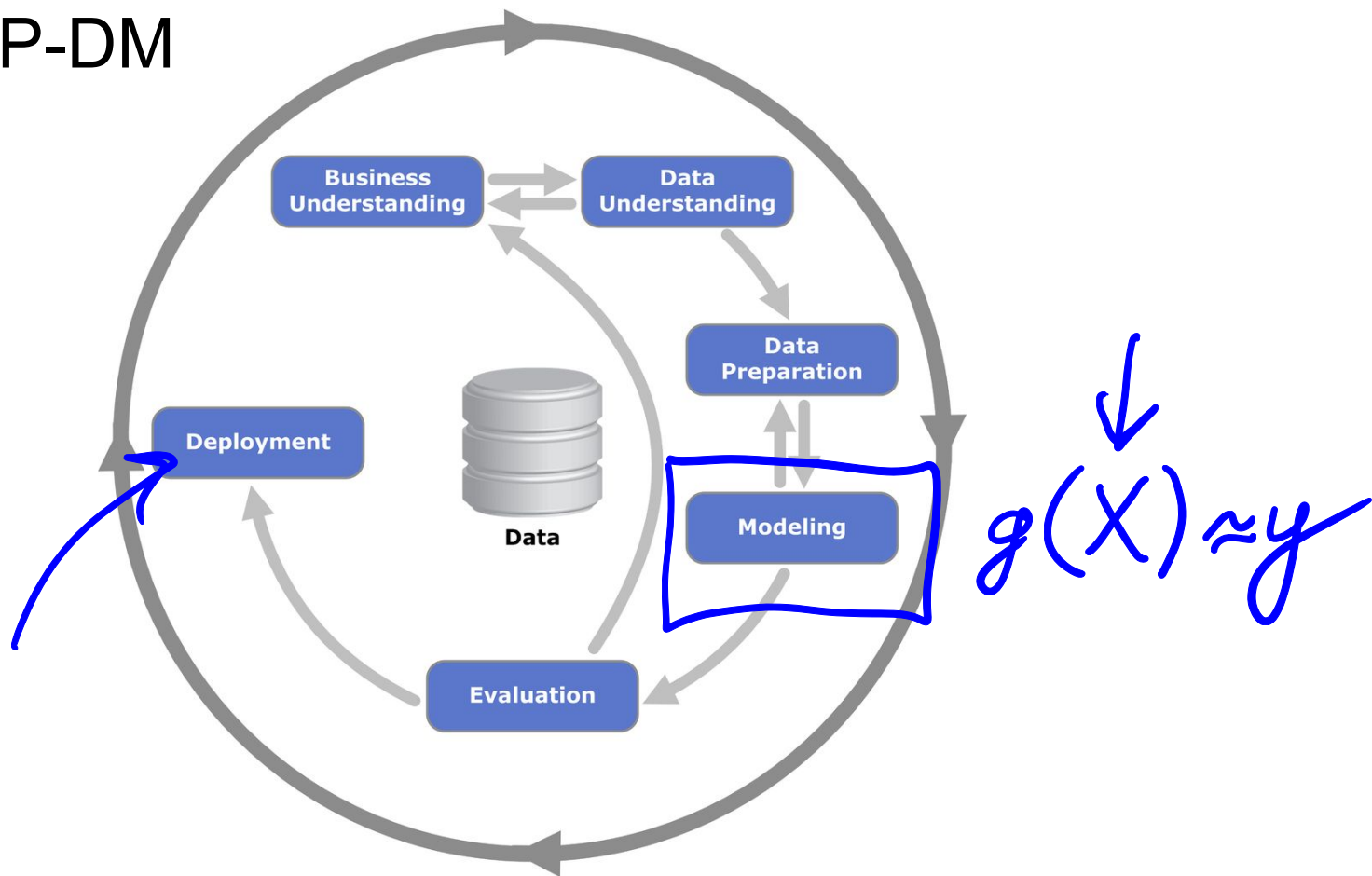
Features (data)	Target (desired output)
[1, 1, 0, 0, 1, 1]	1
[0, 0, 0, 1, 0, 1]	0
[1, 1, 1, 0, 1, 0]	1
[1, 0, 0, 0, 0, 1]	1
[0, 0, 0, 1, 1, 0]	0
[1, 0, 1, 0, 1, 1]	0

1.3 Supervised Machine Learning

A diagram illustrating the supervised machine learning equation $g(X) \approx y$. The equation is centered, with handwritten blue annotations and arrows. The word "feature" is written above the X with a downward arrow pointing to it. The word "model" is written below the g with an upward arrow pointing to it. The word "target" is written below the y with an upward arrow pointing to it.

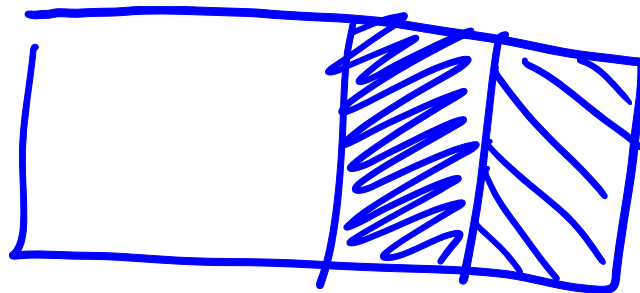
$$g(X) \approx y$$

1.4 CRISP-DM



1.5 Model Selection

1. Split data into train/validation/test
2. Train a model
3. Validate it
4. Select the best model
5. Test it



1.6 Environment

- Install Python, Numpy, Pandas, Matplotlib and Scikit-Learn
- Anaconda is the easiest option
- Creating account on AWS

1.7 Introduction to NumPy

1.8 Linear algebra

- Multiplication
- Formulas are not scary when you implement them



1.9 Introduction to Pandas

Next

- Session 2: Car price prediction project