Large scale machine learning

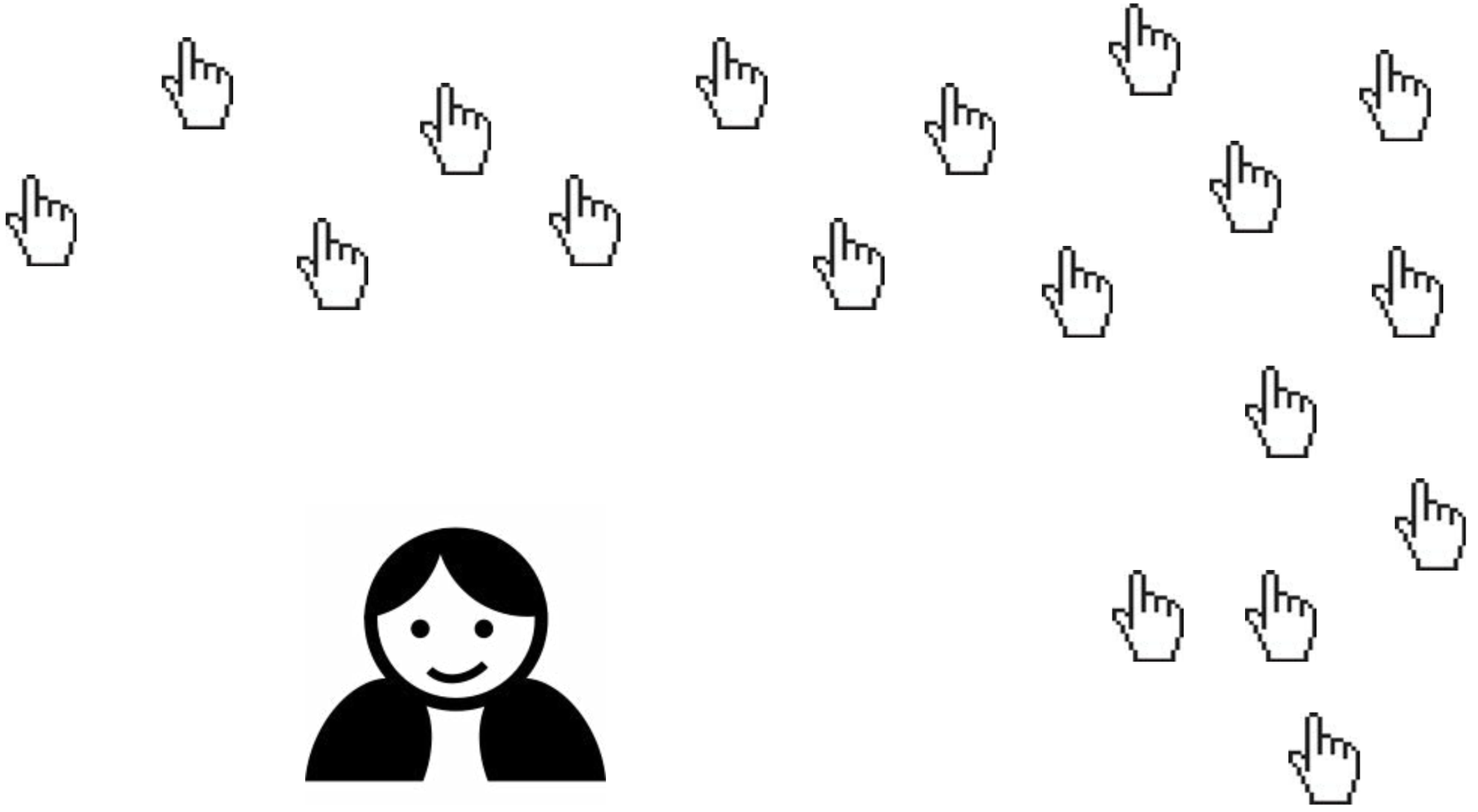
This week you will:

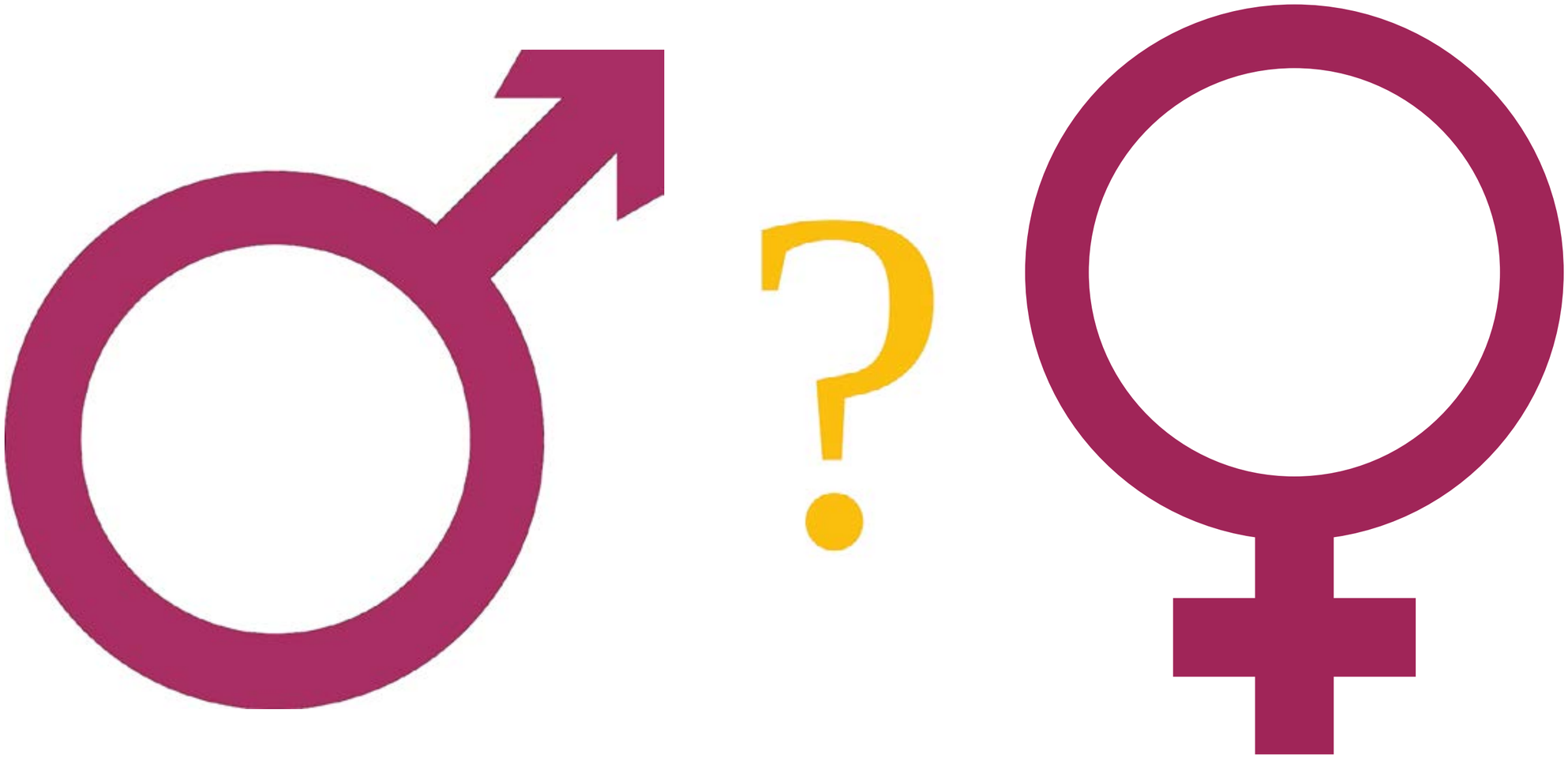
- Learn why you need large scale machine learning?
- Learn how to use Spark MLlib
- Train linear models on big data to make predictions
- Discover some techniques for improving prediction quality

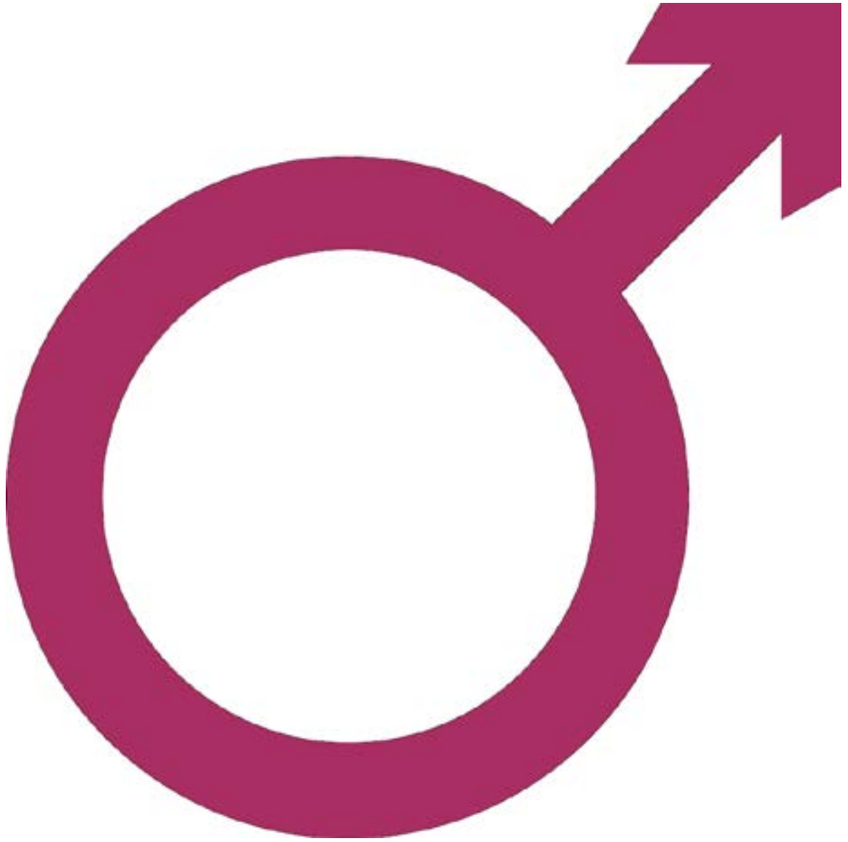
In this video you will learn:

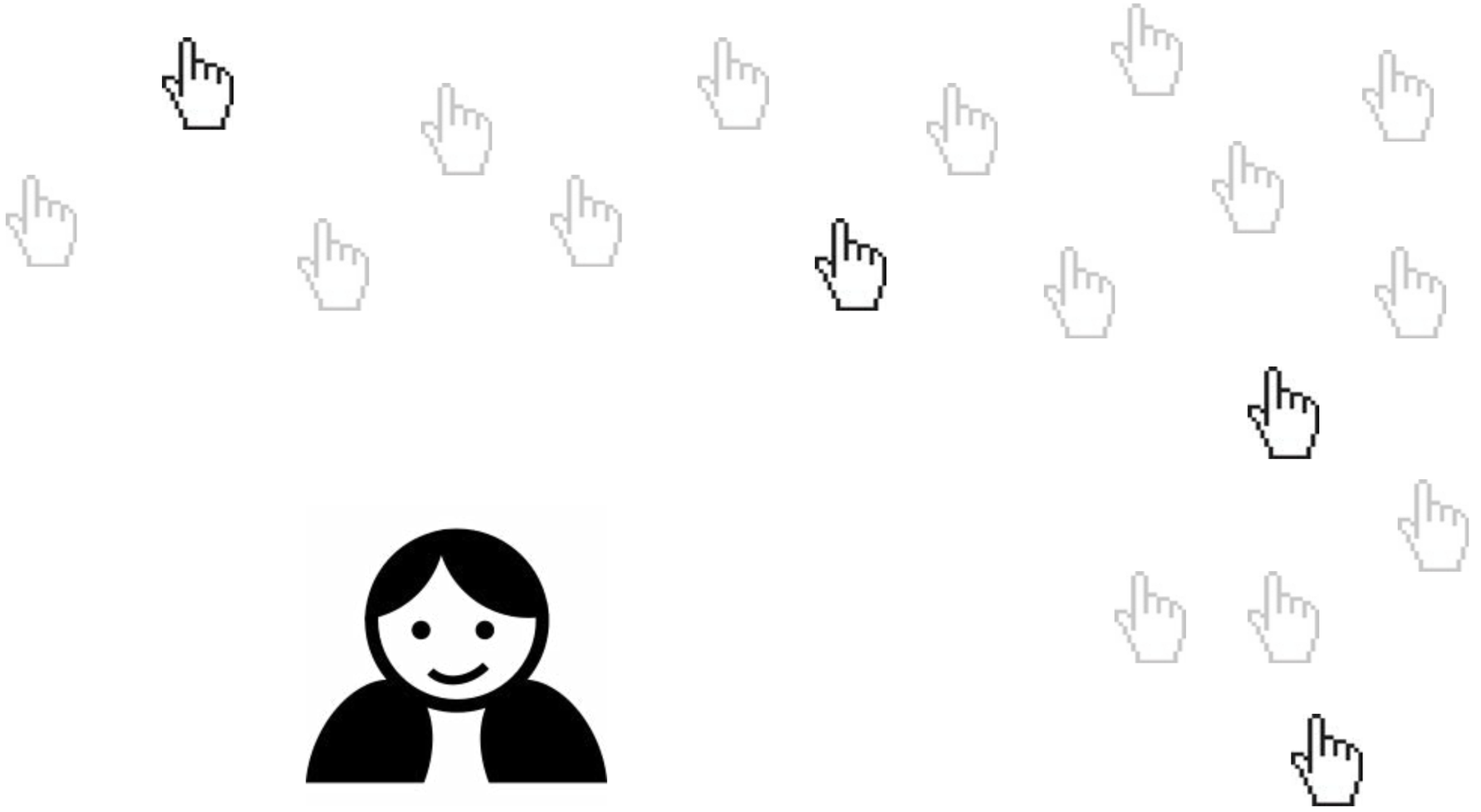
- How to correctly circumvent the problem of large data
- Why you need to train algorithms on large datasets
- Why you should study Spark MLlib

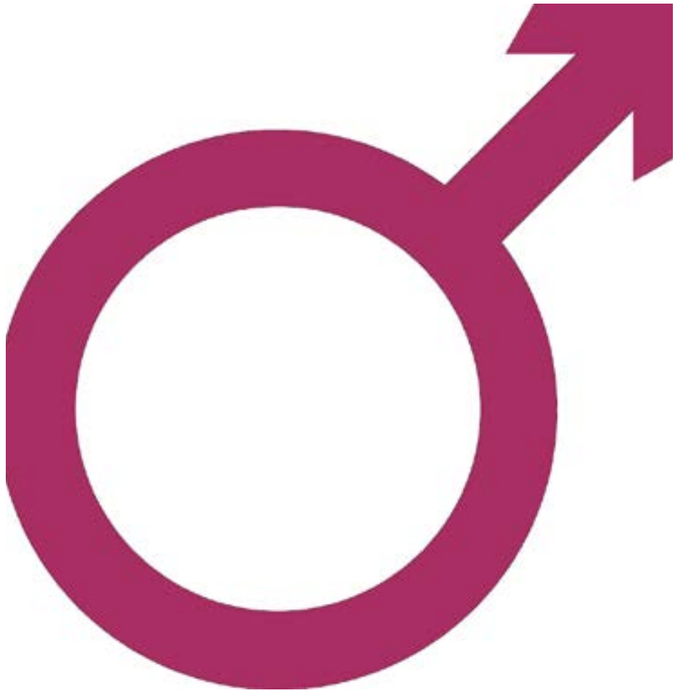
Split

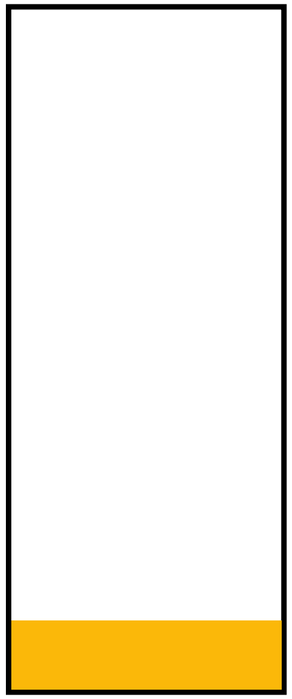








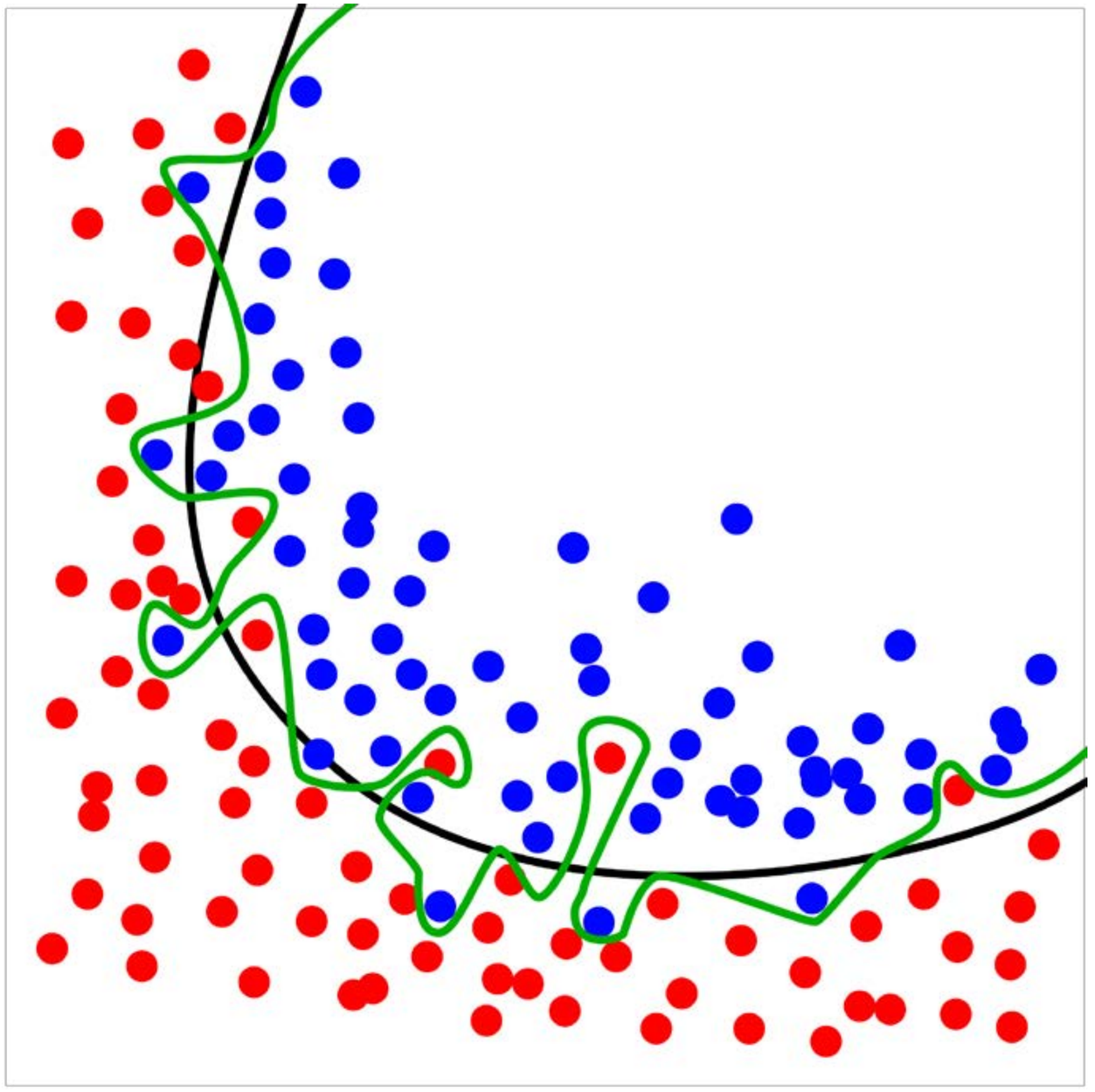


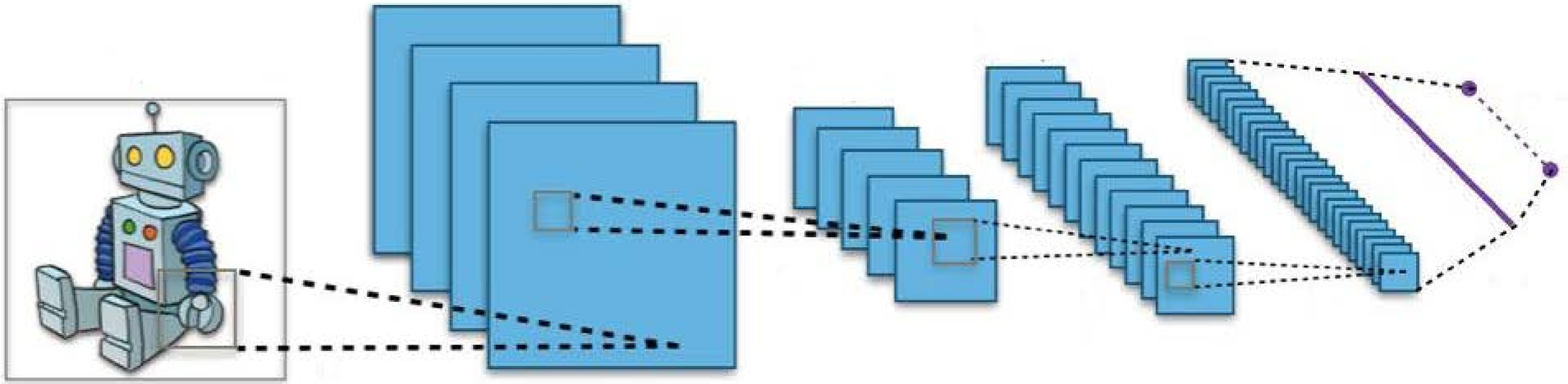


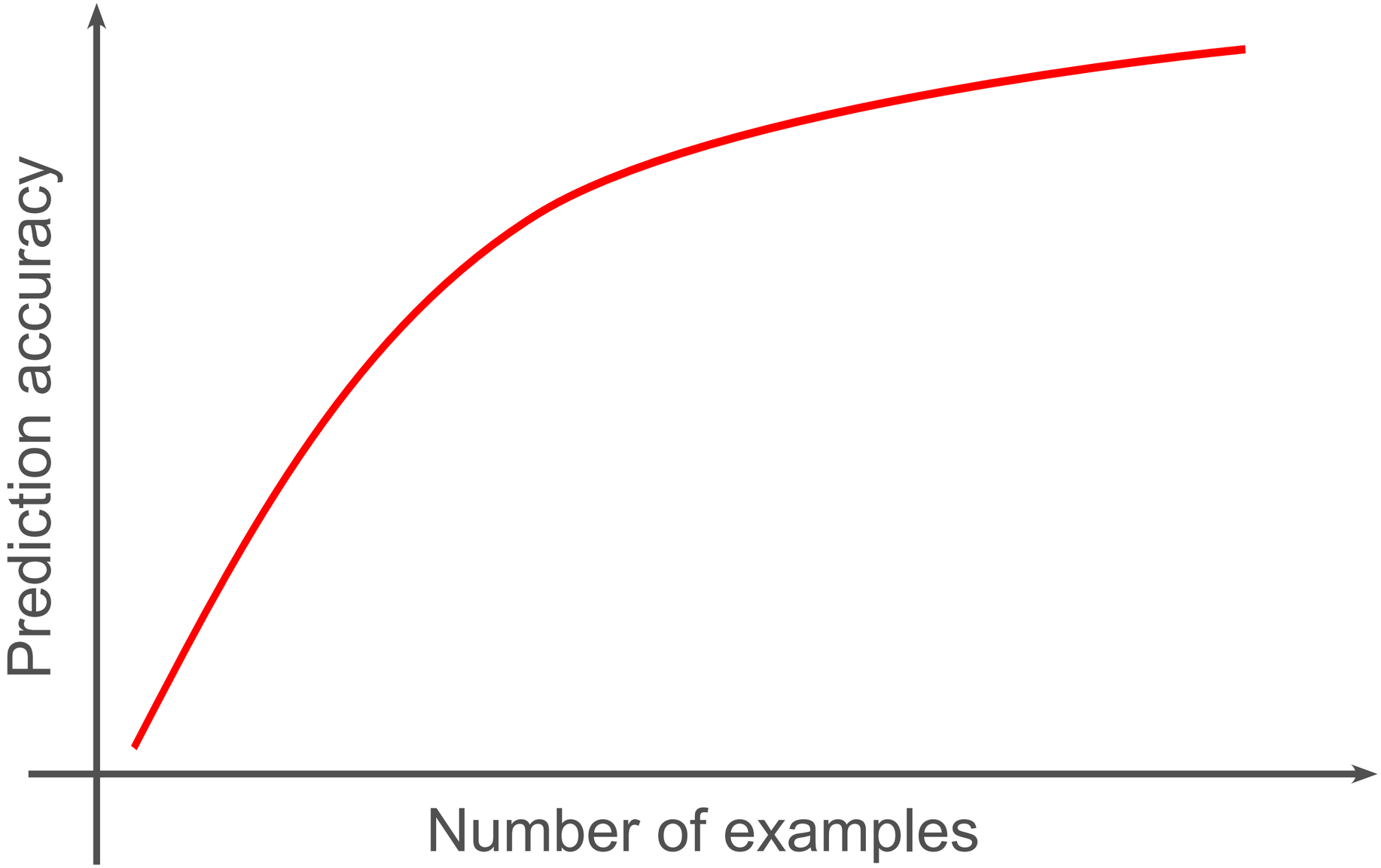
10 %

Create the right subsamples

Why do you need full dataset?






Prediction accuracy
Number of examples

Prediction of rare events


BUNDESREPUBLIK DEUTSCHLAND
5
DEUTSCHE MARK
1951
D

Why do you need a library of machine learning for big data?


SOUND || VISION

MLlib

SQL

Advantages of Spark SQL

- Simplicity
- Integration
- Performance

Spark MLlib advantages

- Unified way of data processing and application of models
- Easy deploymenta on big data
- A large number of third-party libraries

What have you learned today?

- How to correctly bypass
- The problem of large data
- Why you need a large scale machine learning
- Why you should learn Spark MLlib