

Machine Learning

B. N. M. Institute of Technology

Department: Artificial Intelligence and Machine Learning (AI)

Course: Machine Learning (18AI61) effective from the academic year 2018–2019
under VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELAGAVI

Instructor: Pradip Kumar Das

Machine Learning Landscape

What is Machine Learning?

Few Examples

Spam Email Filter

Recommender

*Optical Character
Recognition*

Chatbots

Voice Recognition

*Facial
Recognition*

Forecasting

*Autonomous
Vehicle*

*Medical
Diagnostics*

Fraud Detection

and many more...

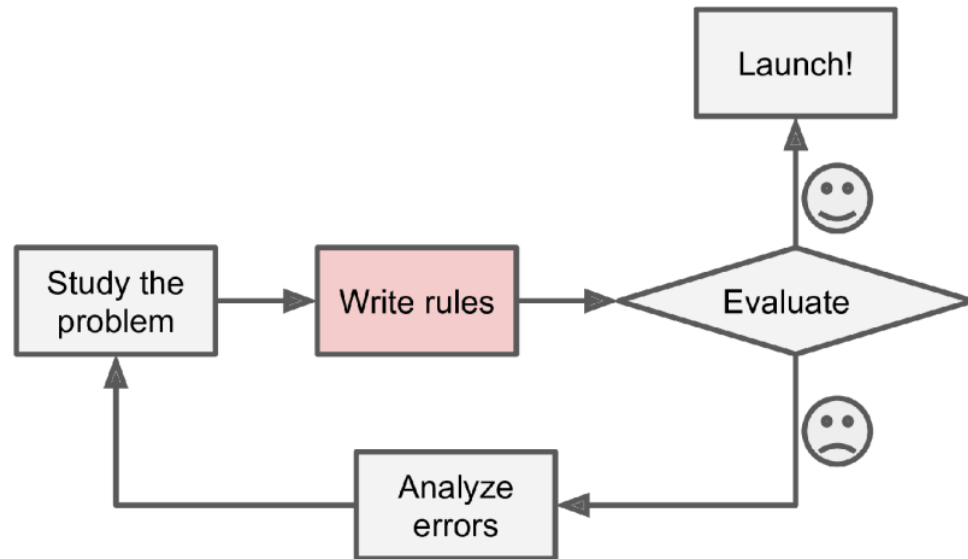
Some Definitions

"Science of programming computer to learn from data"

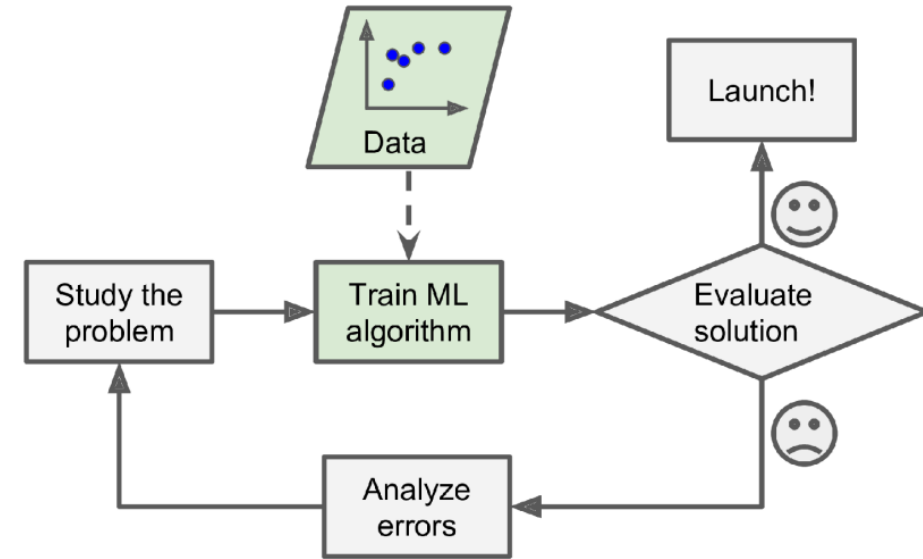
"field of study that gives computers the ability to learn without being explicitly programmed"

"A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E "

Why is Machine Learning Required?

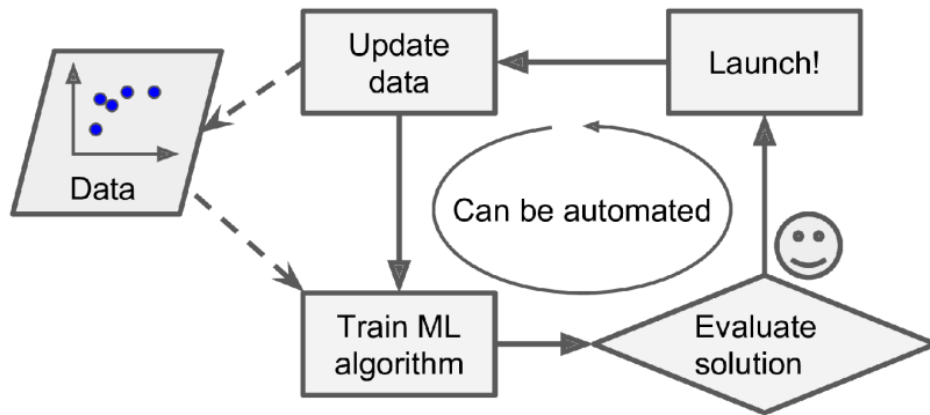


Traditional program writing approach

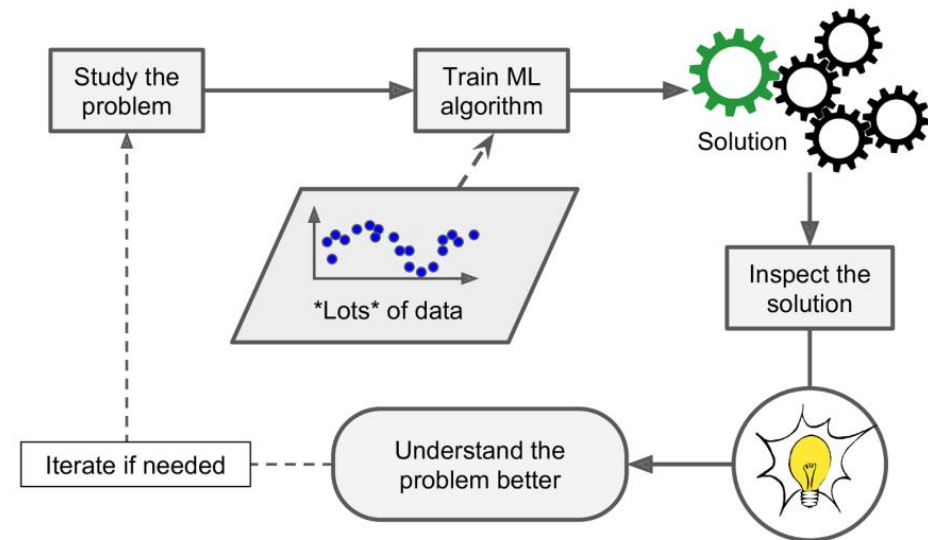


The Machine Learning approach

Why is Machine Learning Required? (Cont.)



Automatically adapting change



Discovering patterns for human learning

Machine Learning is a Good Option for

- Problems that require a lot of fine-tuning and long list of rules
- Problems that traditional approaches do not yield good results
- Problems where data distribution changes over time
- Getting insights about complex problems from large volume of data

Types of Machine Learning Systems

Whether learning requires human supervision

- Supervised Learning
- Semi-supervised Systems
- Unsupervised Learning
- Reinforcement Learning

Whether learning is incremental

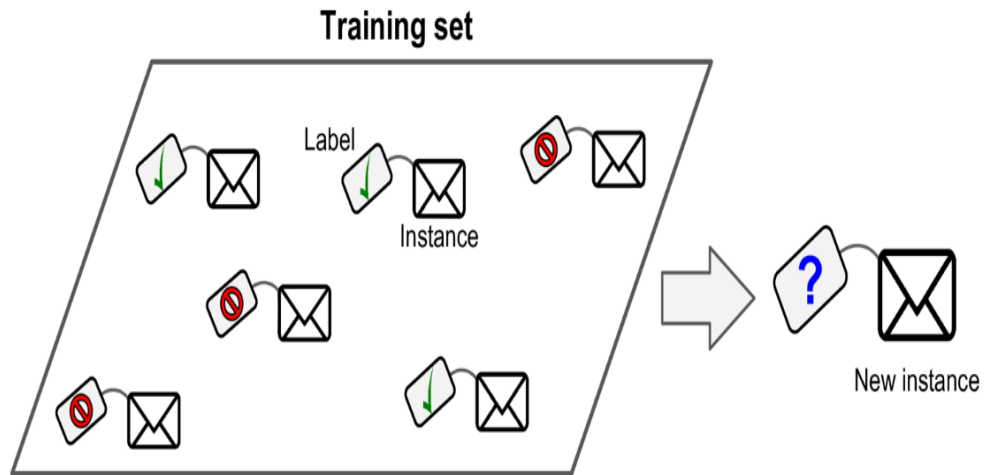
- Batch Learning
- Online Learning

Whether learning is to compare or to detect pattern and/or predict

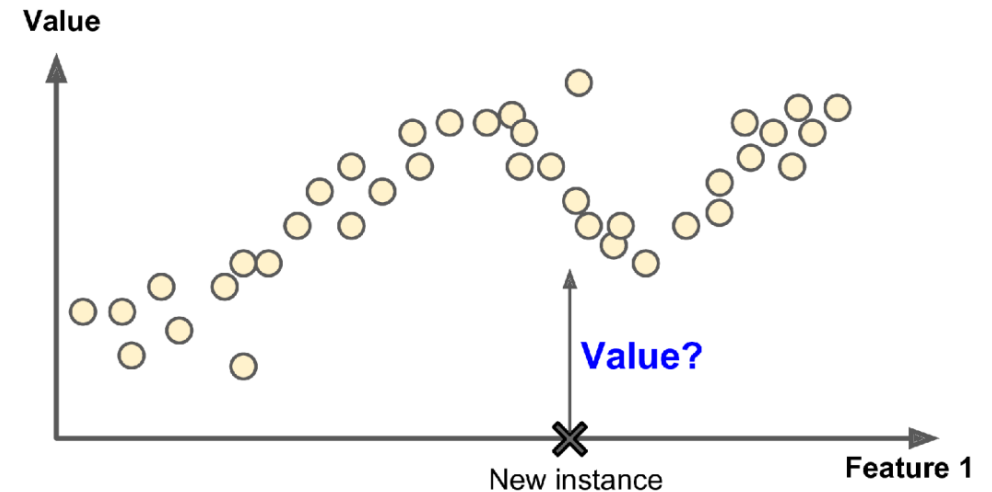
- Instance-based Learning
- Model-based Learning

Types of Machine Learning Systems (Cont.)

Classification & Regression in Supervised Learning



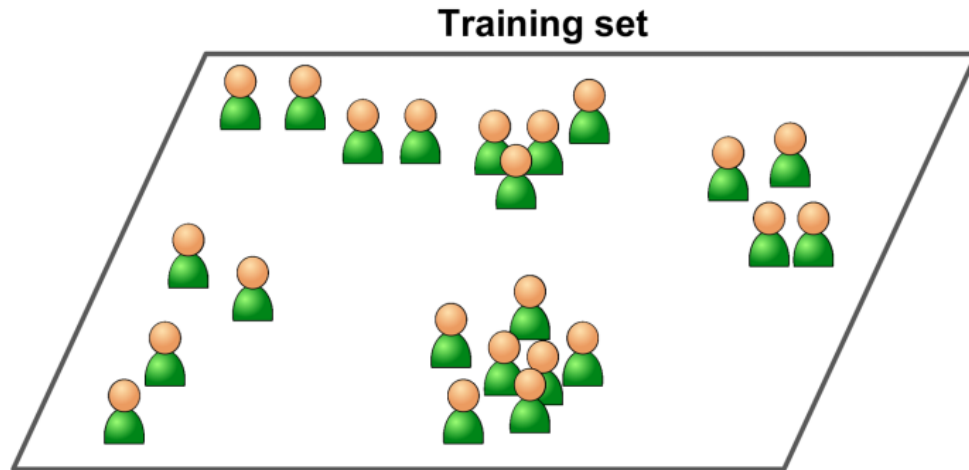
A labeled training set for spam classification



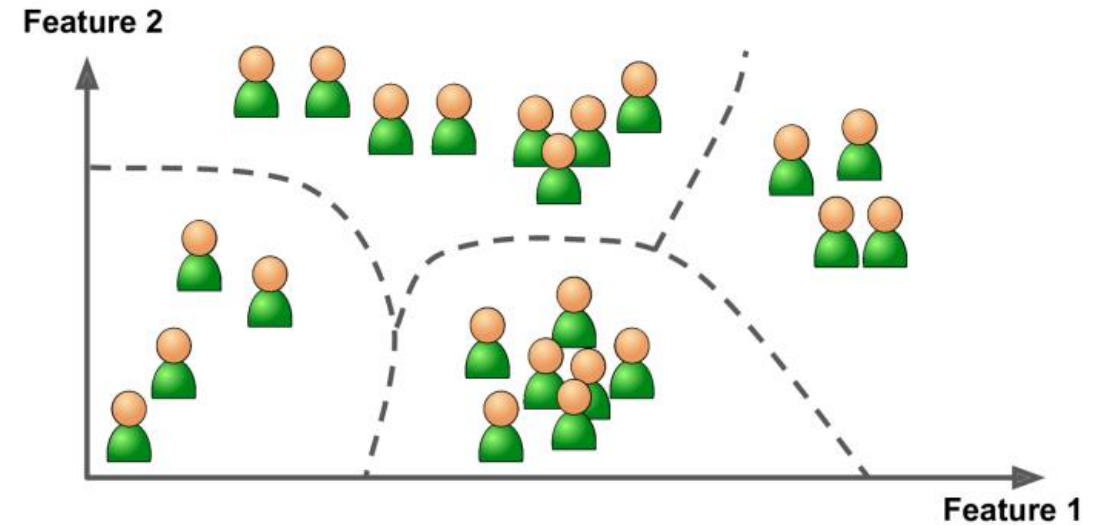
Predicting a continuous value in regression problem

Types of Machine Learning Systems (Cont.)

Clustering in Unsupervised Learning



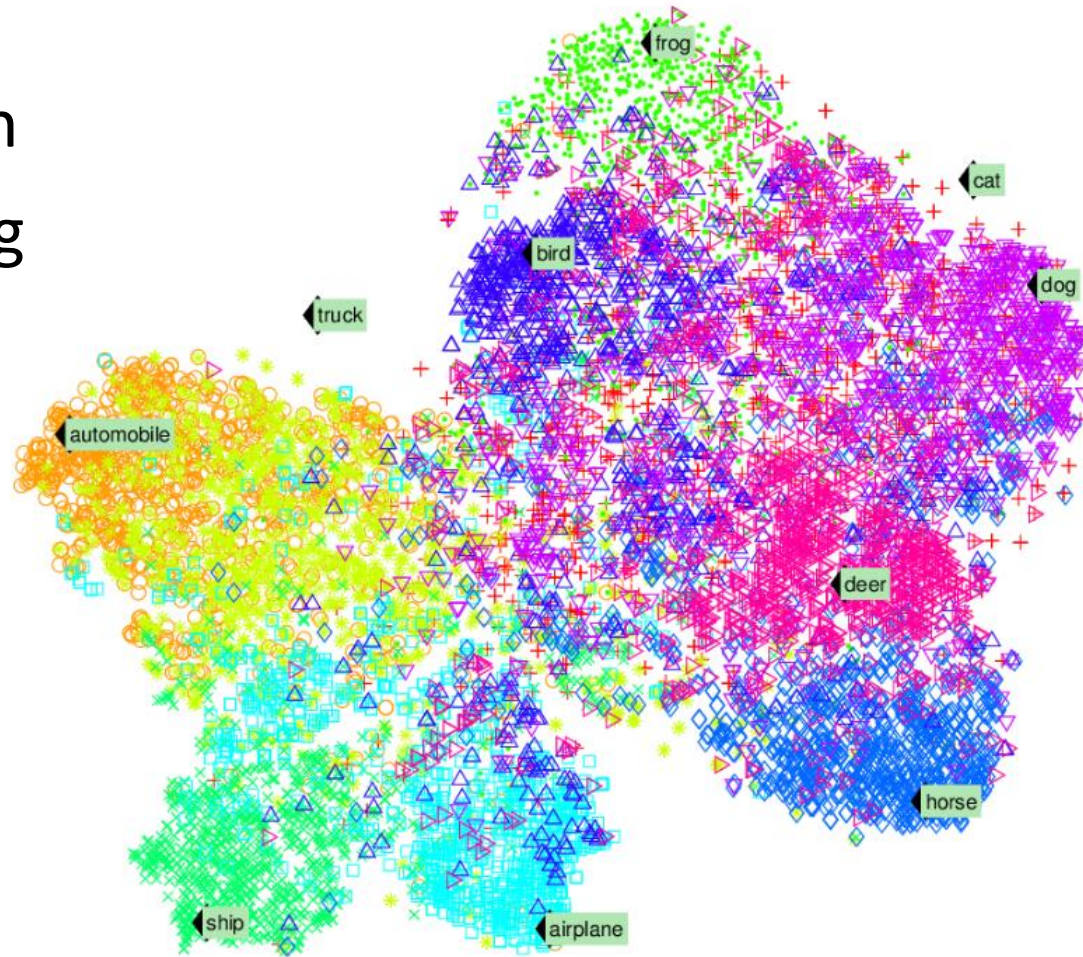
An unlabeled training set for clustering



Detecting groups of similar visitor in clustering

Types of Machine Learning Systems (Cont.)

Cluster Visualization in Unsupervised Learning



Cluster visualization in unsupervised learning

Types of Machine Learning Systems (Cont.)

Dimensionality Reduction in Unsupervised Learning

- Simplifying data without losing too much information
- Merging correlated features into one
- Feature extraction

Types of Machine Learning Systems (Cont.)

Anomaly and Novelty Detection in Unsupervised Learning



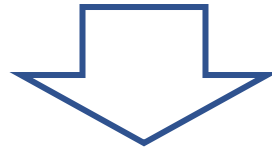
Detection of an anomaly

Types of Machine Learning Systems (Cont.)

Association Rule Learning in Unsupervised Learning

{Potatoes, Onions} => {Burger}

{Barbeque sauce, Potato chips} => {Steak}

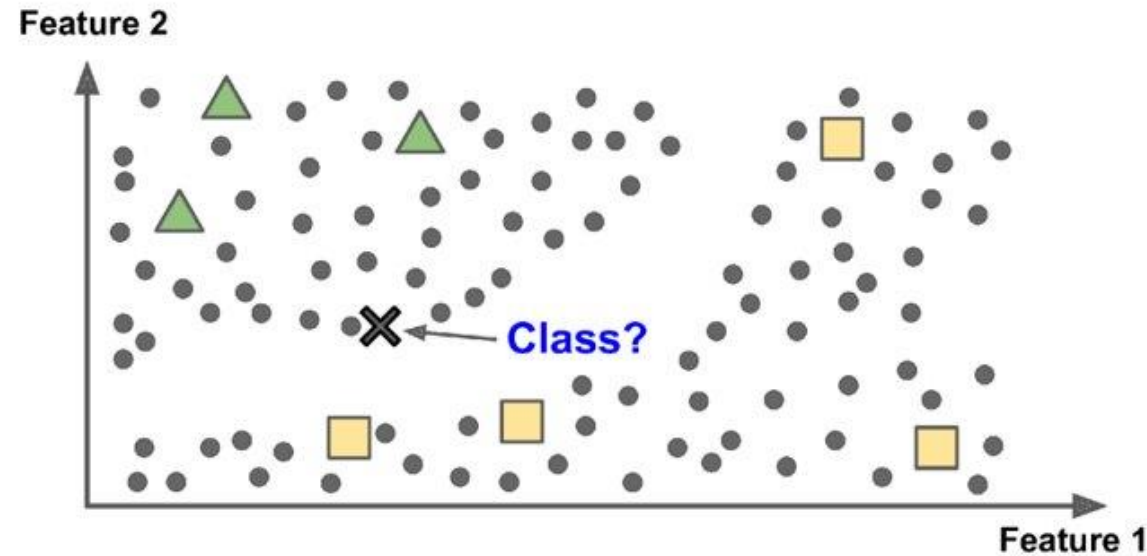


**Results in Promotional Pricing and/or
Appropriate Product Placement**

Sales logs revealing customers' buying pattern

Types of Machine Learning Systems (Cont.)

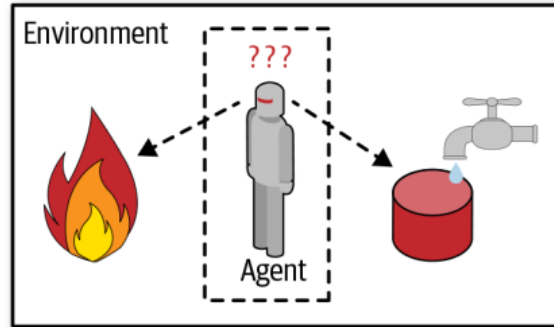
Classification in Semisupervised Learning



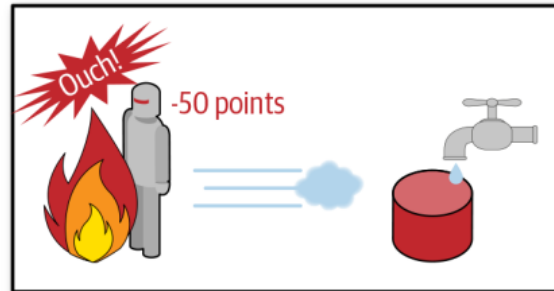
Few labeled examples help classifying new instances

Types of Machine Learning Systems (Cont.)

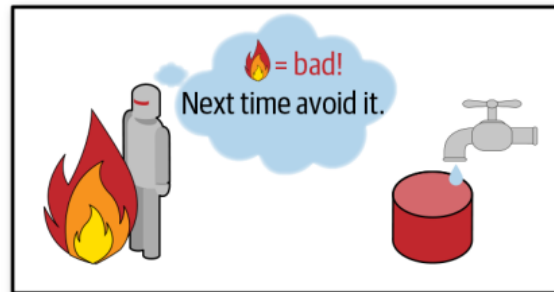
Building Strategy in Reinforcement Learning



- 1 Observe
- 2 Select action using policy



- 3 Action!
- 4 Get reward or penalty

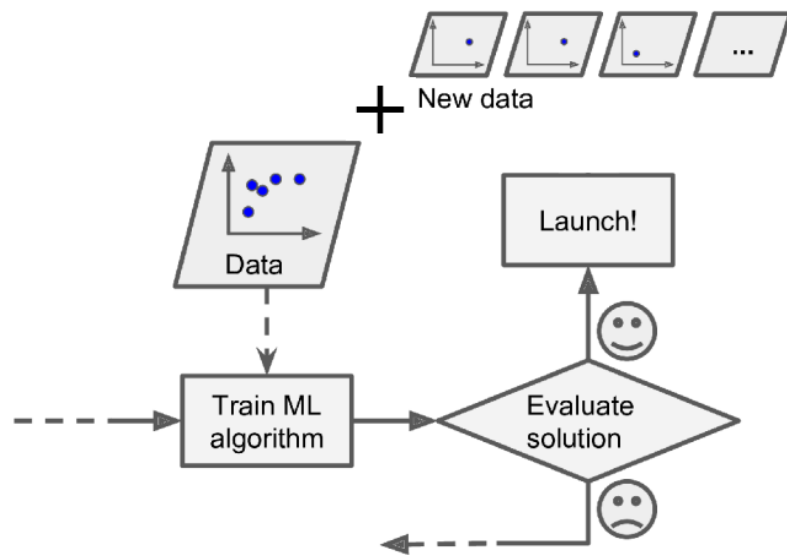


- 5 Update policy (learning step)
- 6 Iterate until an optimal policy is found

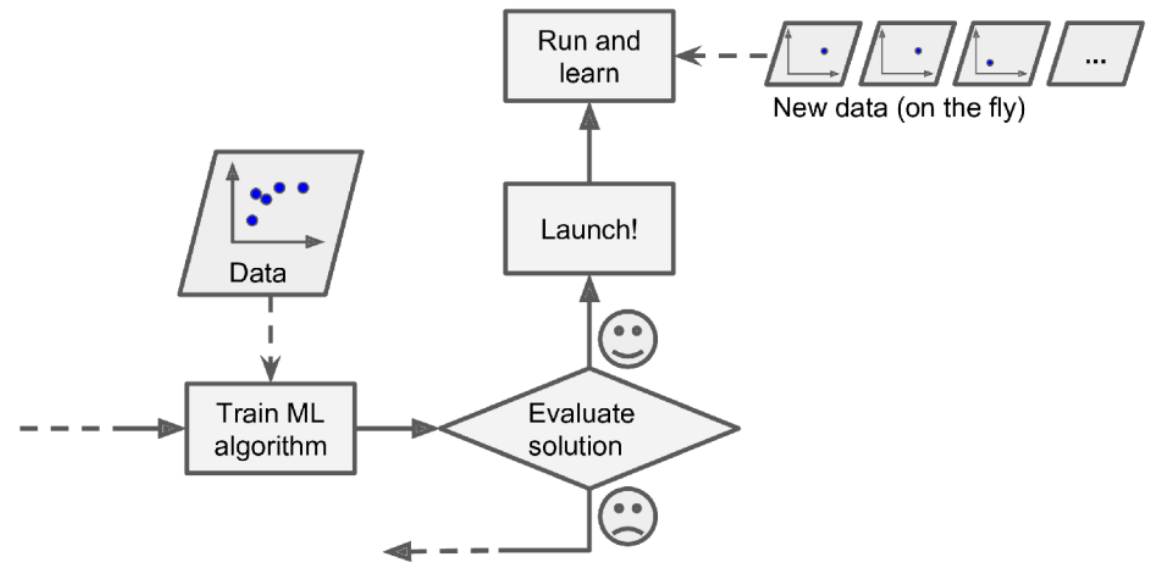
Receiving reward or penalty against its actions and fine-tuning its strategy

Types of Machine Learning Systems (Cont.)

Batch and Online Learning

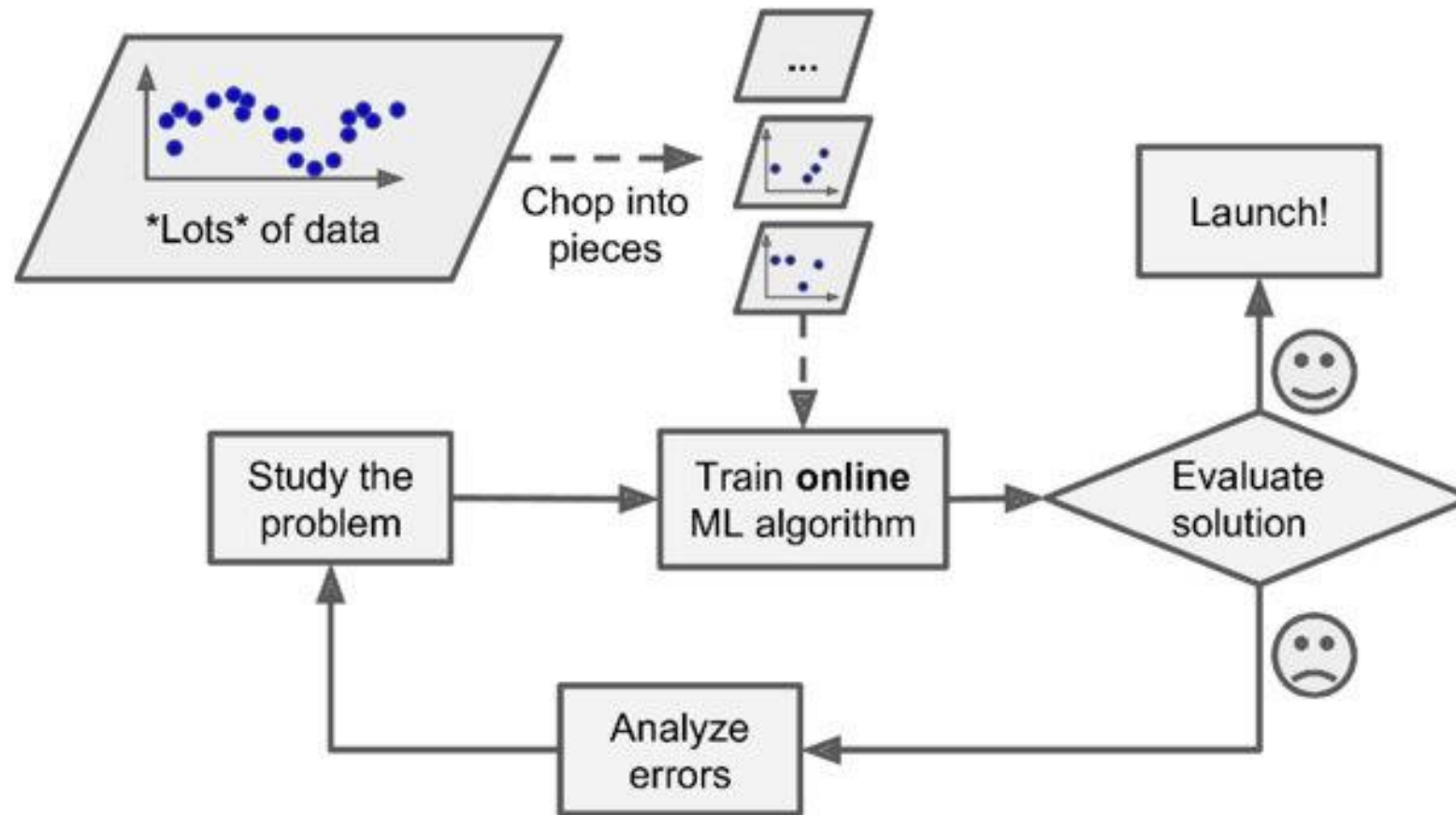


*Relearning from scratch on full dataset (old and new data)
in Batch Learning*



Learning is incremental for new data in Online Learning

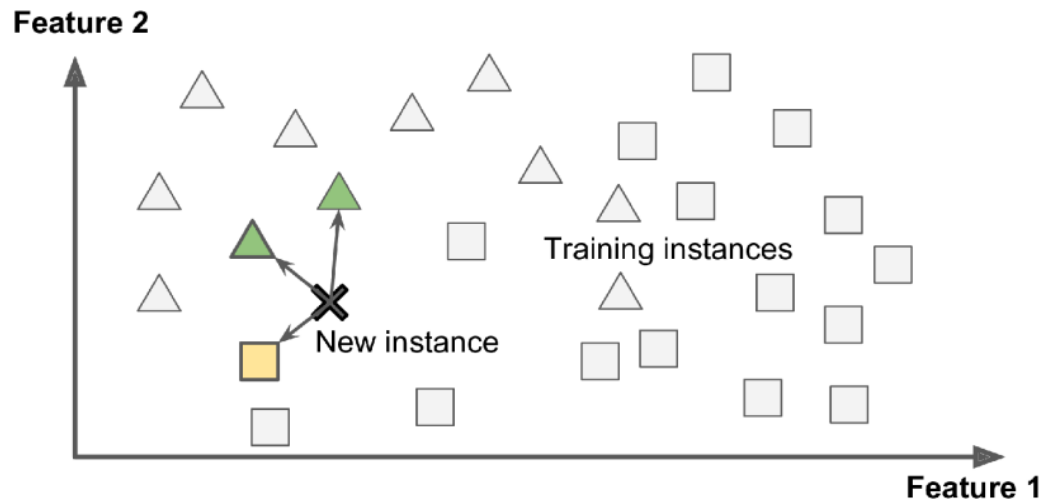
Types of Machine Learning Systems (Cont.)



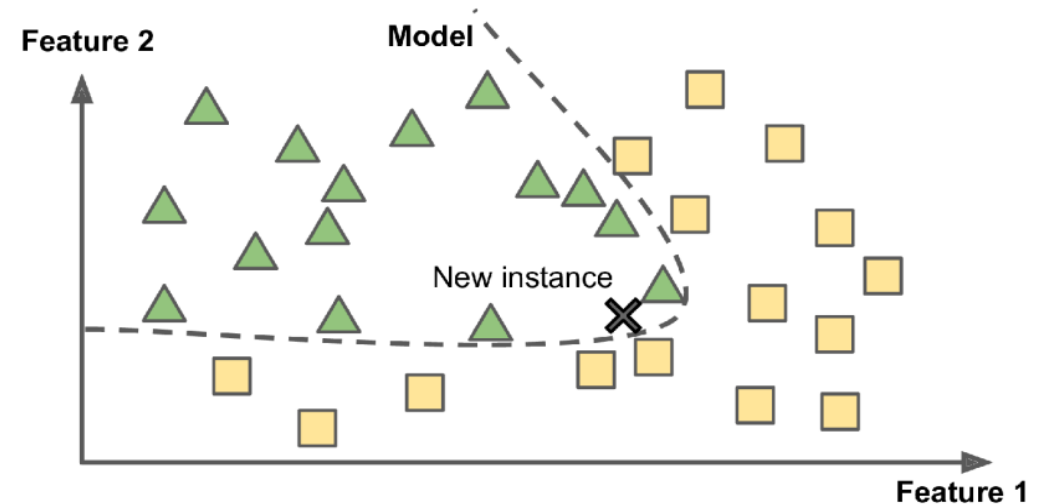
Handling huge dataset in Online Learning

Types of Machine Learning Systems (Cont.)

Instance-based or Model-based Learning systems on how they generalized



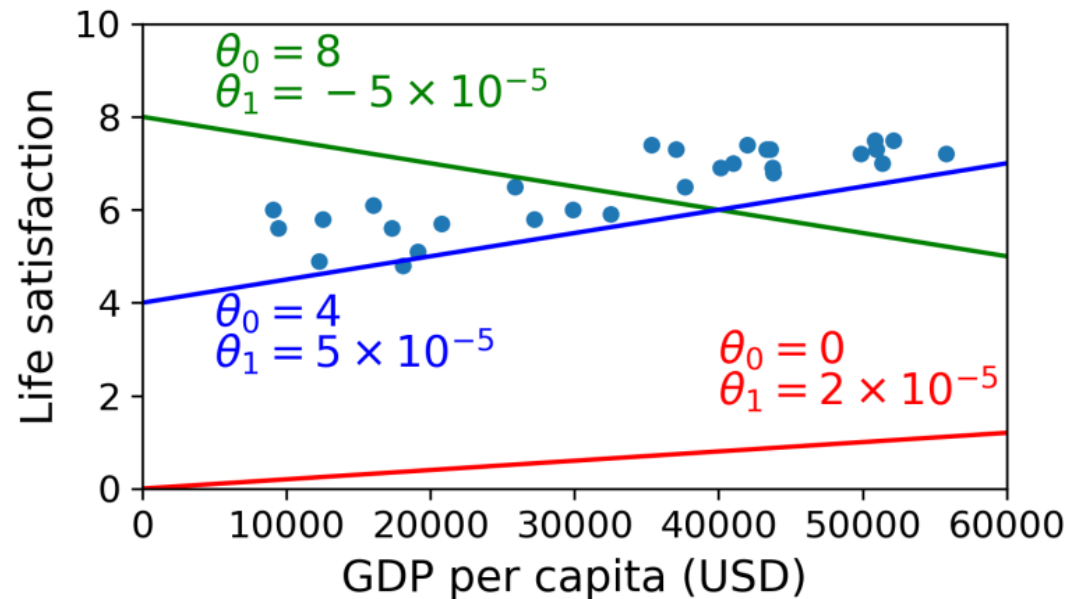
Instance-based Learning is by heart and it then generalizes to new cases by using similarity measures



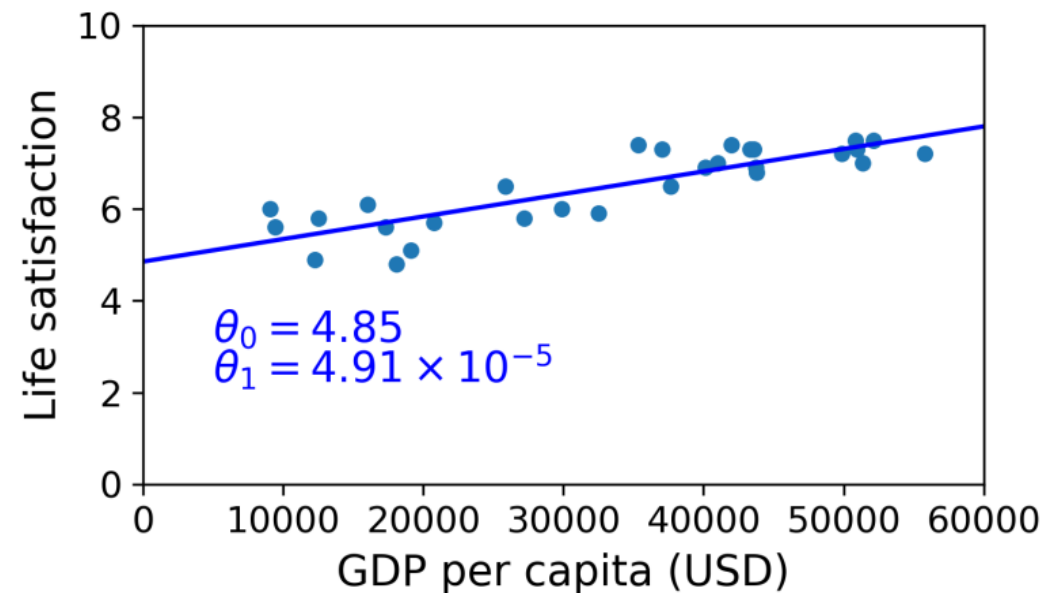
Model is created in Model-based Learning and then it make predictions

Types of Machine Learning Systems (Cont.)

Fitting Model to Data in Model-based Learning



Possible linear models

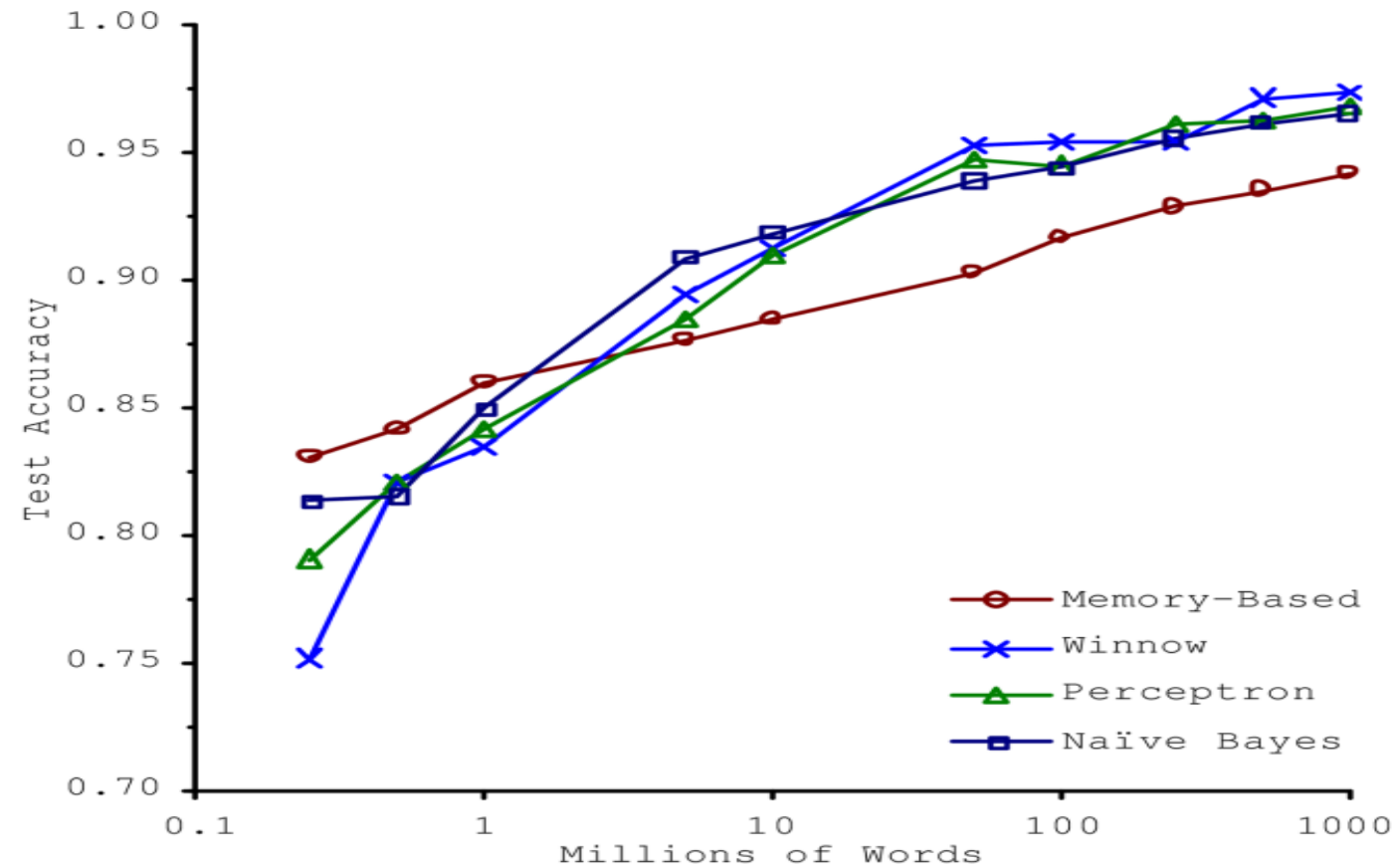


Fitted linear model

Main Challenges in Machine Learning

Insufficient Quantity of Training Data

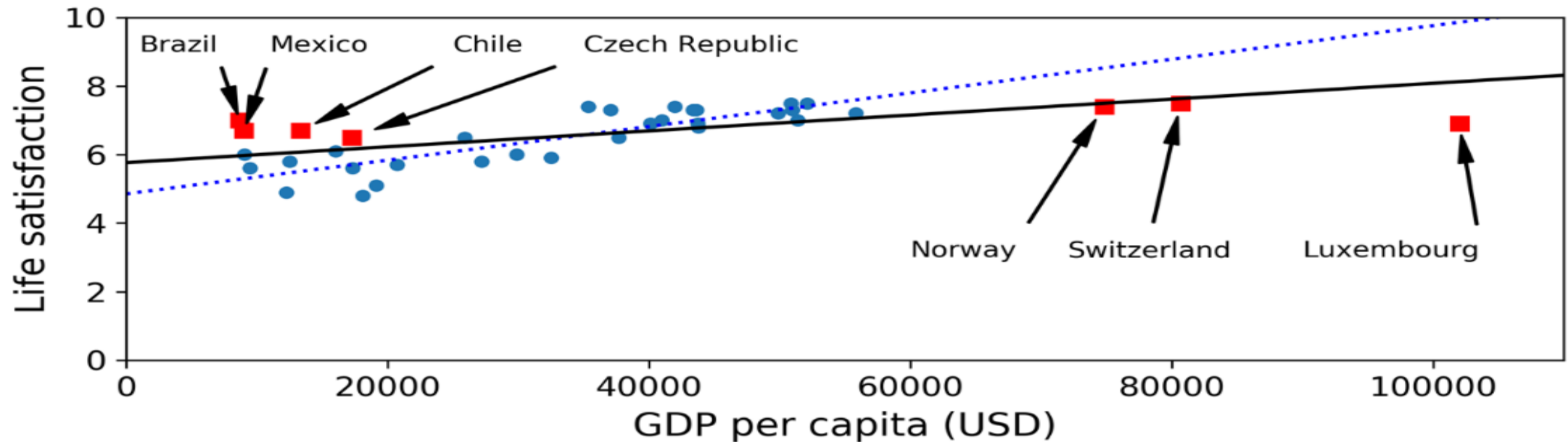
Trade-off between spending time and money on algorithm development and spending these on corpus development



Main Challenges in Machine Learning (Cont.)

Nonrepresentative Training Data

- Sampling Noise
- Sampling Bias



A better model fitted over more representative training data

Main Challenges in Machine Learning (Cont.)

Poor-Quality Data

- Errors
- Missing values
- Outliers

Main Challenges in Machine Learning (Cont.)

Irrelevant Features

Applying Feature Engineering involving the following steps

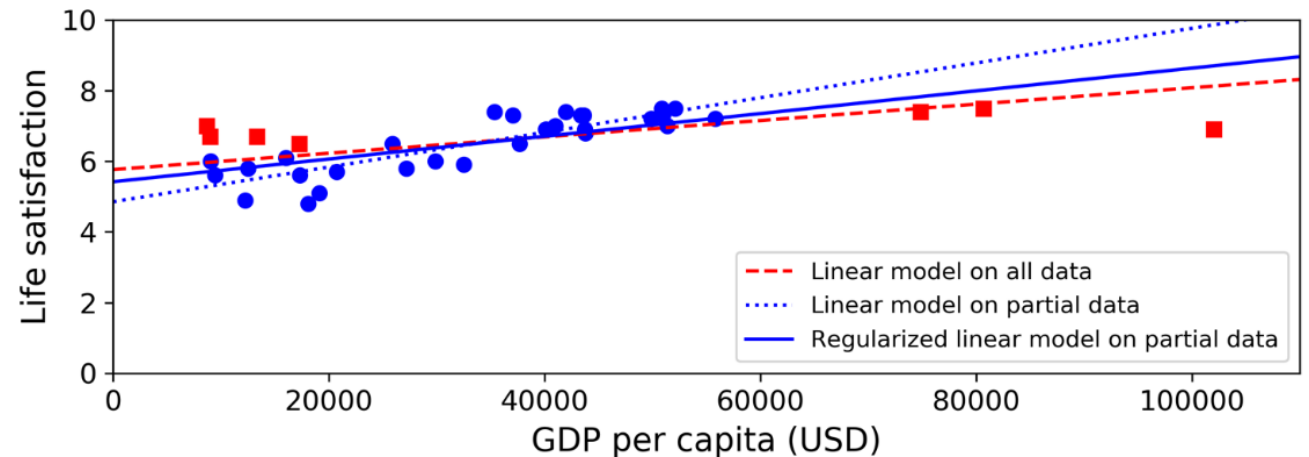
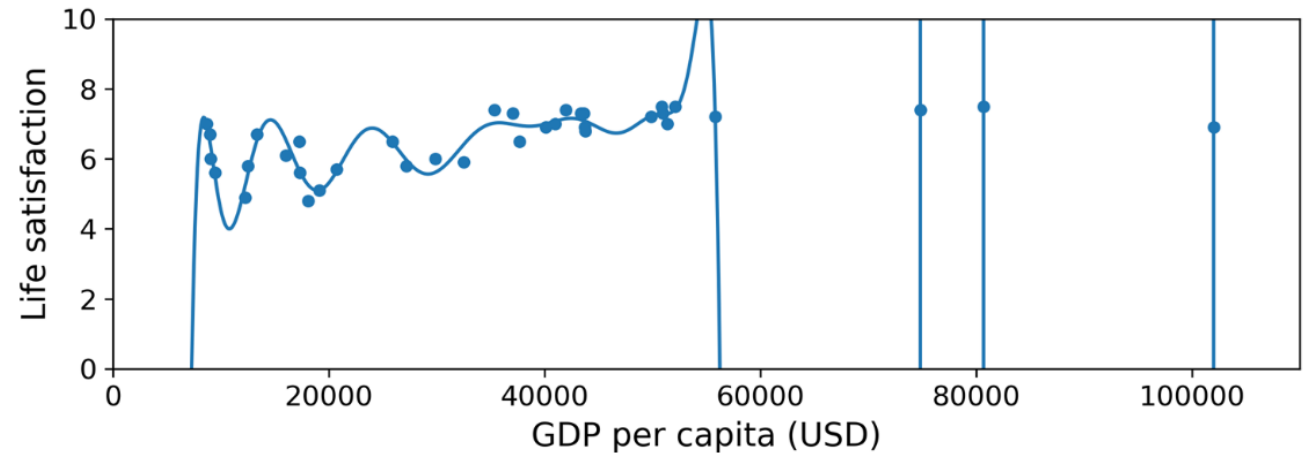
- Feature selection
- Feature extraction
- Creating new features

Main Challenges in Machine Learning (Cont.)

Overfitting the Training Data

Resolving by

- Simplifying model
- Applying constraints (regularization)
- Gathering more training data
- Fixing data quality issues



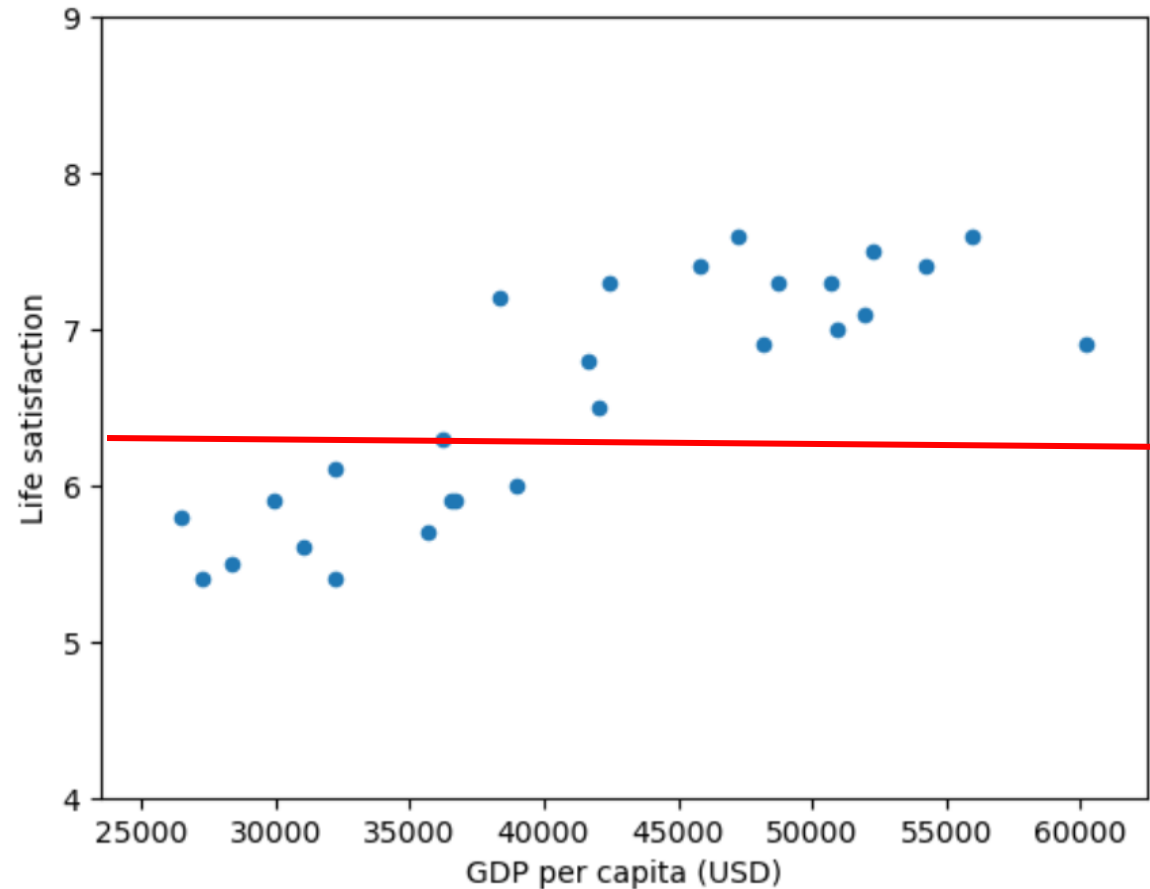
Overfitting and applying regularization to avoid it

Main Challenges in Machine Learning (Cont.)

Underfitting the Training Data

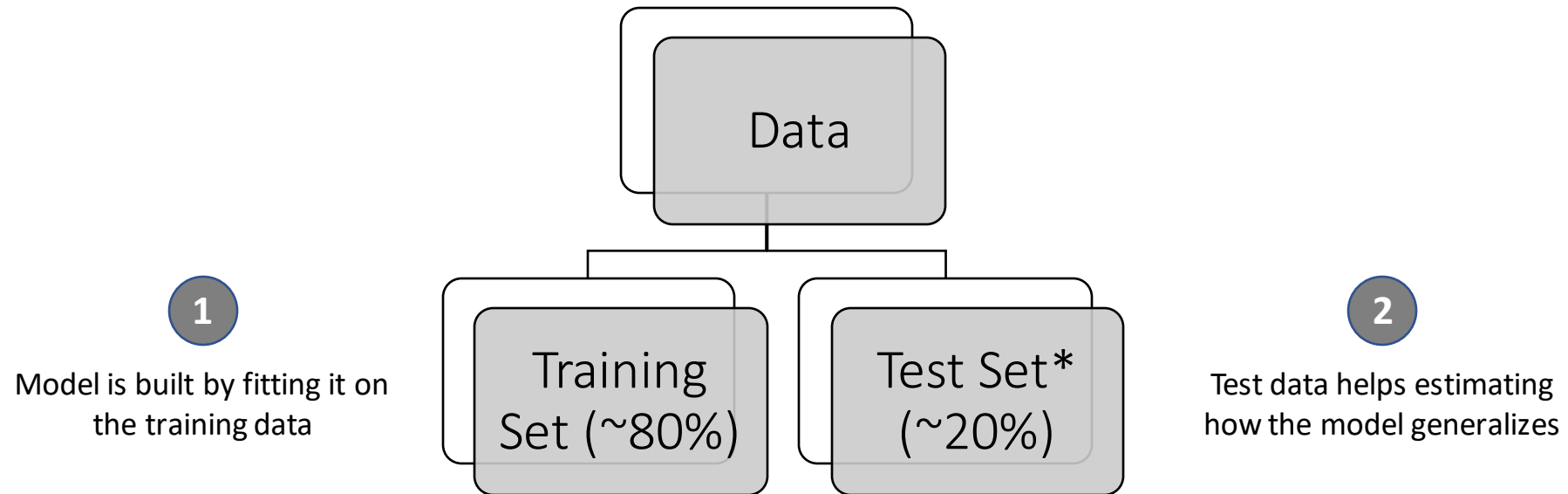
Resolving by

- Selecting more powerful model
- Feeding better features
- Reducing constraints or regularization



Testing and Validating

Ensuring Model Generalizes Well

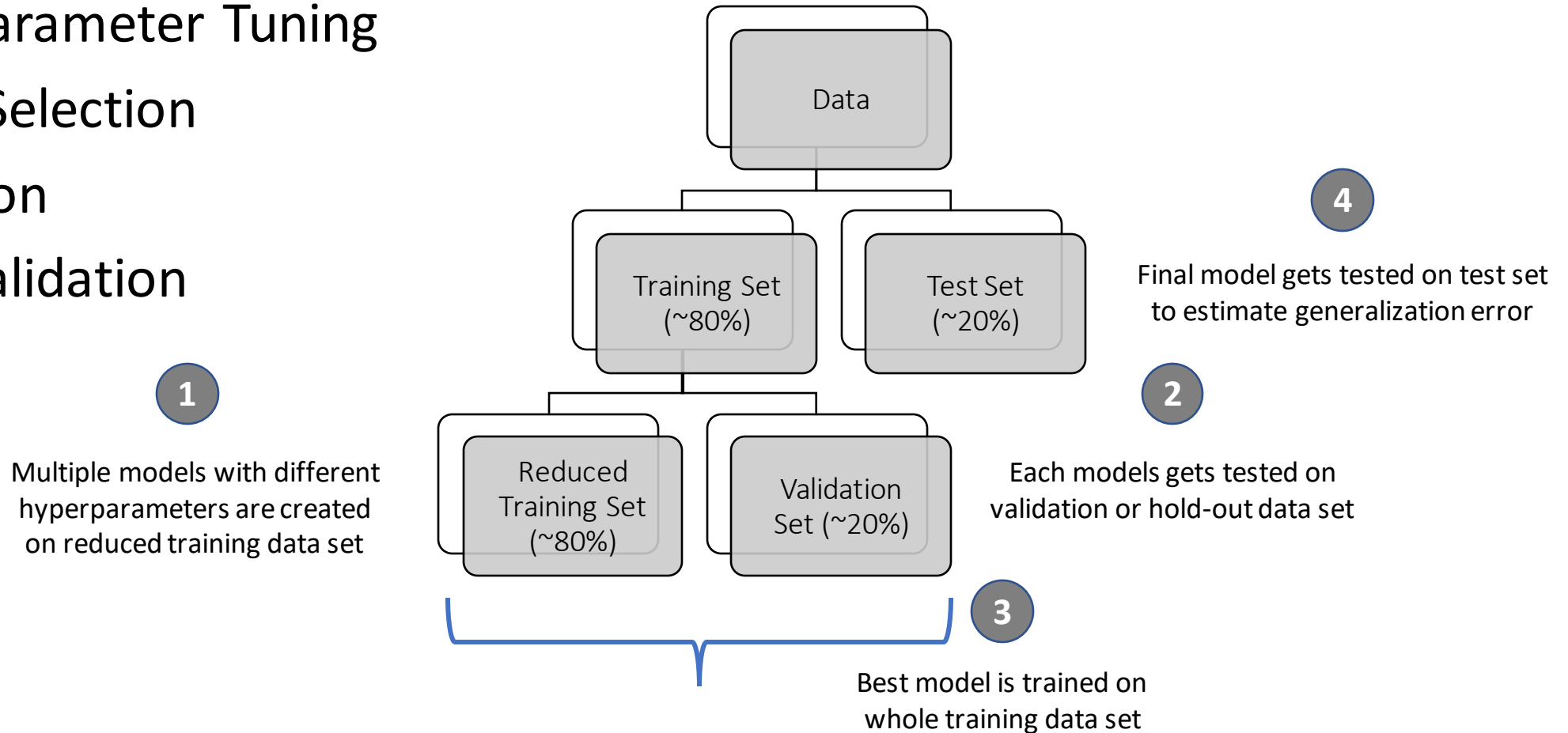


**a.k.a. Hold-out set*

A better model fitted over more representative training data

Testing and Validating (Cont.)

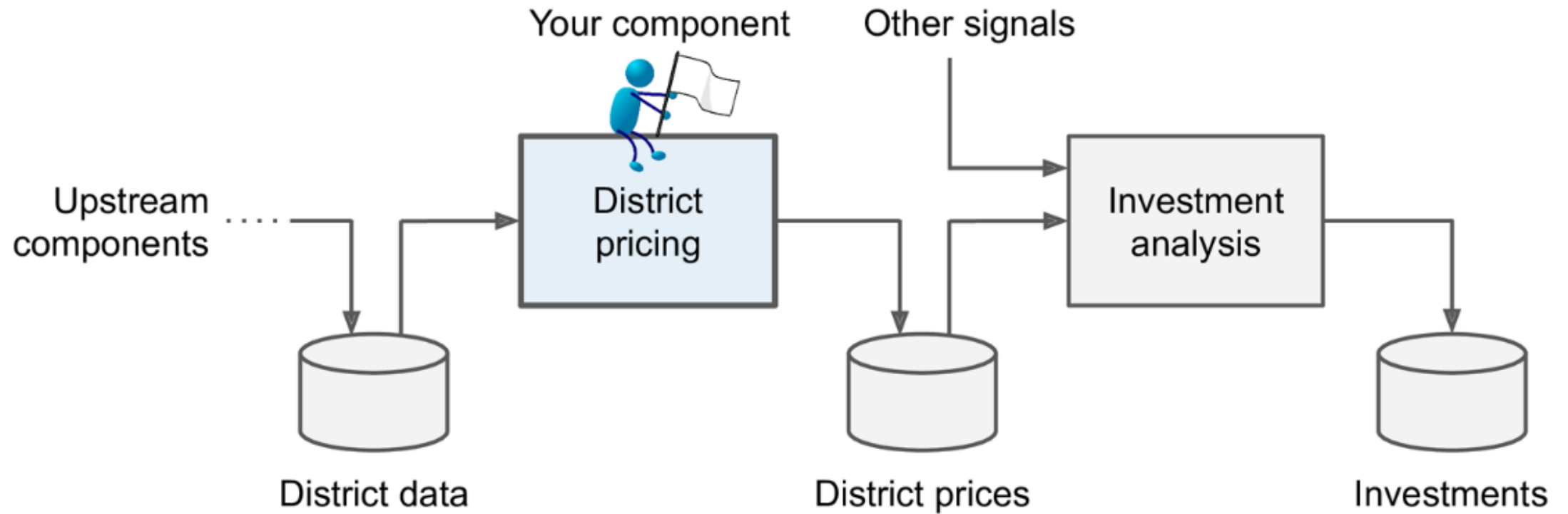
- Hyperparameter Tuning
- Model Selection
- Validation
- Cross-validation



No Free Lunch (NFL) Theorem

- No model that is *a priori* guarantees to be a better one
- Evaluating all models is only way to know which one works best – *Not practical*
- Hence, making *reasonable assumption about data* and *evaluating few reasonable models* is better option

The Pipeline



High-level machine learning pipeline for real estate investment