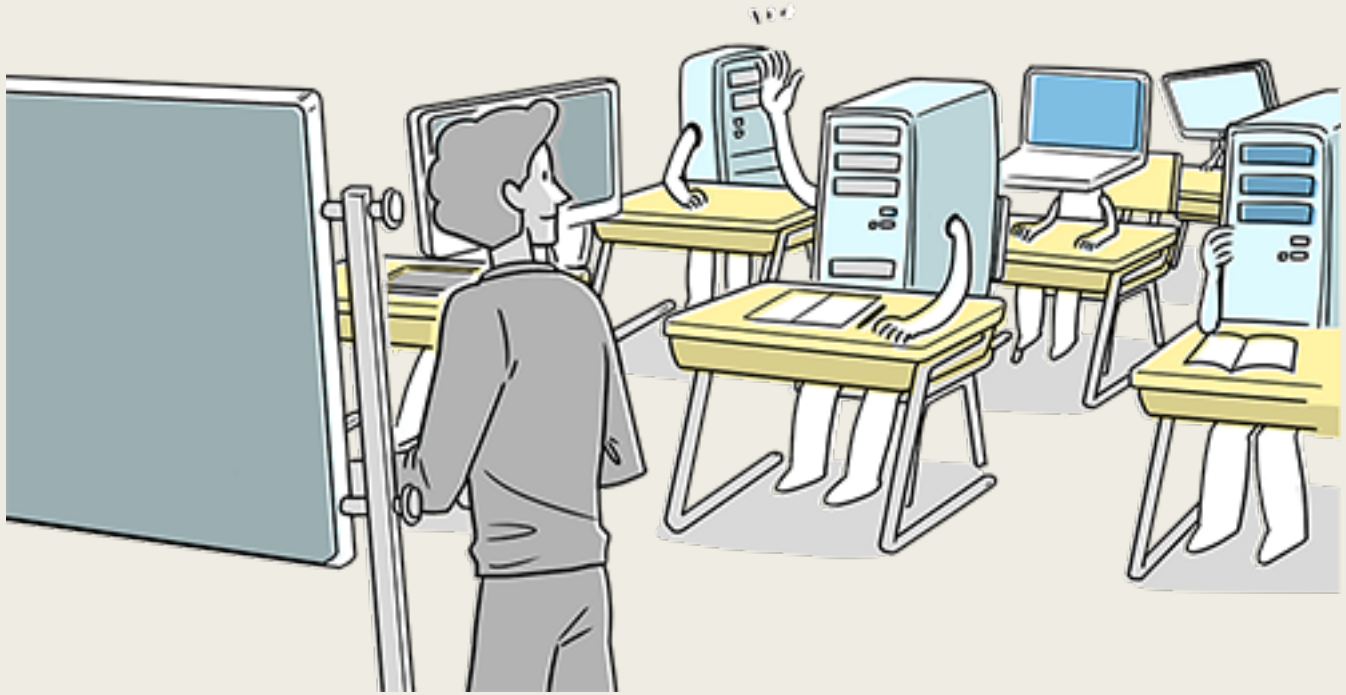
Sorawit Saengkyongam **sorawitj@gmail.com**



INTRODUCTION TO MACHINE LEARNING

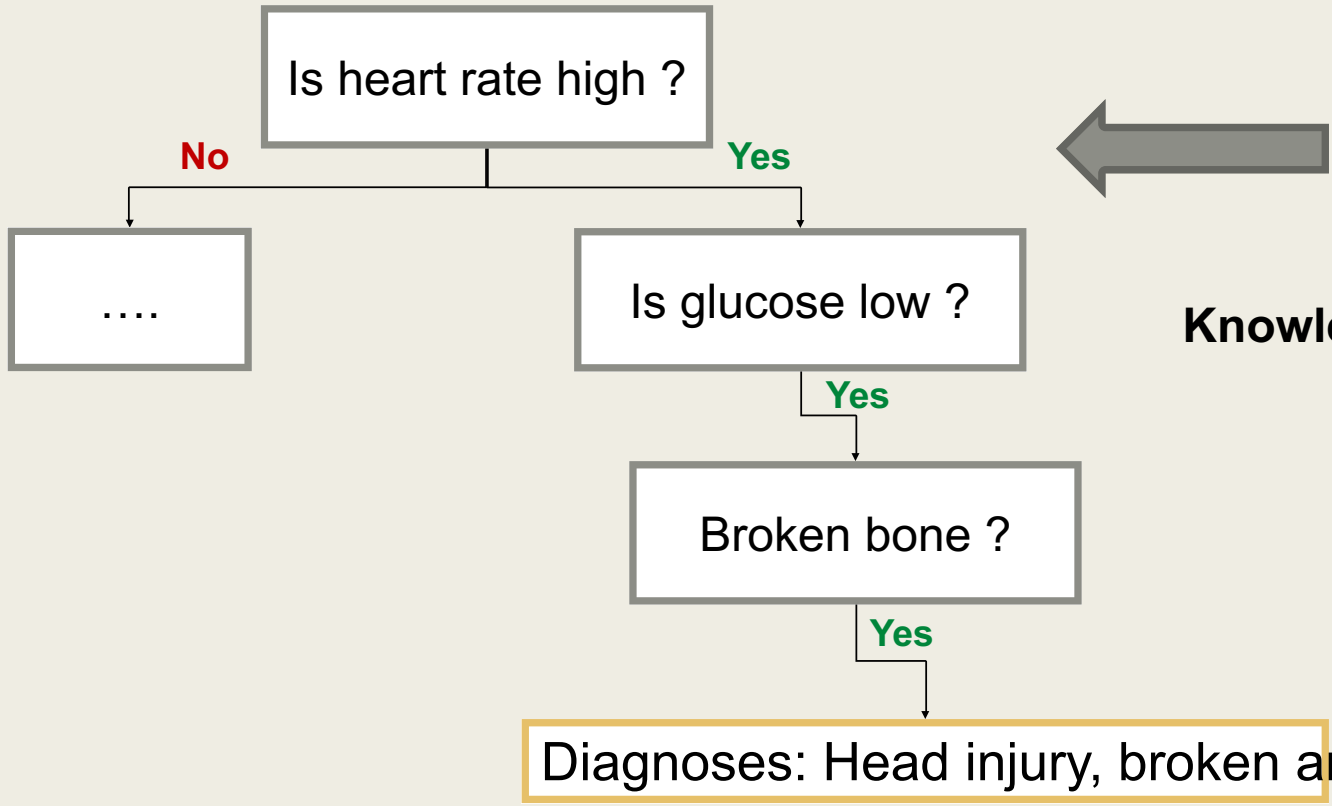
WHAT IS MACHINE LEARNING?

Machine Learning

- A subfield of Artificial Intelligence (AI)
- The field of study that gives computers the ability to learn without being explicitly programmed — Arthur Samuel

What is "NOT" machine learning

AI system for medical diagnosis



Knowledge from an expert

What is machine learning

Can we learn those rules from the data ?

Heart rate	Glucose	Broken bone	….	Diagnoses
High	Low	Yes		Head Injury, Broken arm
Normal	High	No		Diabetes
….	….	….		….

Supervised machine learning: Features **X**

Labels **Y**

Model Output

What is machine learning

- How would you teach someone to ___ ?

- Classify if an email is spam
- Estimate how many likes your post will get
- Recommend a movie to a friend
- Group customers into segments

What is machine learning

- Let’s listen from the world’s expert
- http://ta.virot.me/fb-ai-explainer/

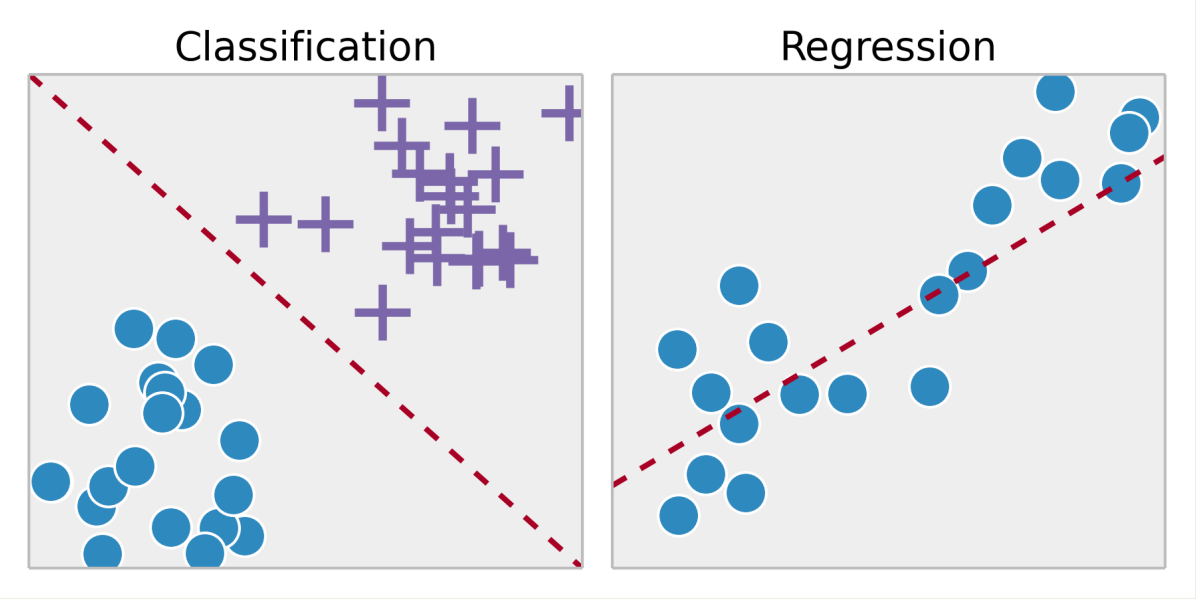
WHAT ARE MACHINE LEARNING USE CASES?

Agoda’s use cases

- Pay-Per-Click (PPC) Bidding Algorithm
- Search and Recommender Systems
- Dynamic Pricing
- Fraud Detection
- Photos Auto-tagging

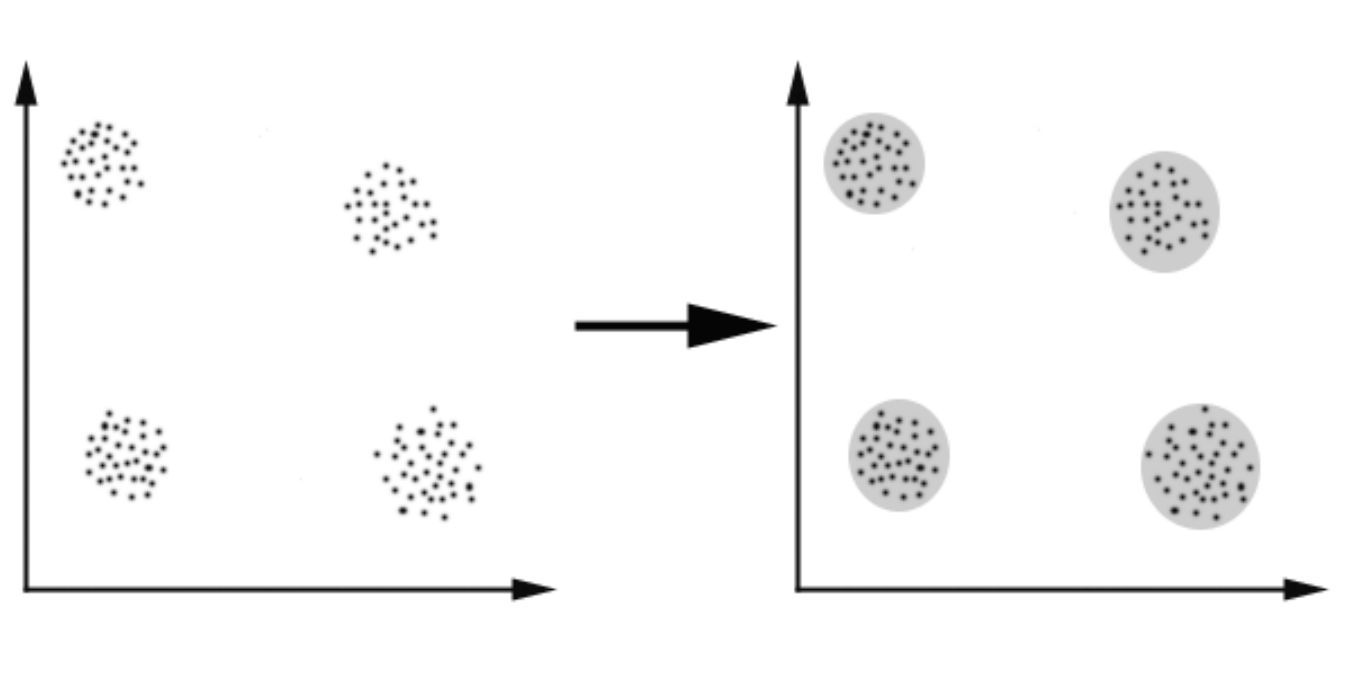
Type of machine learning

- **Supervised Learning**
  - The target values (**labels**) are provided
  - The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.
  - **Classification** (for discrete outcome) and **Regression** (for continuous outcome)
  - Example: Image classification



Type of machine learning

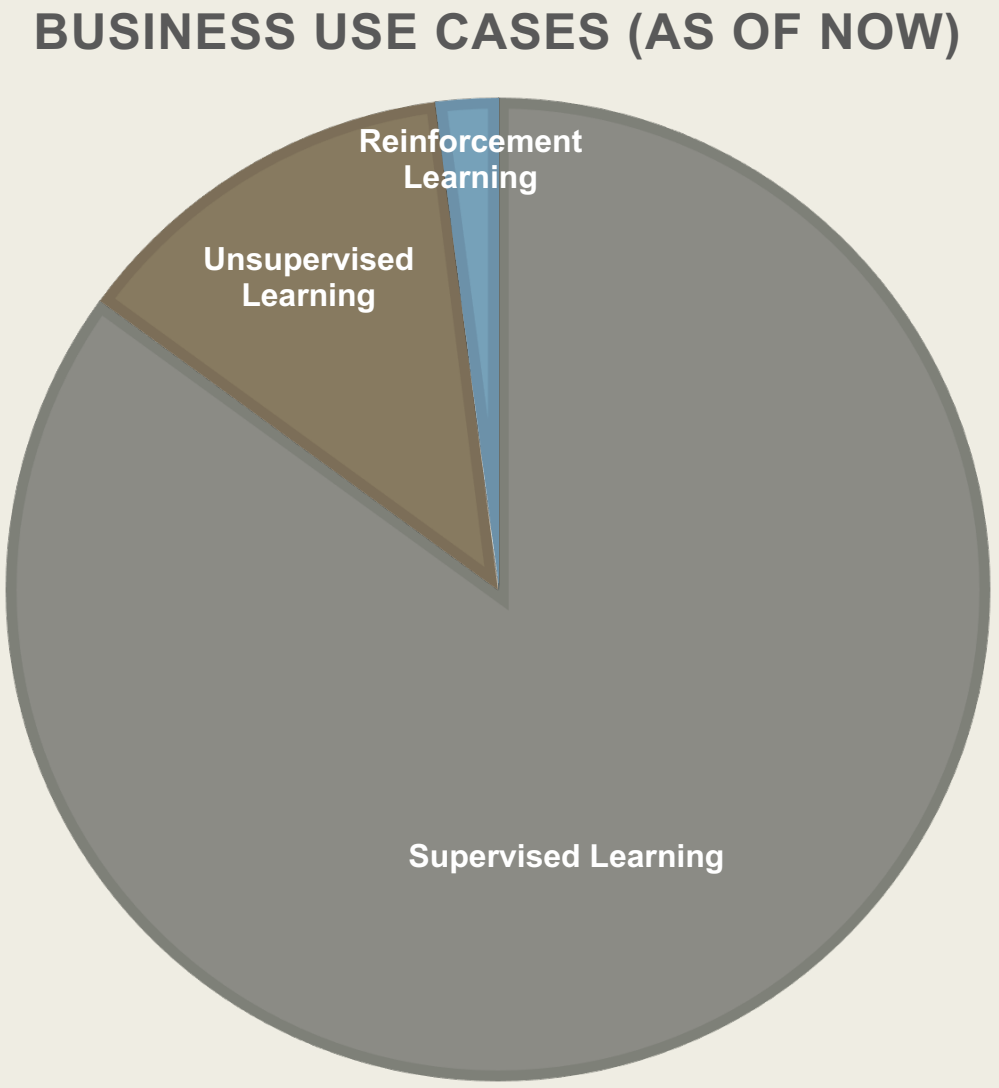
- **Unsupervised Learning**
  - No target values specified
  - The goal is to model the underlying structure or distribution in the data in order to learn more about the data
  - Example: Customer Segmentation

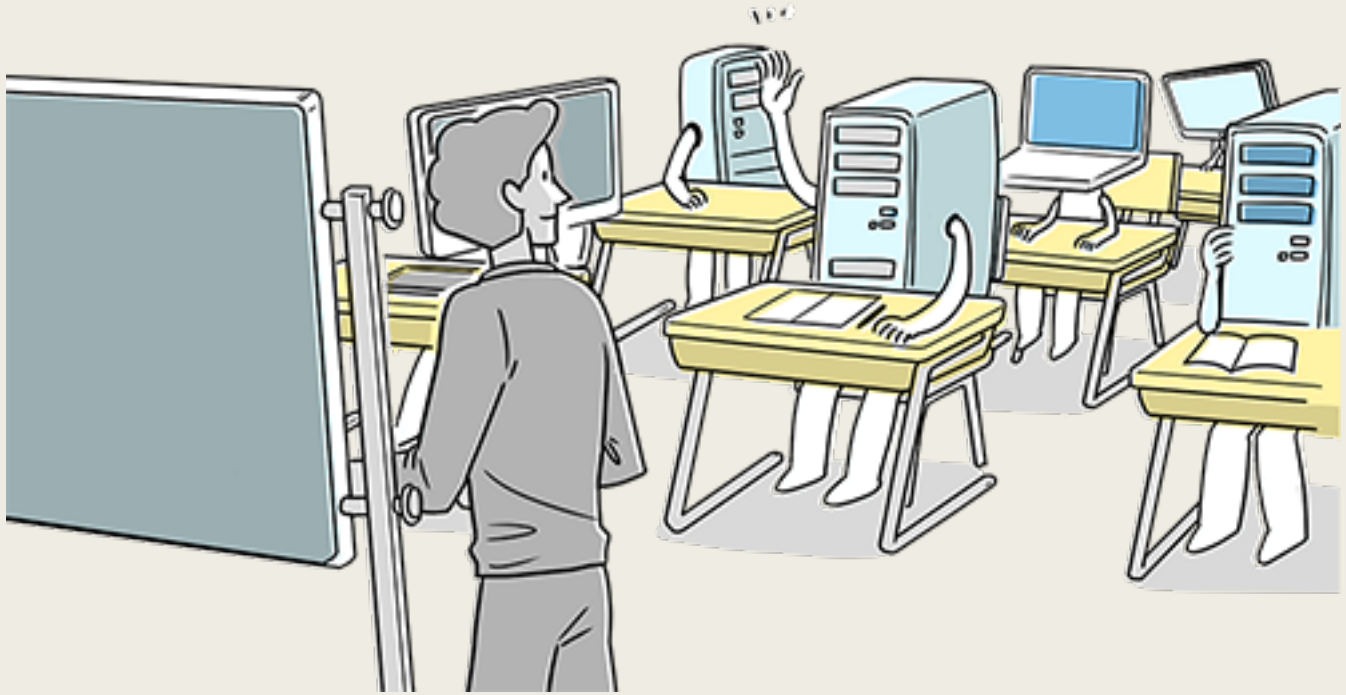


Type of machine learning

- **Reinforcement Learning**
  - The method aims at using observations gathered from the interaction with the environment to take actions that would maximize the reward
  - Example: AlphaGo

Type of machine learning

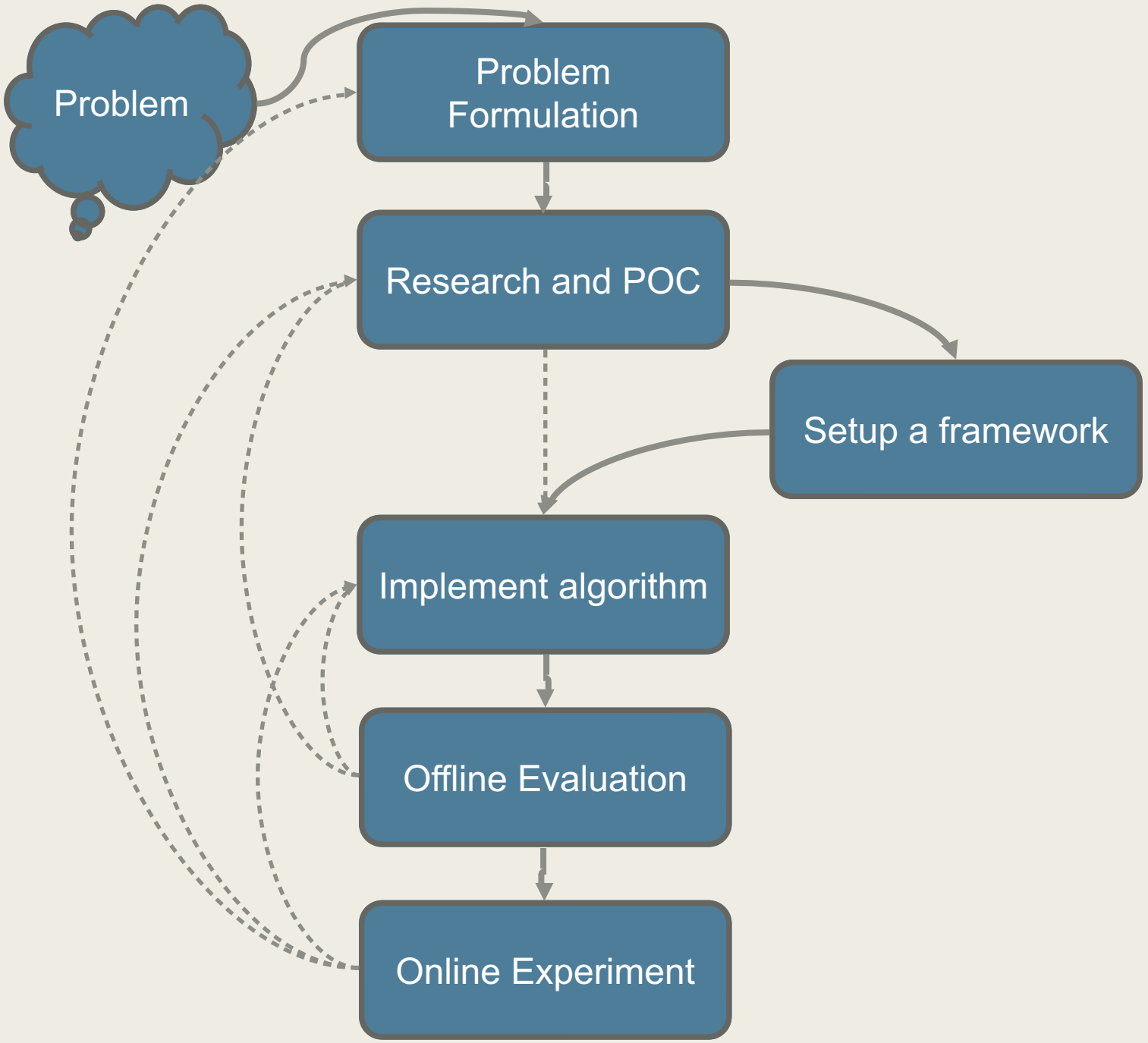




MACHINE LEARNING PROCESS

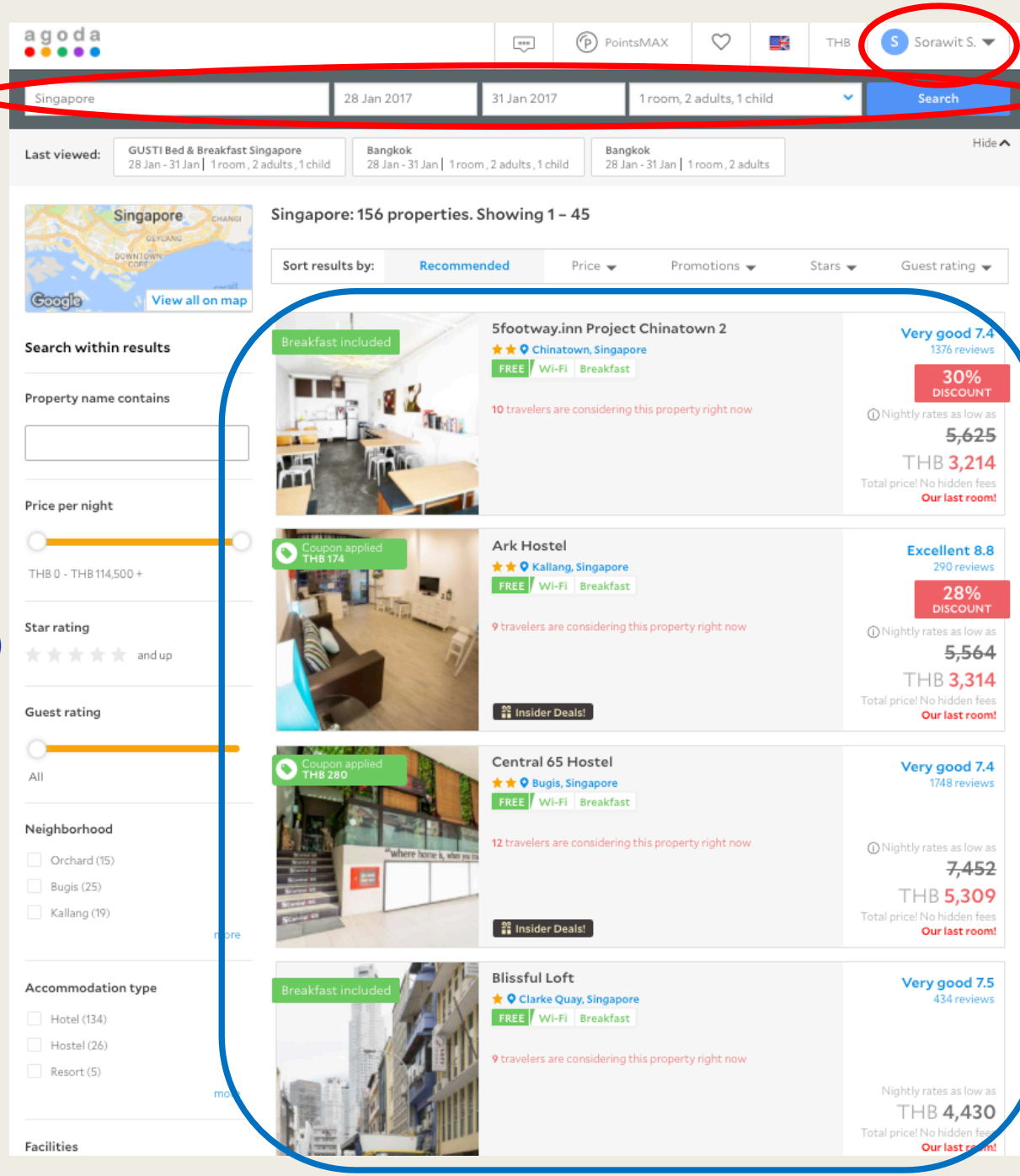
Onboarding Machine Learning

- How do we start adopting machine learning in practice ?
- What is a process for developing ML products ?


Problem
Problem Formulation
Research and POC
Setup a framework
Implement algorithm
Offline Evaluation
Online Experiment

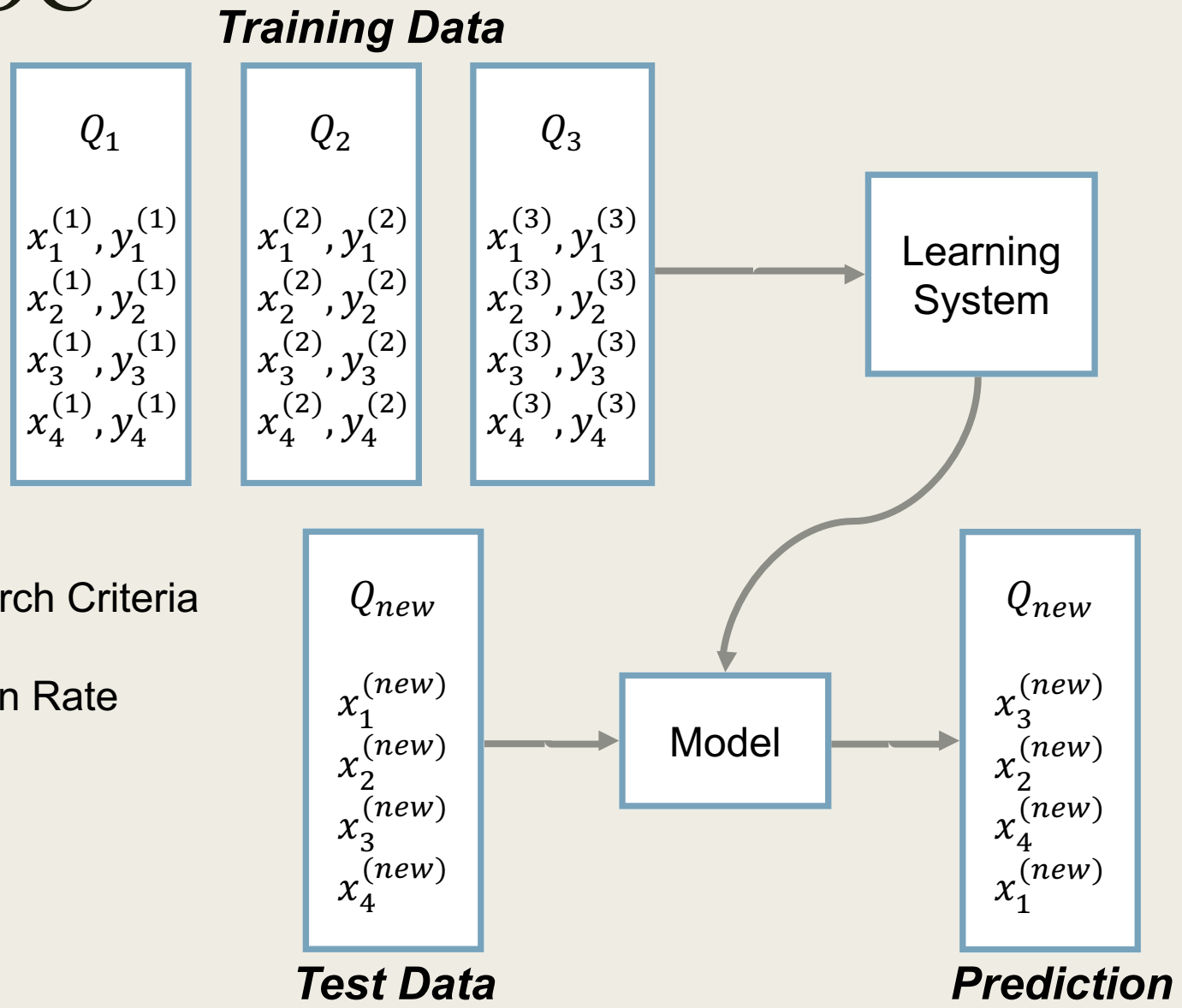
Problem Formulation

- **Objective**
  - *Maximizing reward by providing relevant search results*
- **Input Output**
  - ***Input**: (User, Search Criteria)*
  - ***Output**: {H1, H2, H3, H4 … } (a **permutation** of hotels)*
- **Measure of success**
  - *Uplift in Conversion rate, Click through rate or Profit ?*
- How would we solve this problem ?



Research and POC

- **Avoid** reinventing the wheel
- **Information Retrieval:**
  - ***Learning to rank***



Notation

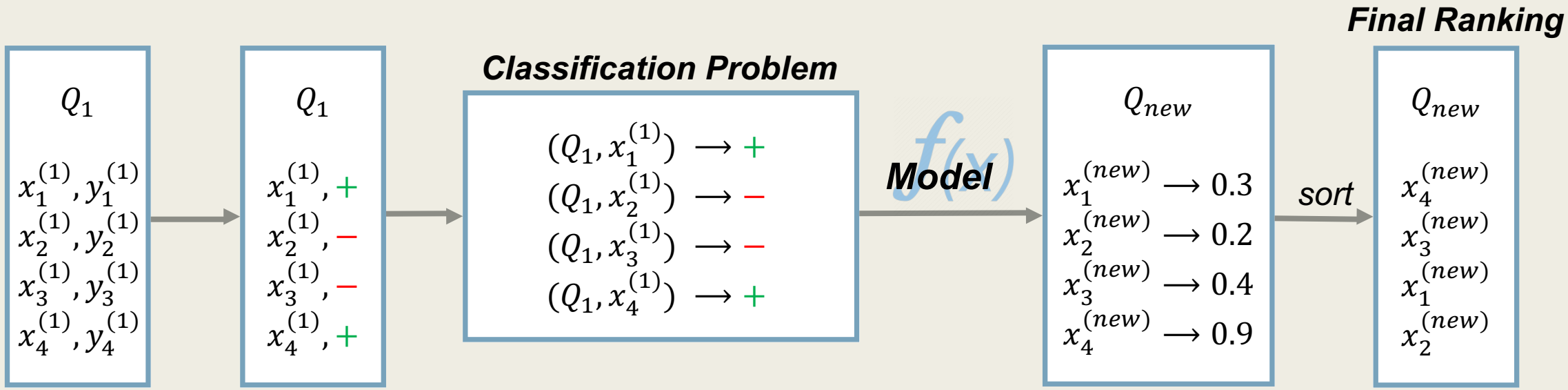
Q → User, Search Criteria

X → Hotel

Y → Conversion Rate

Research and POC

- **Point-wise Approach**
  - Assign a score for each Query-Hotel pair



Research and POC

- **Evaluation metrics: each problem has its own appropriate set of metrics**

Information Retrieval Metrics

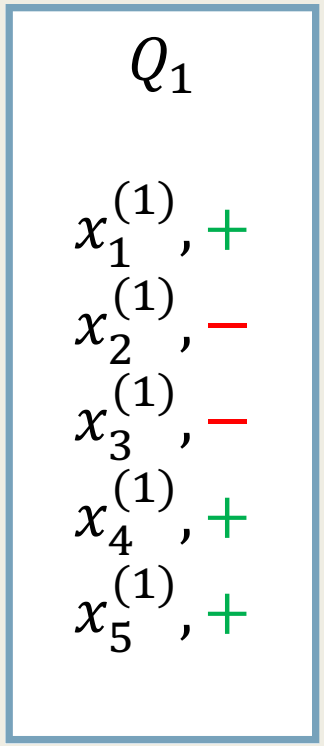
- Precision at K: Compute % relevant in top K, ignores hotels ranked lower than K

Ex.

Prec@3 = 1/3

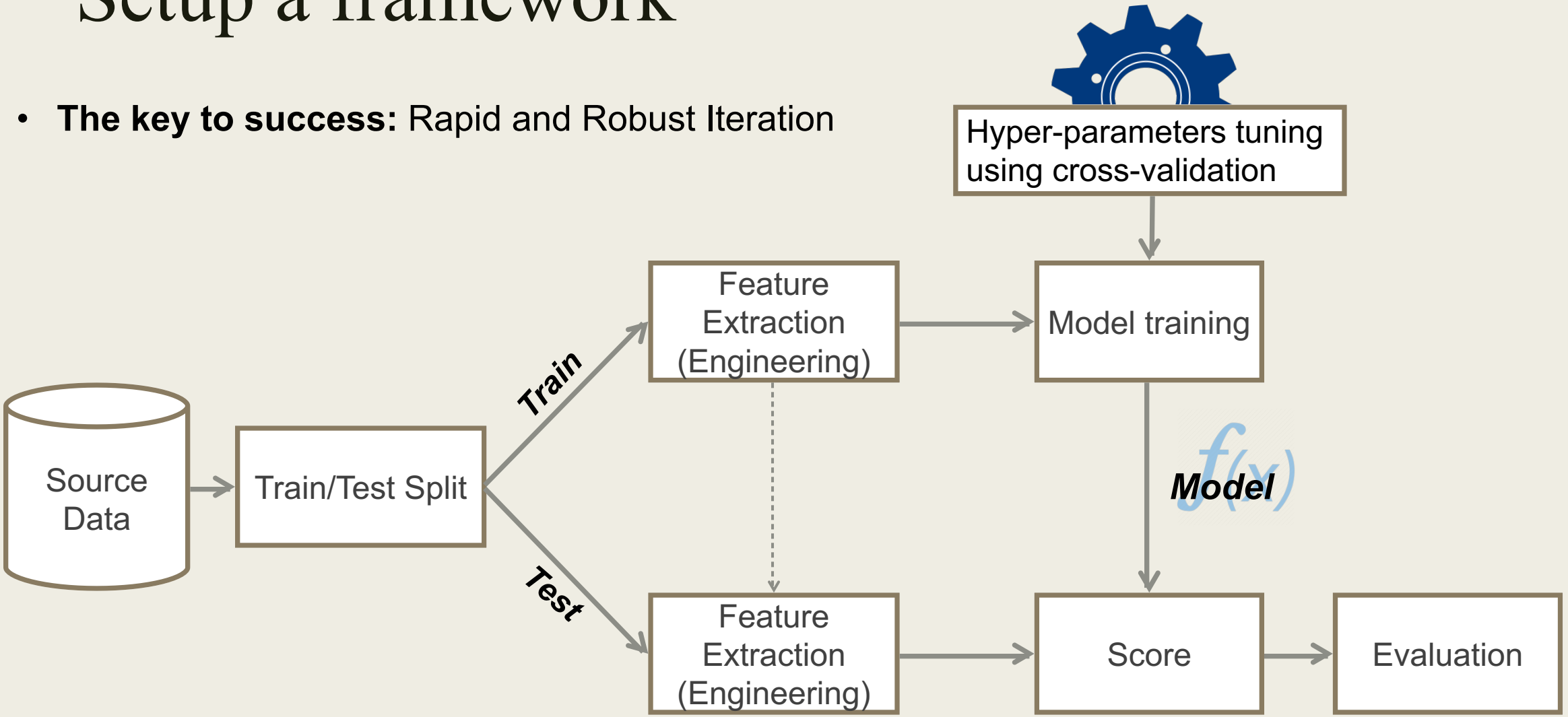
Prec@4 = 2/4

Prec@5 = 3/5



Setup a framework

- **The key to success:** Rapid and Robust Iteration



Implement algorithm

Feature Engineering

- Domain knowledge plays an important role

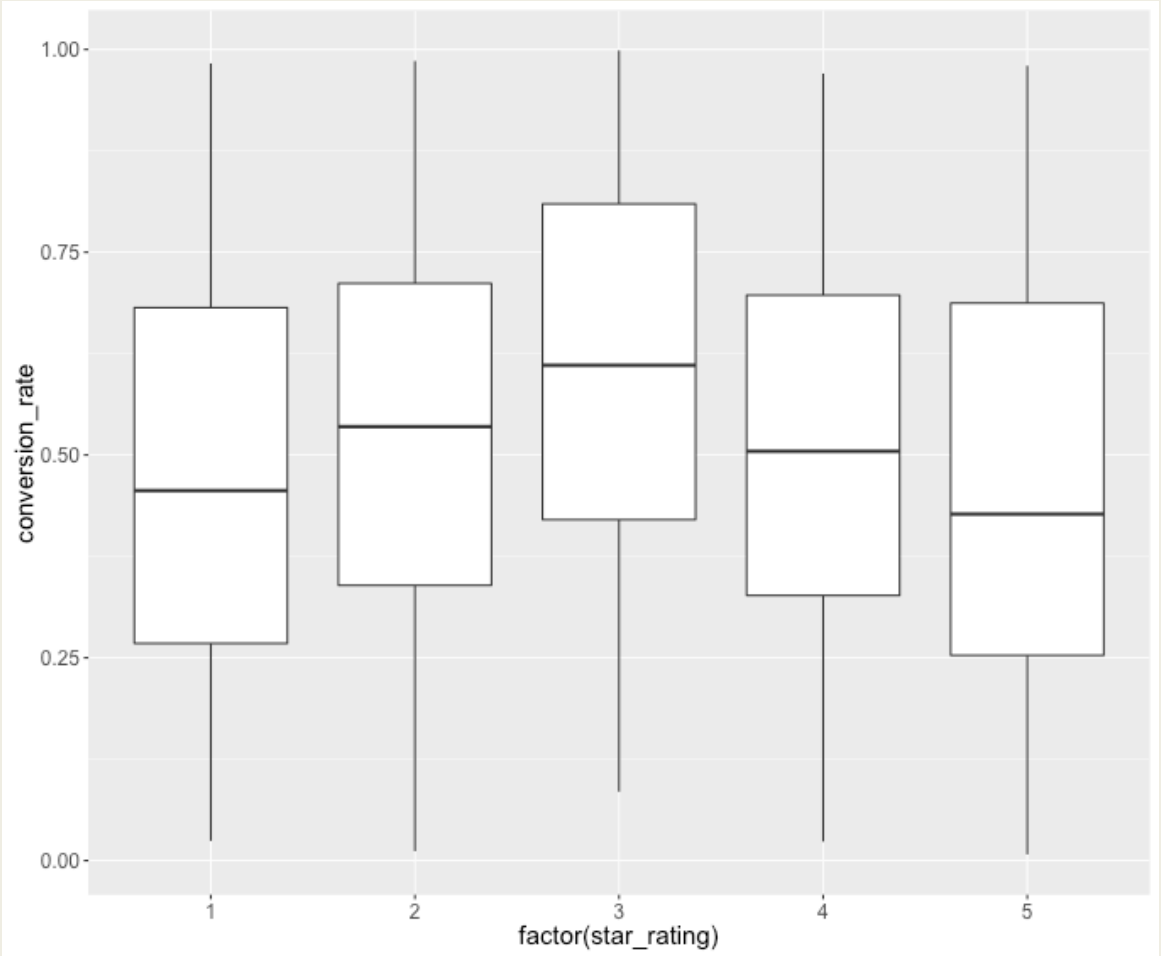
- ***Hotel features***
  - hotel_id, star rating
  - price
  - review score

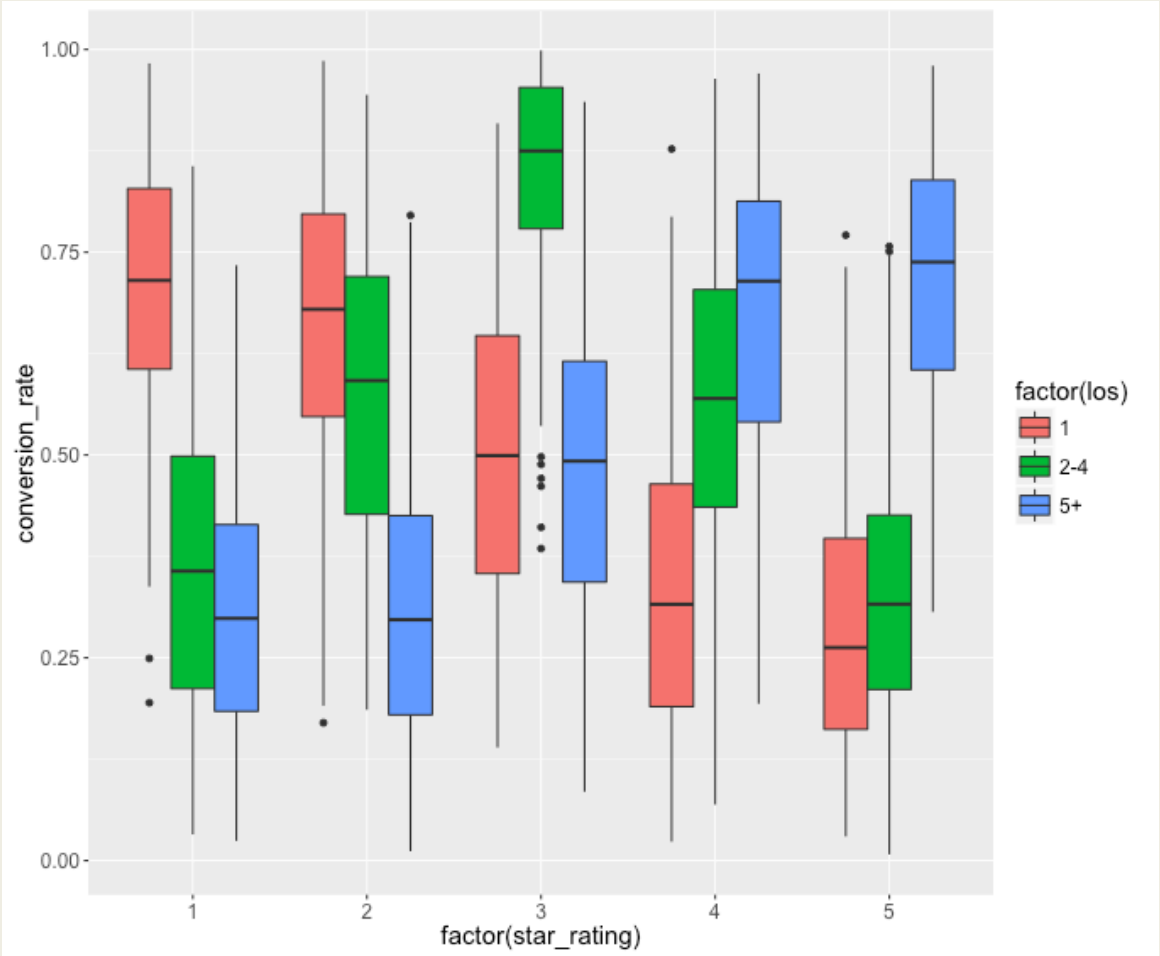
- ***Query features***
  - user_id, length of stay
  - user origin
  - historical booking rate of user

- ***Query-Hotel features (the most important)***
  - abs(historical bookings price of user – hotel price)
  - length of stay * star rating
  - hotel_id * user_id

Implement algorithm

- Feature exploration: ***identify potential predictors (features)***



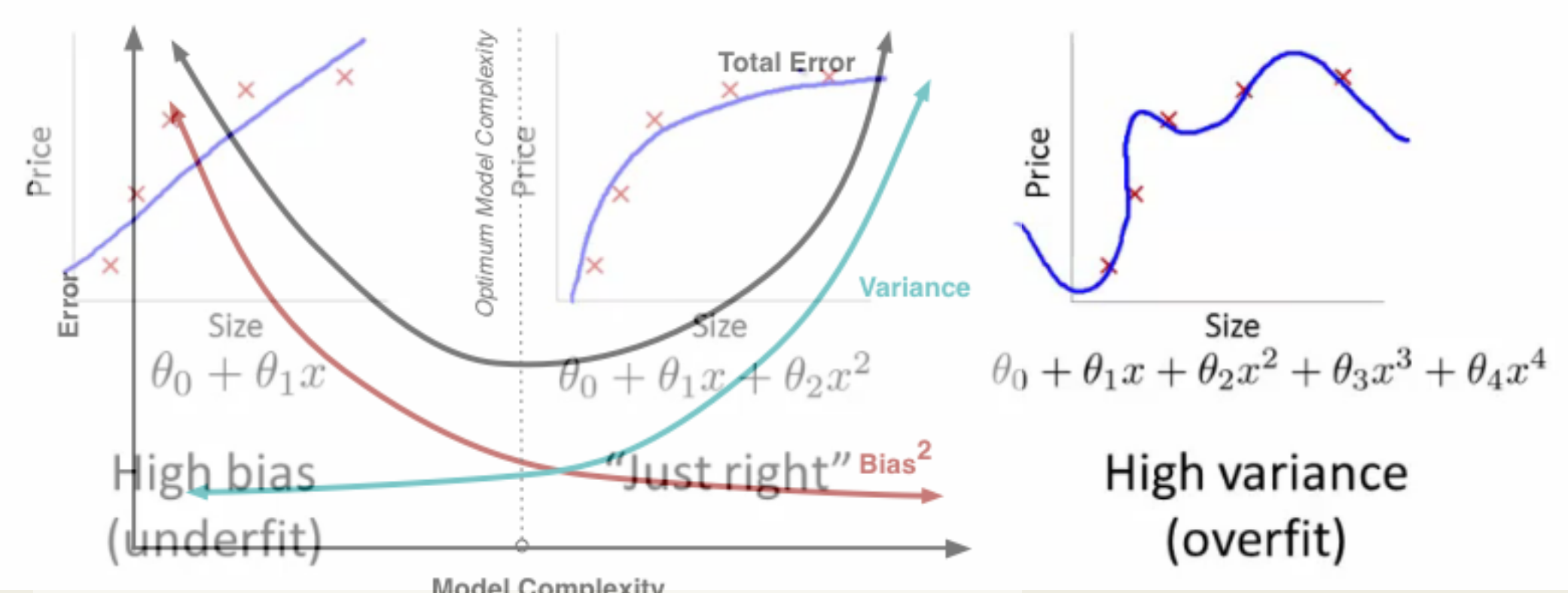


Implement algorithm

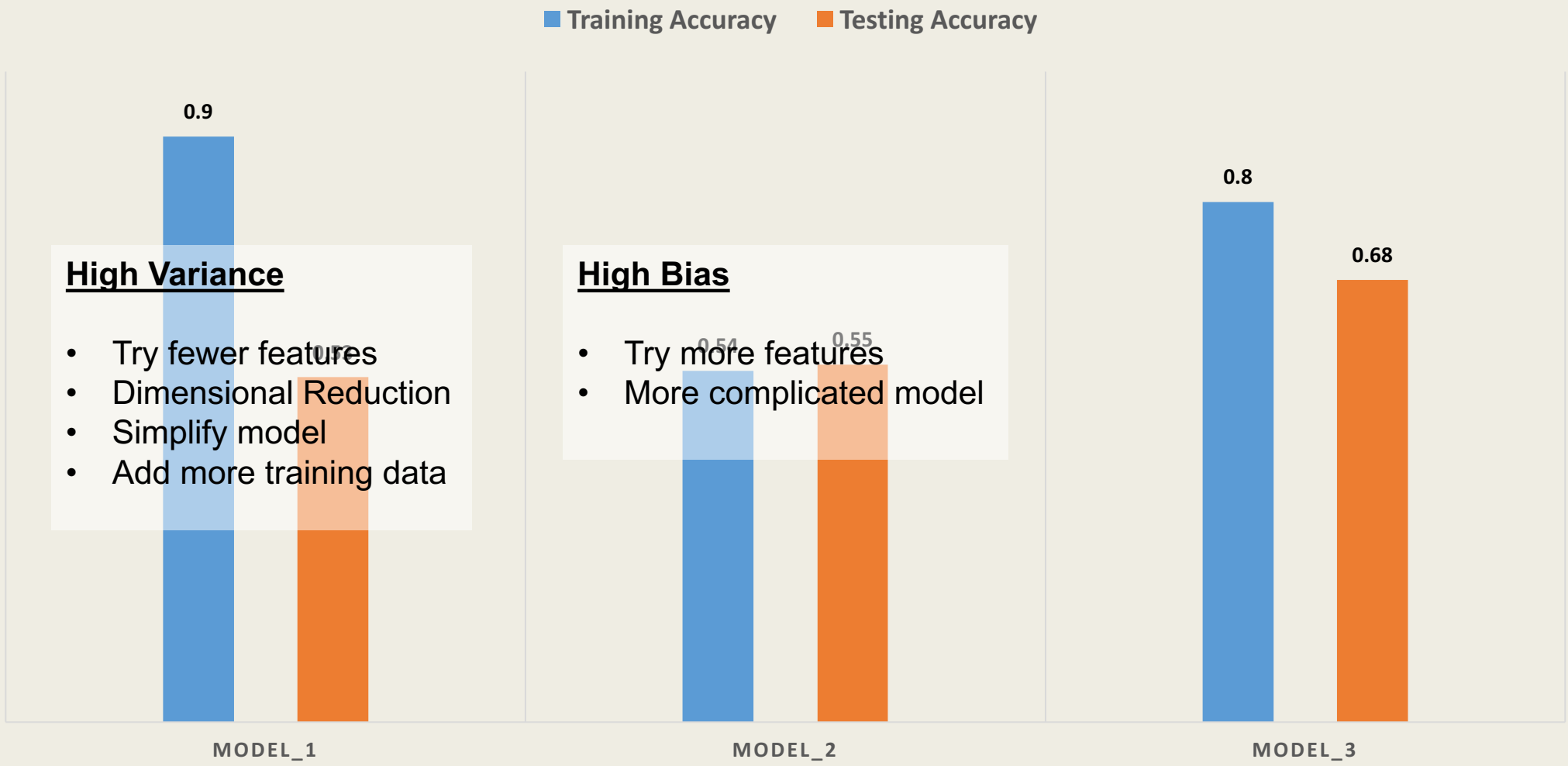
- Pick your candidate algorithms
  - **Always always start with Baseline**: Rank by popularity
  - Generalized Linear Model
  - SVM
  - Gradient Tree Boosting
  - Factorization Machine
  - Neural Network

Implement algorithm

- **Model selection tip**: understand **bias-variance** trade off



Implement algorithm



High Variance

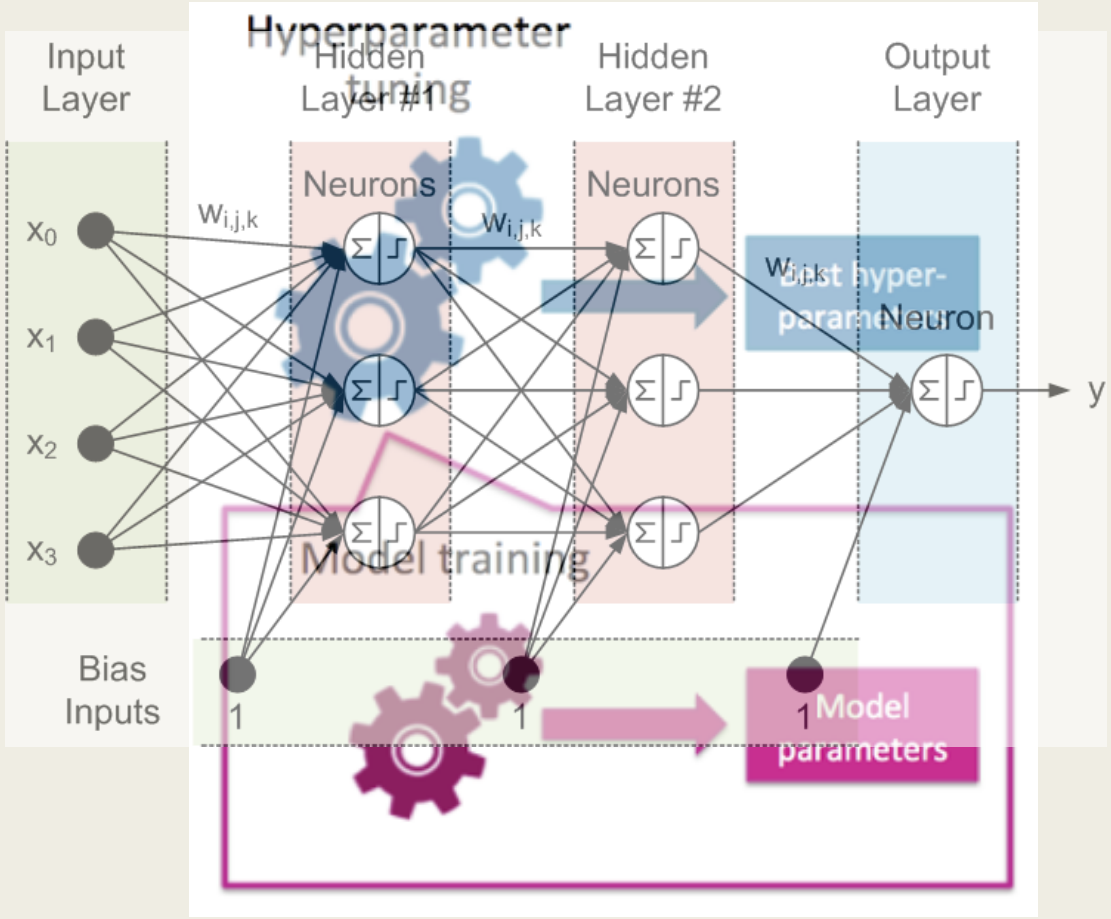
- Try fewer features
- Dimensional Reduction
- Simplify model
- Add more training data

High Bias

- Try more features
- More complicated model

Implement algorithm

- **Super important:** Understand hyper-parameters of your picked algorithms



Model Parameters

- Parameters that we optimize during training process

Hyper-parameters

- Parameters that express “higher-level” properties of the model such as its complexity

*It actually boils down to the **bias-variance** trade-off*

Tuning hyper-parameters

- Grid search with cross-validation
- Bayesian Optimization

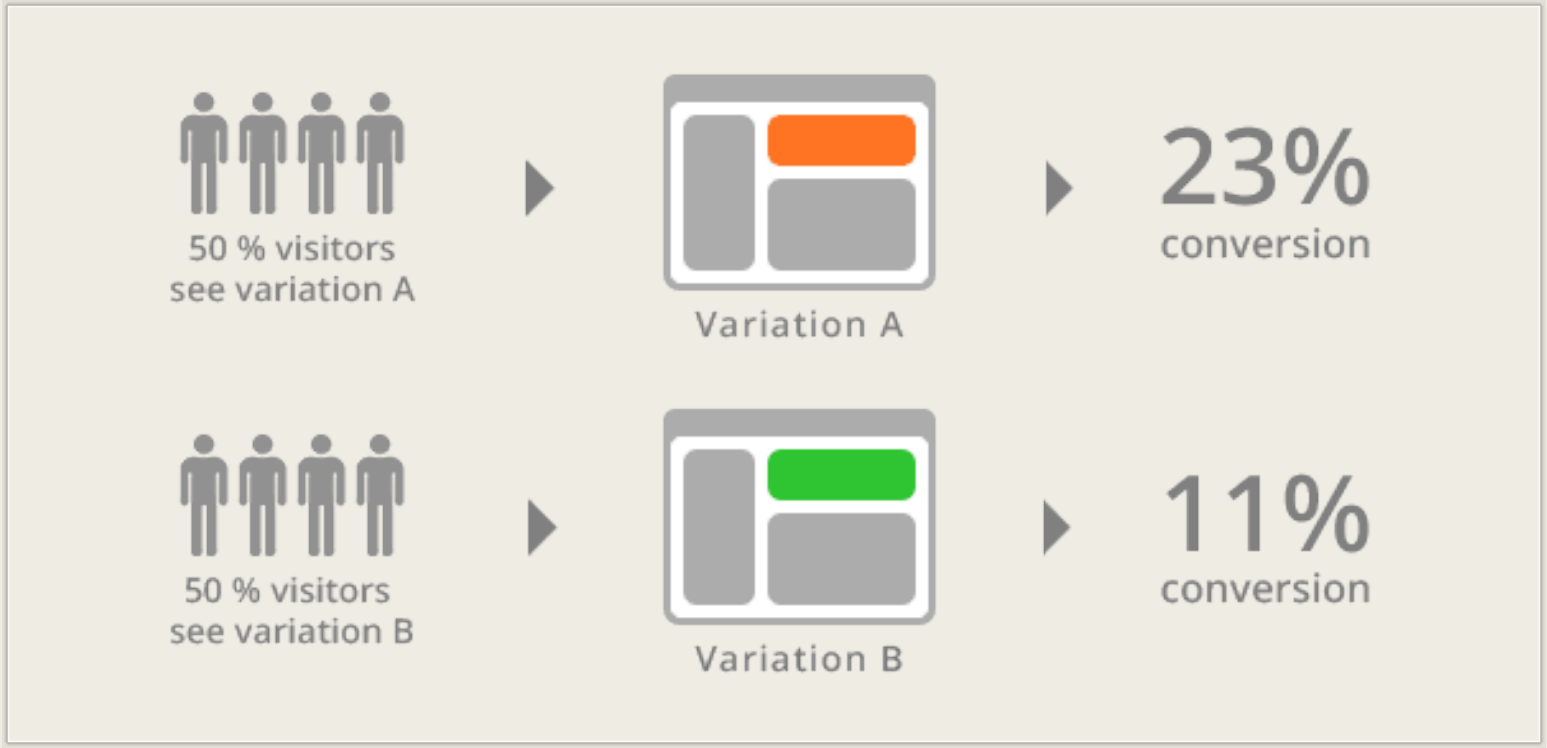
Offline Evaluation (Back testing)

- Remember that what only matters is **“Generalized Error”**
  - Train/Test splitting logic is very crucial
  - Make sure there is no information leakage

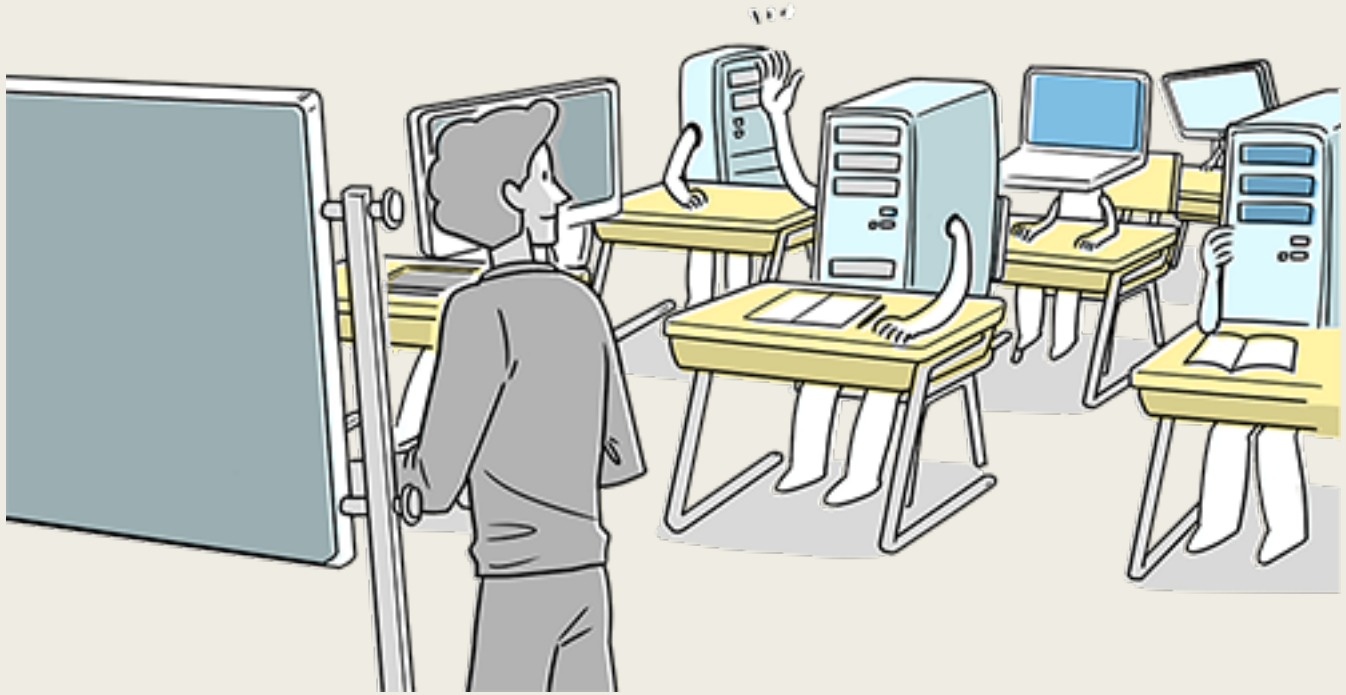
- Defining the right **metrics**
  - Indication to make decision on follow-up online experiments
  - A critical issue is how offline metrics correlate with online test results

Online Experiment

- A/B Testing



- Feedback Loop
  - *Observe how users interact with your product and make use of that data*

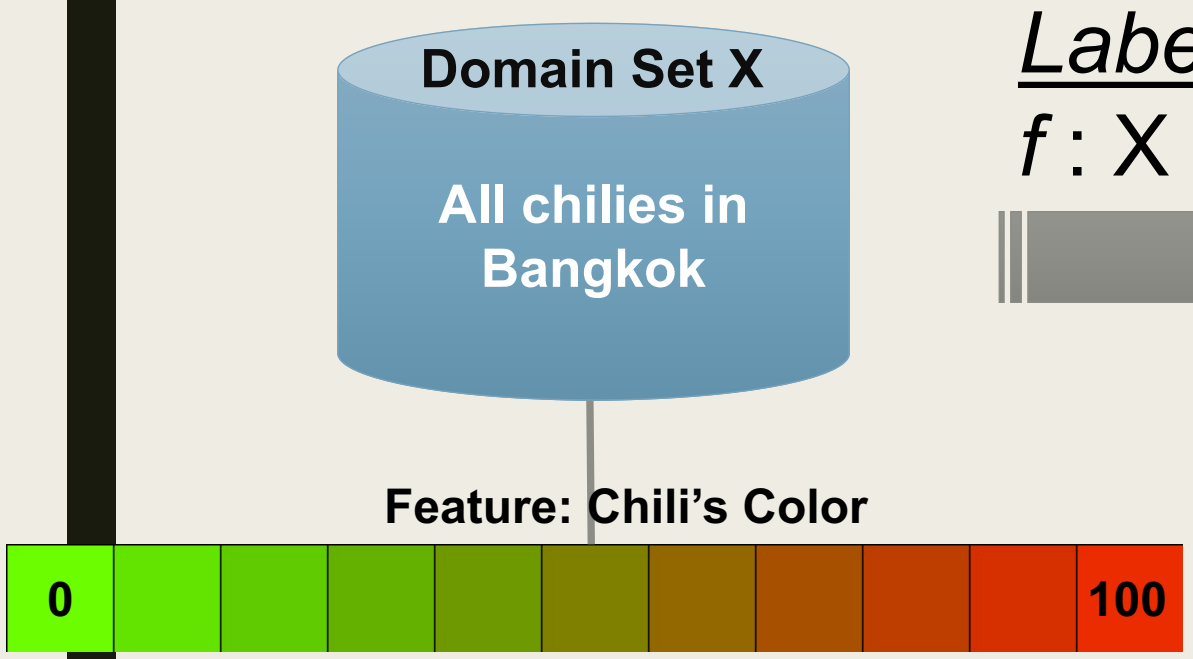


INTRODUCTION TO SUPERVISED LEARNING

Let’s start with example

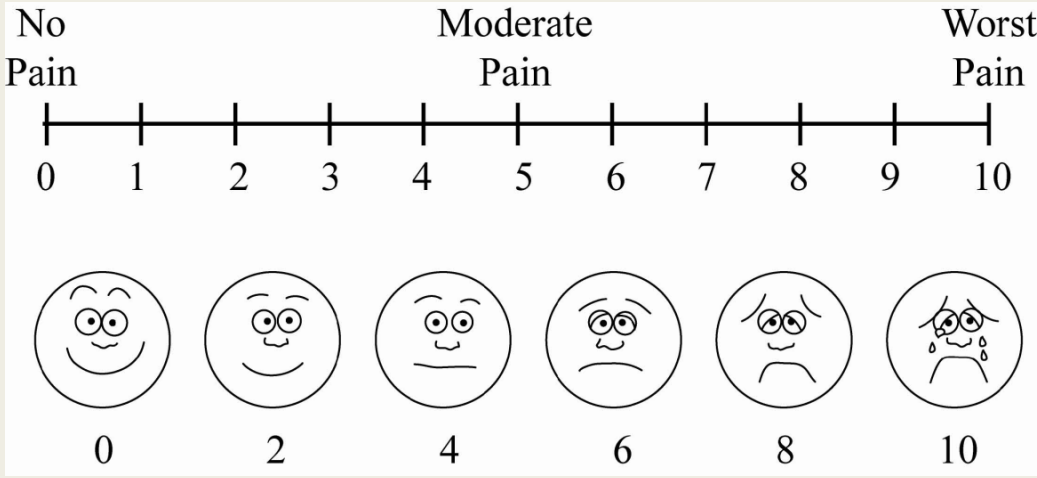
LOVE

Learning Framework



Labeling function

$f : X \rightarrow Y$

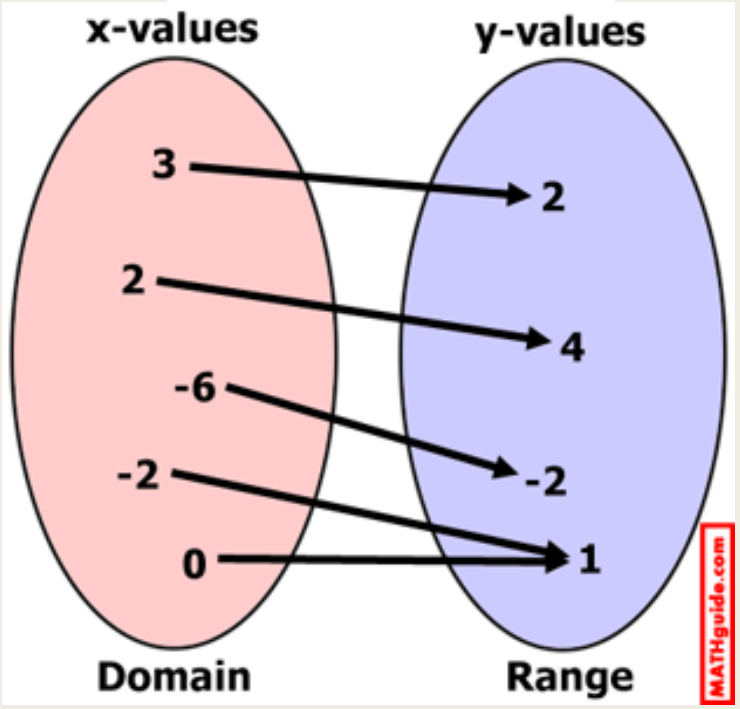


Learning Framework

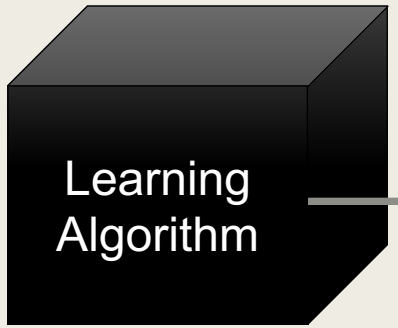
SUPERVISED LEARNING IS ALL ABOUT FINDING THE **LABELING FUNCTION**

What people think I do

What I actually do



Learning Framework



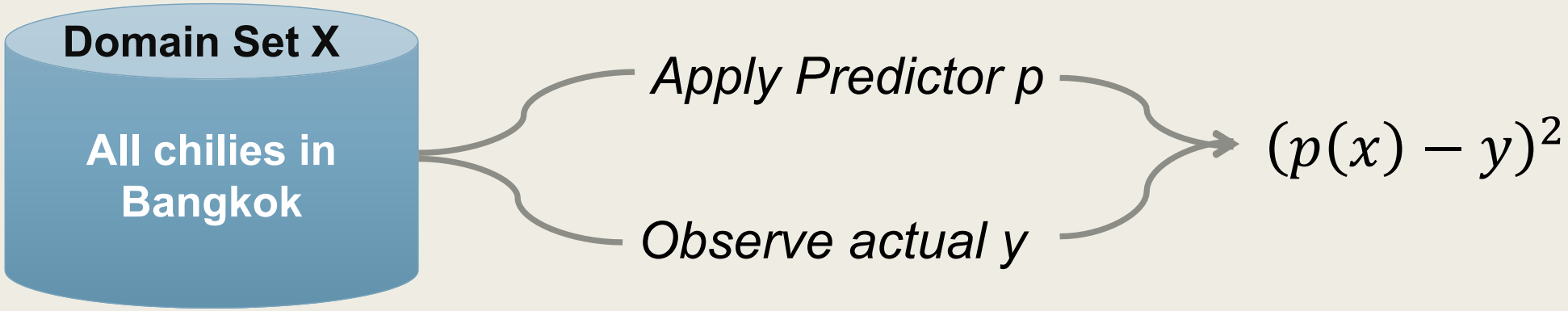
Predictor

$p: X \rightarrow Y$ best represent

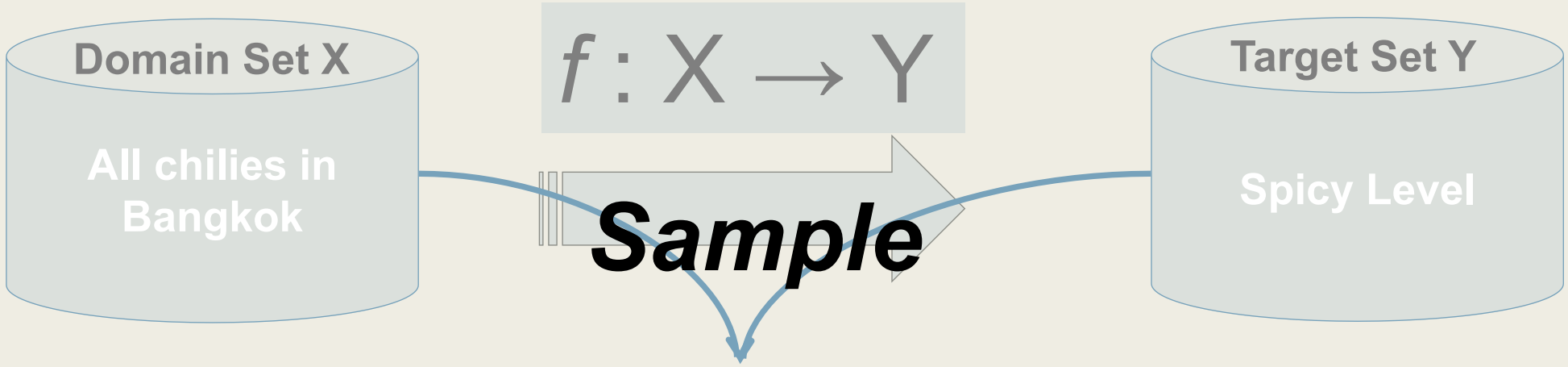
Labeling function f

Measure of success: Generalized Error

$$E_{(x,\,y)\sim D}\,(p(x)-y)^2$$



Learning Framework



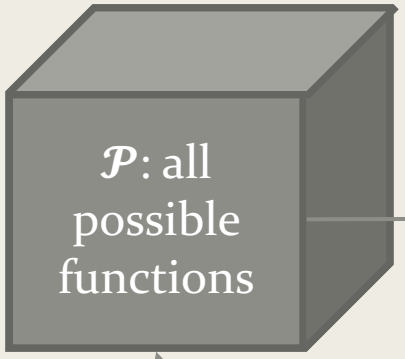
Training Data	
Color (X)	**Spicy Level (Y)**
2	2
5	2
12	3
37	5
48	6
67	6
88	7
95	9

Empirical Risk Minimization

Define: $\boldsymbol{\mathcal{P}}$ *as all possible functions mapping from* $X \rightarrow Y$

Find best p in $\boldsymbol{\mathcal{P}}$ that has the lowest Error on ***Training Data***

$$Min_{p \in \mathcal{P}} \sum_{i=1}^{n} (p(x) - y)^2$$



Pick p

Training Data	
Color (X)	**Spicy Level (Y)**
2	2
5	2
12	3
37	5
48	6
67	6
88	7
95	9

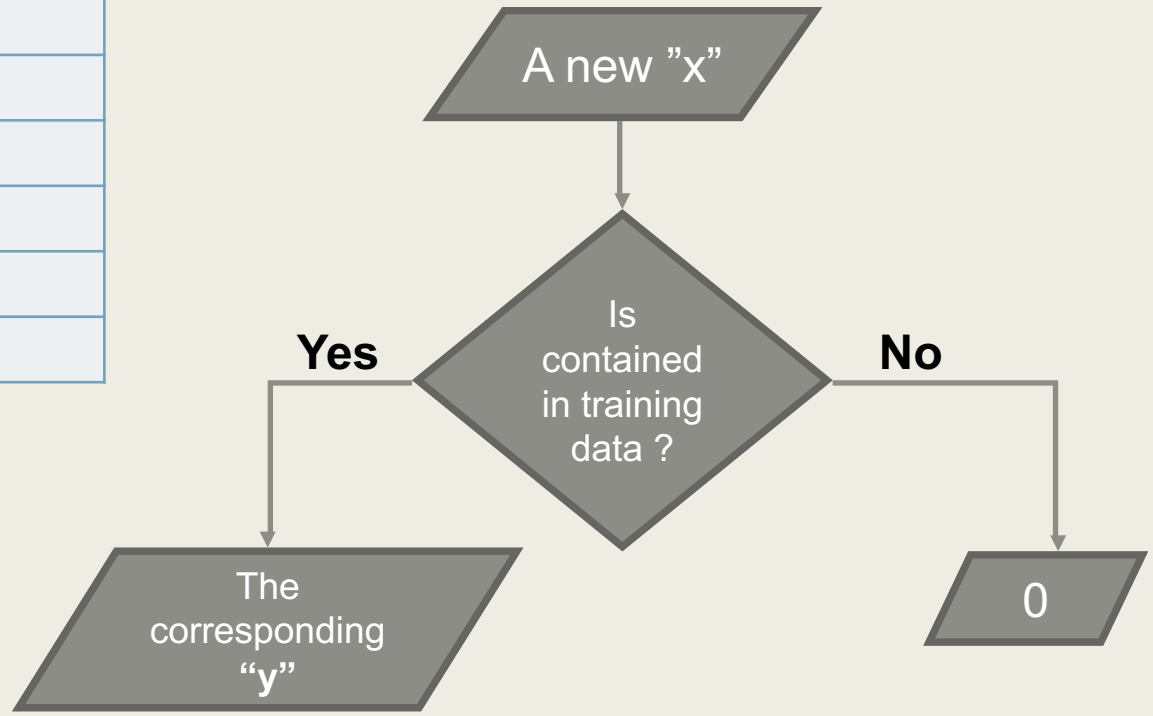
Calculate

$$\sum_{i=1}^{n} \underline{\textit{Training Error}} \; (p(x) - y)^2$$

Empirical Risk Minimization

Training Data	
Color (X)	**Spicy Level (Y)**
2	2
5	2
12	3
37	5
48	6
67	6
88	7
95	9

$$p(x) = \begin{cases} y_i, & if\ x = x_i\ for\ some\ i \\ 0, & otherwise \end{cases}$$



Empirical Risk Minimization

Training Error

$$\sum_{i=1}^{n} (p(x) - y)^2 = 0$$

Generalized Error

HUGE!!

OVER FITTING

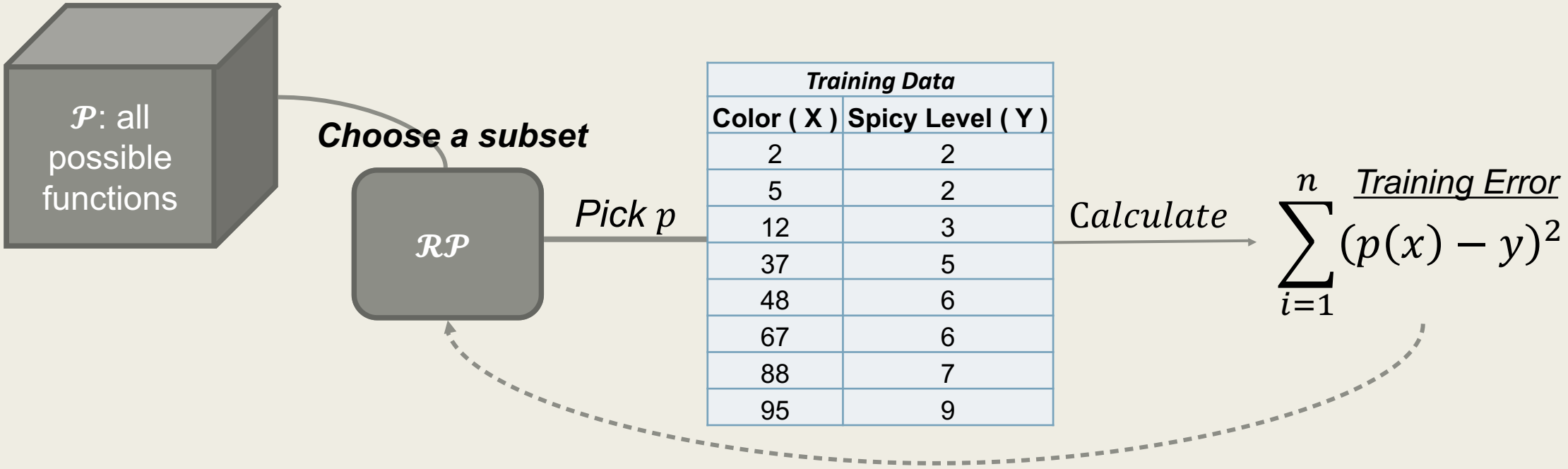
Bias (Assumption)

Apply the ERM learning rule over a **restricted** set

Define: $\mathcal{RP}$ *as a chosen subset of* $\mathcal{P}$

Find best p in $\mathcal{RP}$ that has the lowest Error on $Training\ Data$

$$Min_{p\ \epsilon\ \mathcal{RP}} \sum_{i=1}^{n} (p(x) - y)^2$$



Linear Regression

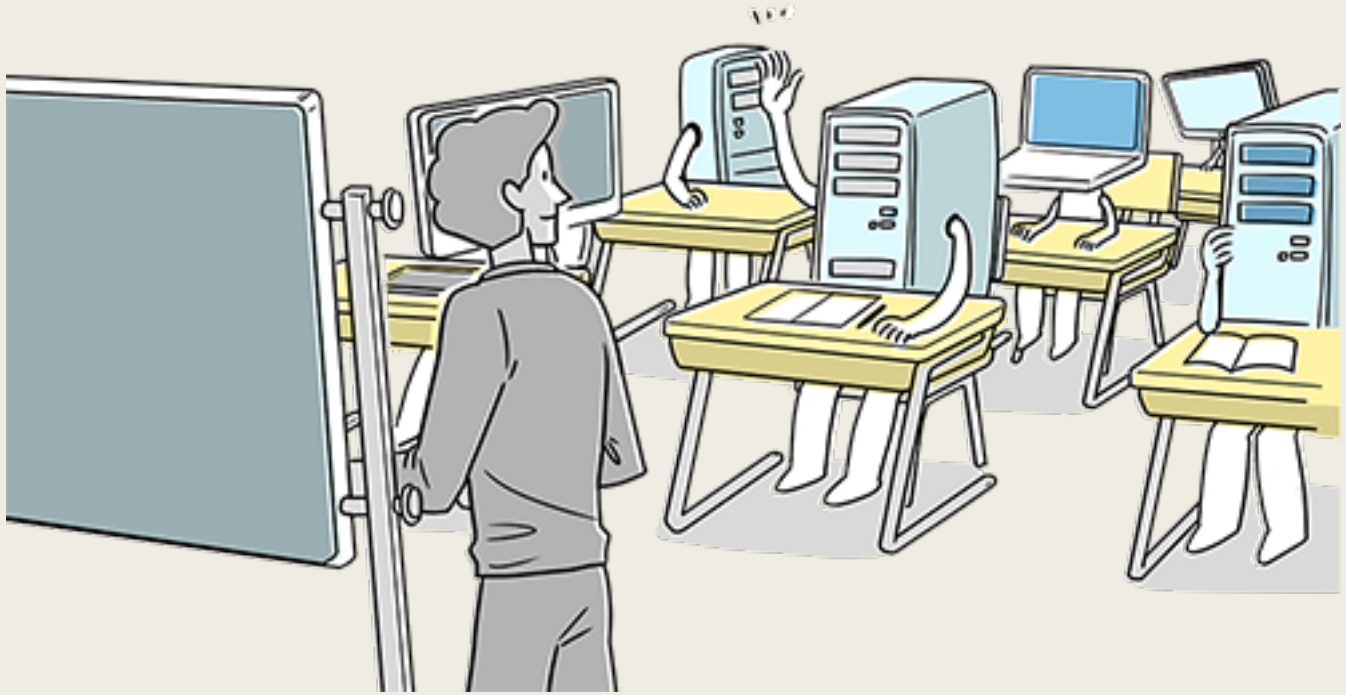
Pick $\mathcal{RP}$ as: $p(x) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} x$, so we only search for best $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ value

Training Error

$$\sum_{i=1}^{n} (p(x) - y)^2 = Small$$

Generalized Error

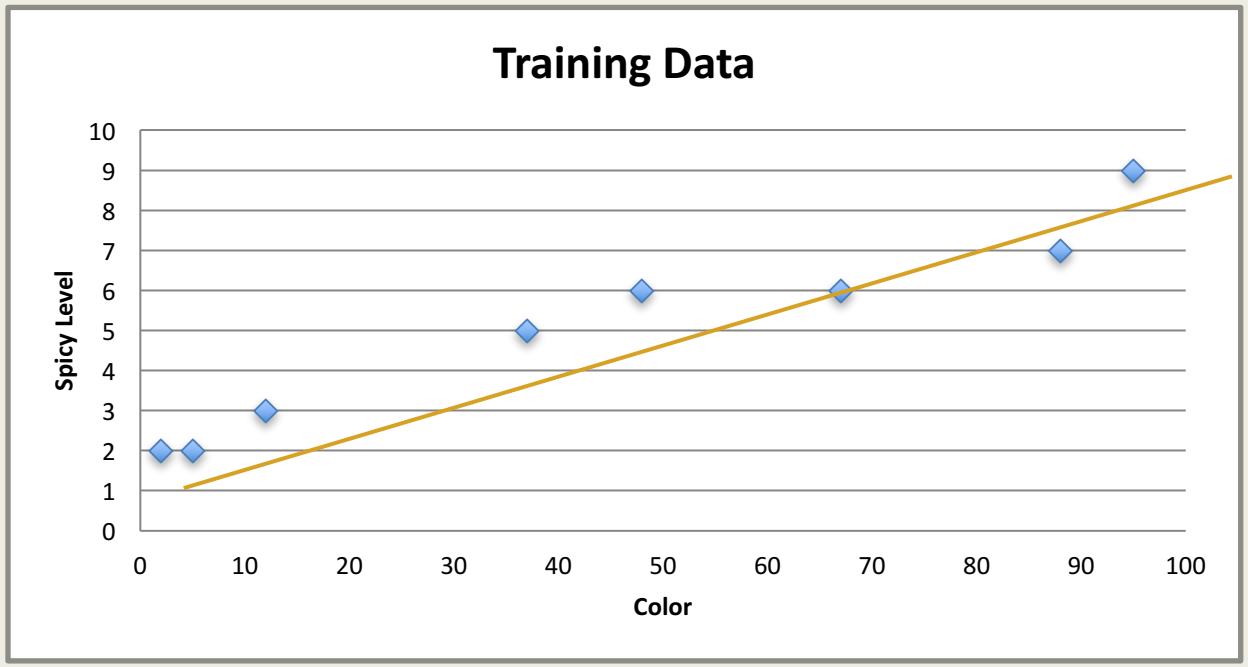
Small



LINEAR REGRESSION

Linear Regression

Model representation: $p(x) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} x$



Cost Function (Training Error)

$$\frac{1}{2m}\sum_{i=1}^{m}\left(p\left(x^i\right) - y^i\right)^2$$