Google Cloud Academy

The Path to Partner Technical Readiness
Professional Machine Learning Engineer Certification

Session 1: Welcome to the Professional Machine Learning engineer Certification Program

We will Begin in:

<<15:00->>

Instructor: Ben Ahmed

Instructor: Ben Ahmed

Working in IT industry for 12 years. A mix of roles, hands on engineering roles in ML (NLP, CNN’s, DNN’s, VertexAI, Python, TensorFlow, PyTorch, CNN’s,) and MLOps solution designing

Years of experience instructing

a Authorized Google Cloud Trainer, I have now been running workshops with Google Customers for over 4 Years for Google. Conducted +25 Workshops, +600 Attendees

Other pertinent tech roles

Machine learning Engineer for Banking (HSBC) & Oil & Gas (BP), and for Startups,

The following course materials are copyright protected materials.

They may not be reproduced or distributed and may only be used by students attending this Google Cloud Partner Learning Services program.

Session logistics

- If you have a question, please enter it in the Meet chat or use the Meet raise hand button
- Answers may be deferred until the end of the session

- The session is **not recorded** and **slides will not be distributed**
  - You'll have everything you need in the links provided in the Meet chat
  - Please copy them to a local text file as they appear in the chat
  - Chat is not persisted between connections should you get disconnected and have to reconnect

Professional Machine Learning Engineer

A Partner Certification Academy program

How to Succeed ?

1) Technical personnel who must be able to:

- Frame ML problems
- Develop ML models
- Architect ML solutions
- Automate and orchestrate ML pipelines
- Design data preparation and processing systems
- Monitor, optimize, and maintain ML solutions

2) Learn the Machine Learning Fundamentals!

3) Understand the context of what you are trying to solve?

- **Do not try and Memorise Questions**
- **Do not try and use Braindumps / Examdumps**

Refer to the Google Cloud Professional Machine Learning Engineer certification overview.

Are there any prerequisites?

- **No.** Learners are advised to have completed the Google Cloud Big Data and ML Fundamentals course, possess experience with Google Cloud and/or be Professional Data Engineer certified.

How can they sign up?

- Individual learners can visit the schedule page
- Partner organizations can work with their Google Cloud representative for bulk registration and program scheduling.

How long does it take to complete the program?

- Standard*: ~ 100 hours across 6 modules.
- Each **week** the learner will complete:
  - 4-6 hours of on-demand lessons
  - 4-6 hours of hands-on labs
  - 4 hours of live instructor-led workshops

Why should my partners enroll?

- Primarily self-paced, but with instructor support!
- Program is ideal for busy partner learners
- **Free** certification exam voucher (by completing specific requirements)

Refer to the Partner Certification Academy site for additional details.

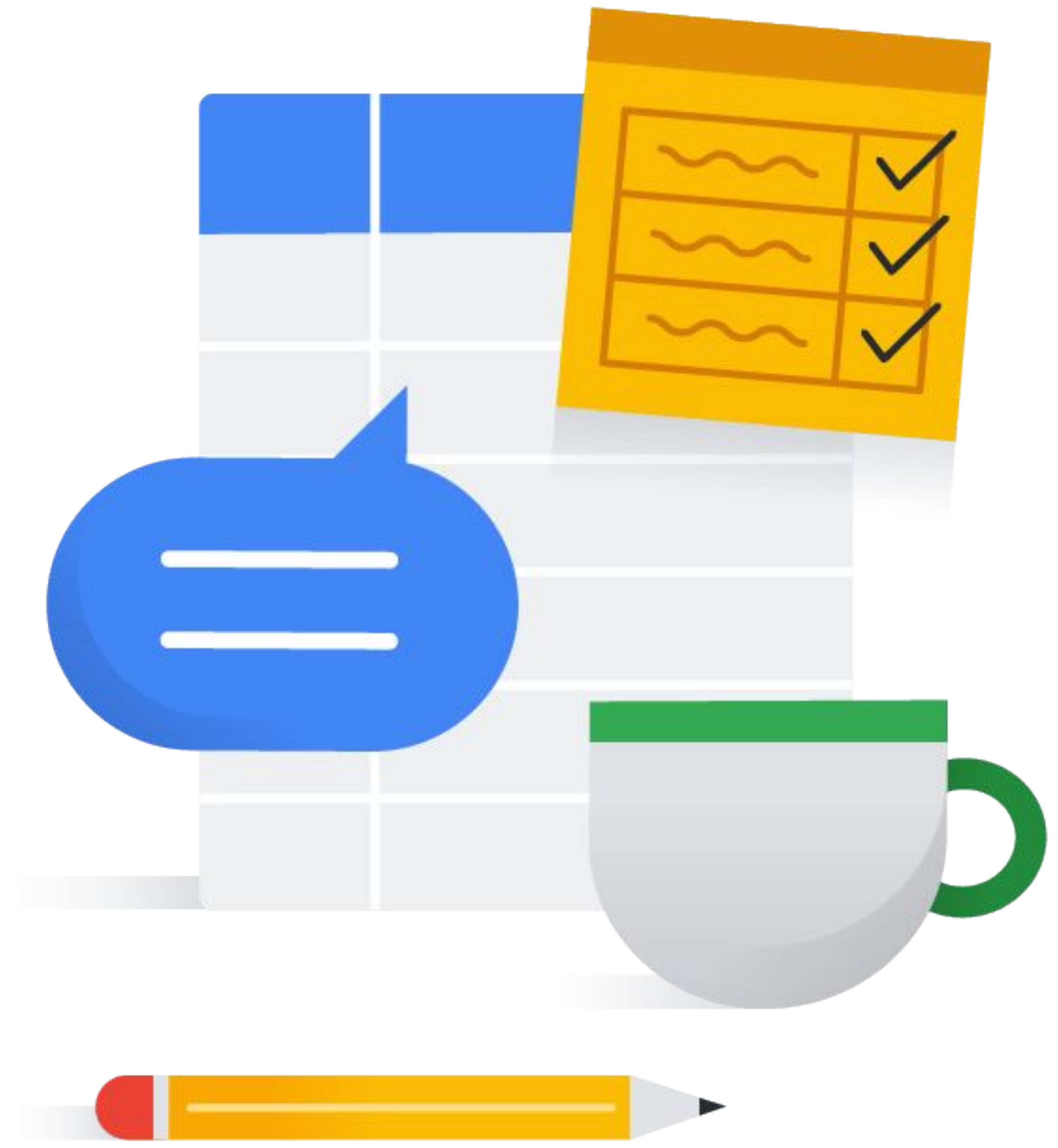
**Standard scheduling encompasses one module per week*

Program issues or concerns?

- Problems with **accessing** Cloud Skills Boost for Partners
  - partner-training@google.com

- Problems with **a lab** (locked out, etc.)
  - support@qwiklabs.com

- Contact your **Google mentor**



Google Cloud
Partner Advantage

Getting started on Partner Advantage: [Partner Advantage Support]

For partners: [Join Partner Advantage]

For individuals: [Get help accessing Partner Advantage]

Google Cloud Academy: Prof Machine Learning Engineer Learning Journey

PRE-WORK	WEEK 0	WEEK 1	WEEK 2	WEEK 3	WEEK 4	WEEK 5	WEEK 6	EXAM
Review: PMLE Certification page **Review the** PMLE Exam Guide **Review the** PMLE Sample Questions **Onboard** page **Enroll in:** PMLE Learning Path	**On-demand training:** ~13 hrs Google Cloud Platform Big Data and Machine Learning Fundamentals **Skill Badge:** ✱ Perform Foundational Data, ML, and AI Tasks in Google Cloud	**On-demand training:** ~13 hrs How Google Does Machine Learning **Skill Badge:** Baseline: Data, ML, AI	**On-demand training:** ~16 hrs Feature Engineering **Skill Badge:** ✱ Transform and Clean your Data with Dataprep by Trifacta on Google Cloud Lab: Interactive Data Exploration with Vertex AI Workbench	**On-demand training:** ~15 hrs Launching into Machine Learning **Skill Badge:** Create ML Models with BigQuery ML	**On-demand training:** ~14 hrs TensorFlow on Google Cloud Lab: Learning TensorFlow: the Hello World of Machine Learning Lab: Running Distributed TensorFlow using Vertex AI Lab: Creating an Object Detection Application Using TensorFlow	**On-demand training** ~10 hrs **Skill Badge:** ✱ Google Cloud Solution II: Data and Machine Learning **Skill Badge:** Explore Machine Learning Models with Explainable AI	**On-demand training/Check Readiness:** ~20 hrs ML Pipelines on Google Cloud **Skill Badge:** Build and Deploy Machine Learning Solutions on Vertex AI **Reattempt the** PMLE Sample Questions **Review the** PMLE Exam Guide	**Week: 7-8 1st attempt** **2nd attempt 14 days after first attempt** **3rd attempt 60 days after 2nd attempt** **Exam** ✱ required to earn exam voucher
		Workshop A: 2 hrs Big Data and Machine Learning on Google Cloud	Workshop A: 2 hrs Data Exploration and Cleaning in Google Cloud	Workshop A: 2 hrs Machine Learning Basics Part 1	Workshop A: 2 hrs TensorFlow and Keras	Workshop A: 2 hrs Advanced Machine Learning with TensorFlow	Workshop A: 2 hrs Introduction to MLOps	
		Workshop B: 2 hrs Demos - How ML works in Google Cloud	Workshop B: 2 hrs Feature Selection and Engineering in Google Cloud	Workshop B: 2 hrs Machine Learning Basics Part 2	Workshop B:2 hrs Optimizing and Scaling Model Training with TensorFlow	Workshop B: 2 hrs Hyperparameter Tuning and Explainability in Google Cloud	Workshop B: 2 hrs MLOps in Google Cloud	

Session 1: agenda

01 Streaming Data in Google Cloud with Pub/Sub and DataFlow

02 How Google Does Machine Learning

03 Machine Learning Basics: Part1- Algorithms

04 Machine Learning Basics: Part2 - Model Metrics

Machine Learning Products in Google Cloud

Build a Custom Model

- Cloud TPUs
- Compute Engine
- Dataproc
- KubeFlow on GKE
- AI Platform -> VertexAI
- BigQuery ML

Build Custom Model (codeless)

AutoML

- AutoML NLP
- AutoML Tables
- AutoML Translation
- AutoML Vision
- AutoML Video Intelligence

Call a Pretrained Model

- Cloud Translation API
- Vision API
- Speech-to-Text API
- Text-to-Speech API
- Video Intelligence API
- Cloud Natural Language API
- Dialogflow

Streaming Data in Google Cloud with Pub/Sub and Dataflow

Instructor: Ben Ahmed

Agenda

Modern Data Pipeline

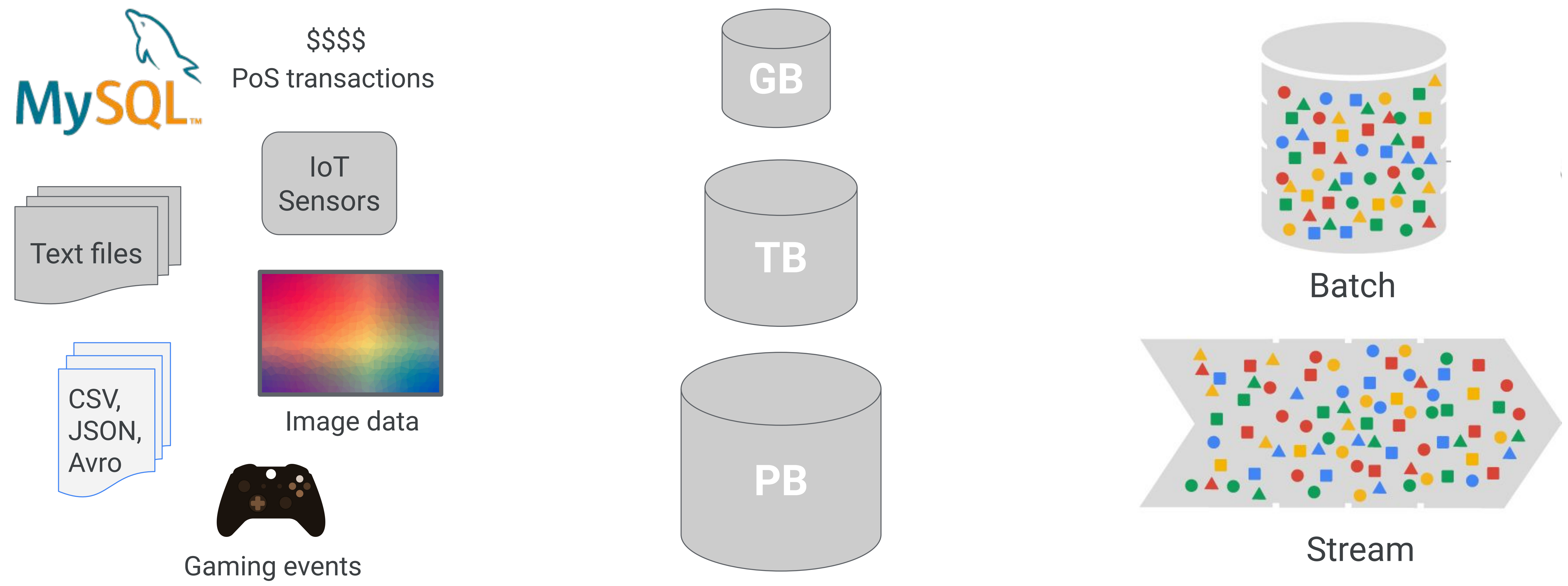
Message-oriented Architectures

Introducing Apache Beam

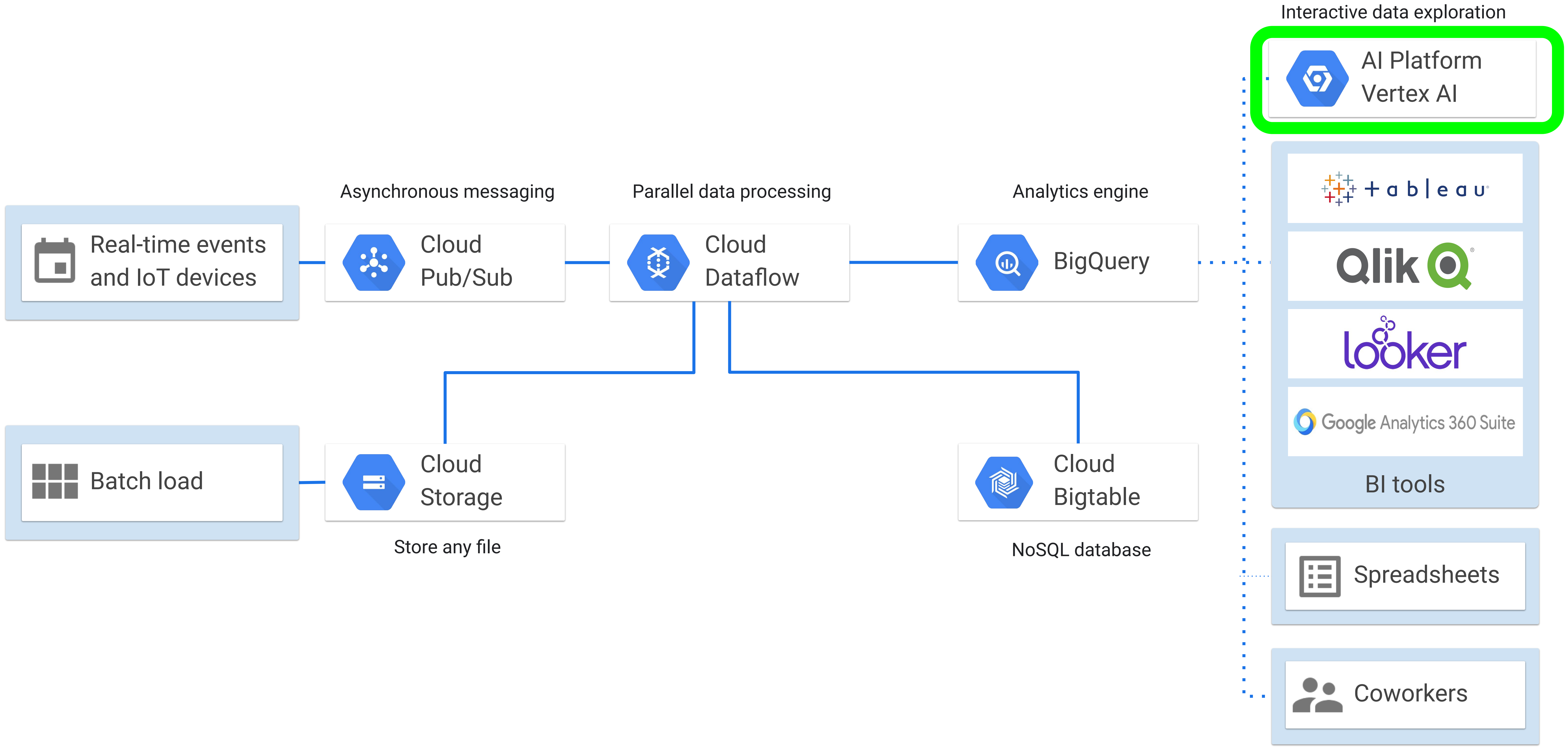
Serverless Data Pipeline

Modern big data pipelines face many challenges

Variety | Volume | Velocity



Modern serverless data pipeline architecture



Agenda

Modern Data Pipeline

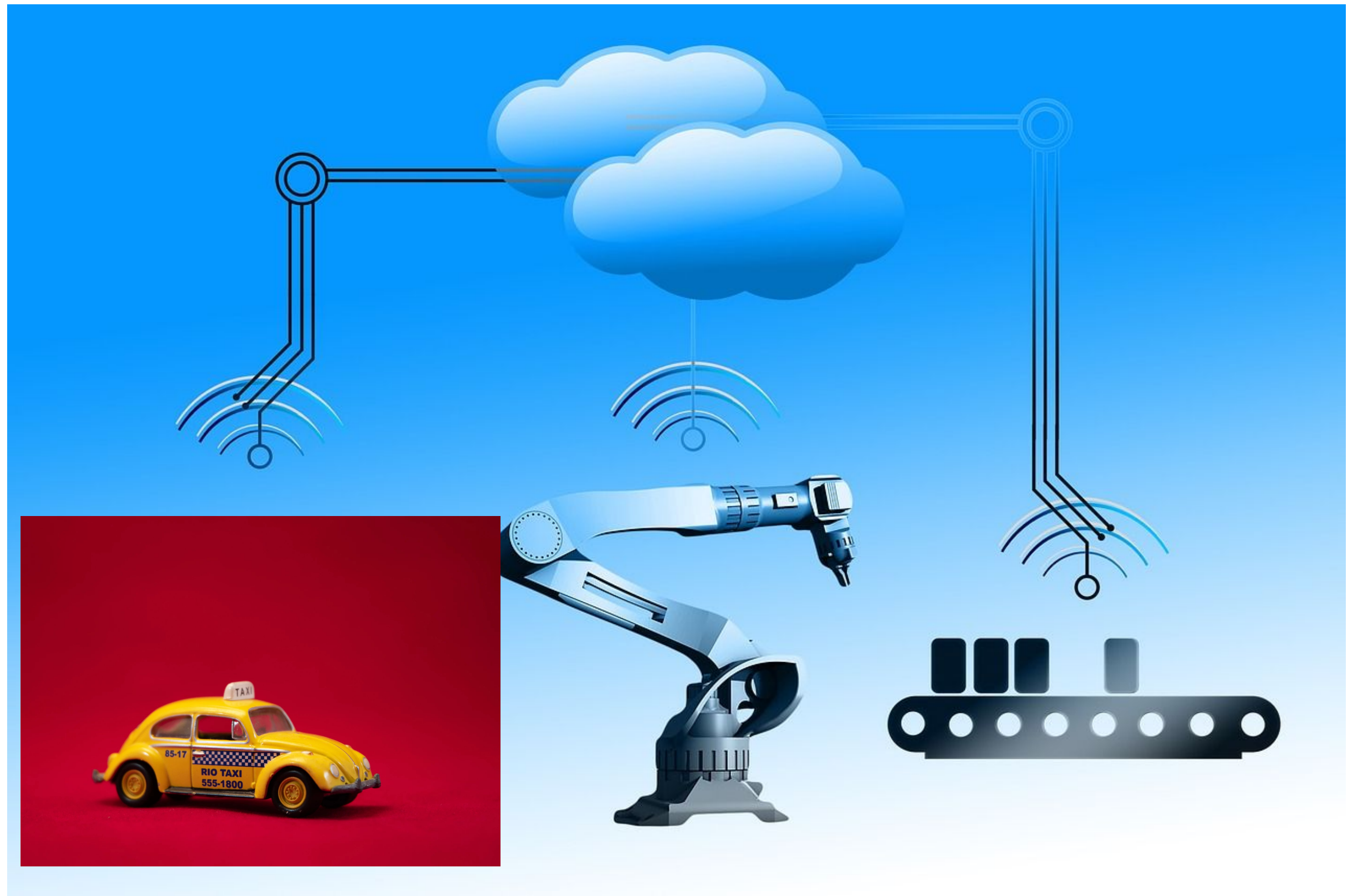
Message-oriented Architectures

Introducing Apache Beam

Serverless Data Pipeline

IoT devices present new challenges to data ingestion

Distributed messages



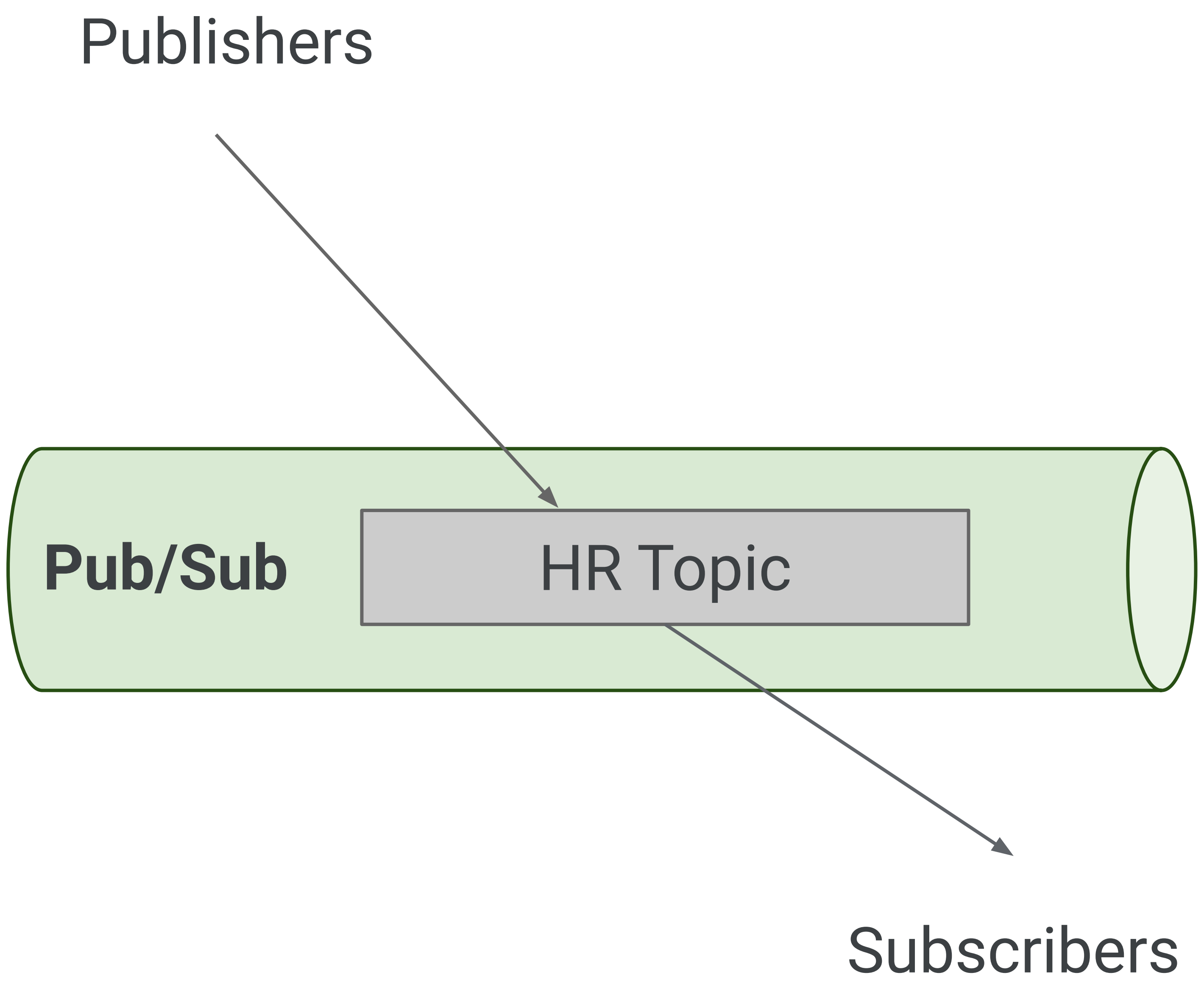
- Data streaming from various processes or devices
- Distributing event notifications (e.g. new user sign up)
- Scale to handle volume
- Reliable (no duplicates)

Pub/Sub offers reliable, real-time messaging

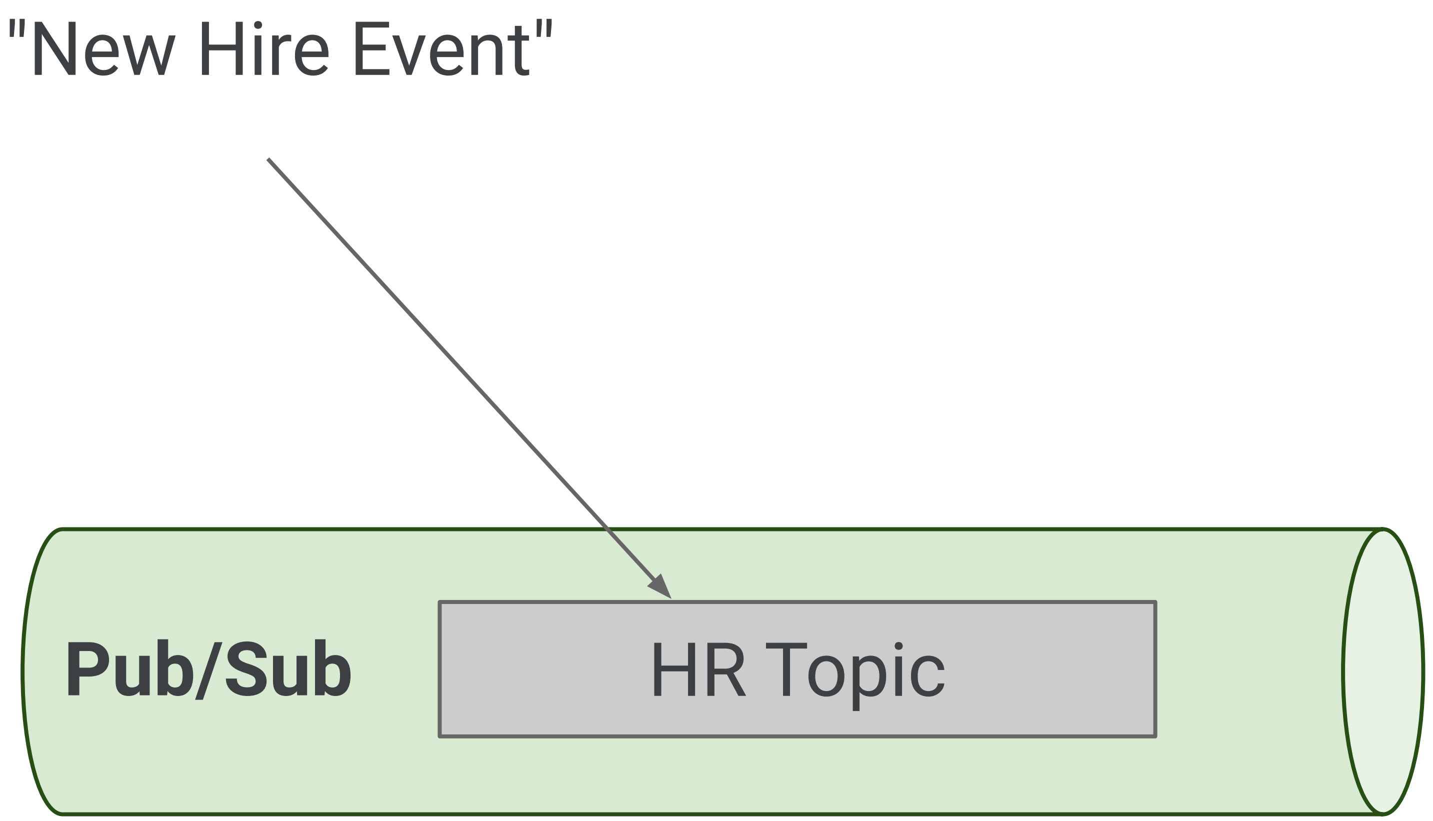
Distributed messaging with Pub/Sub

- At-least-once delivery
- No provisioning, auto-everything
- Open APIs
- Global by default
- End-to-end encryption

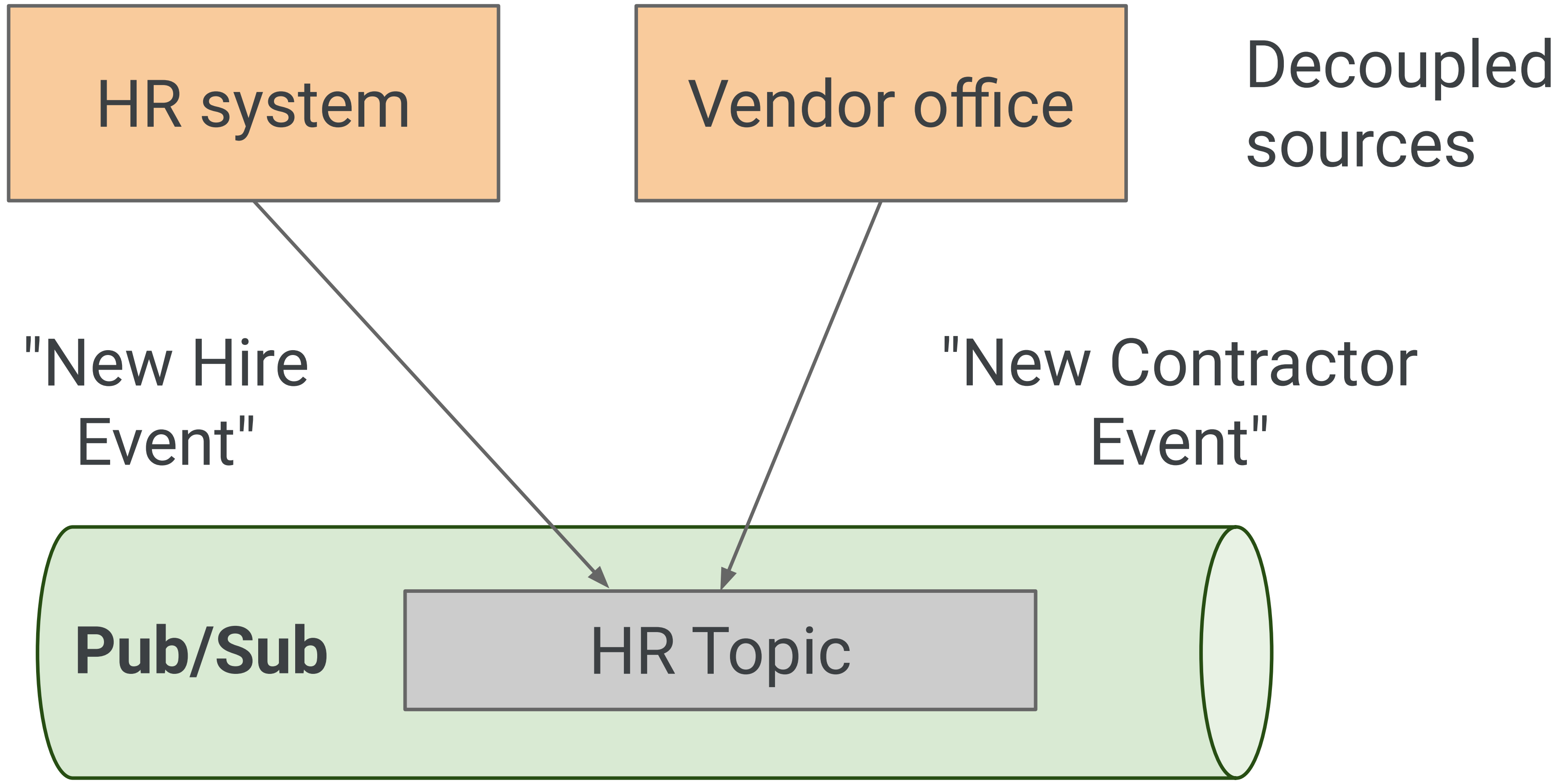
Pub/Sub topics are like radio antennas



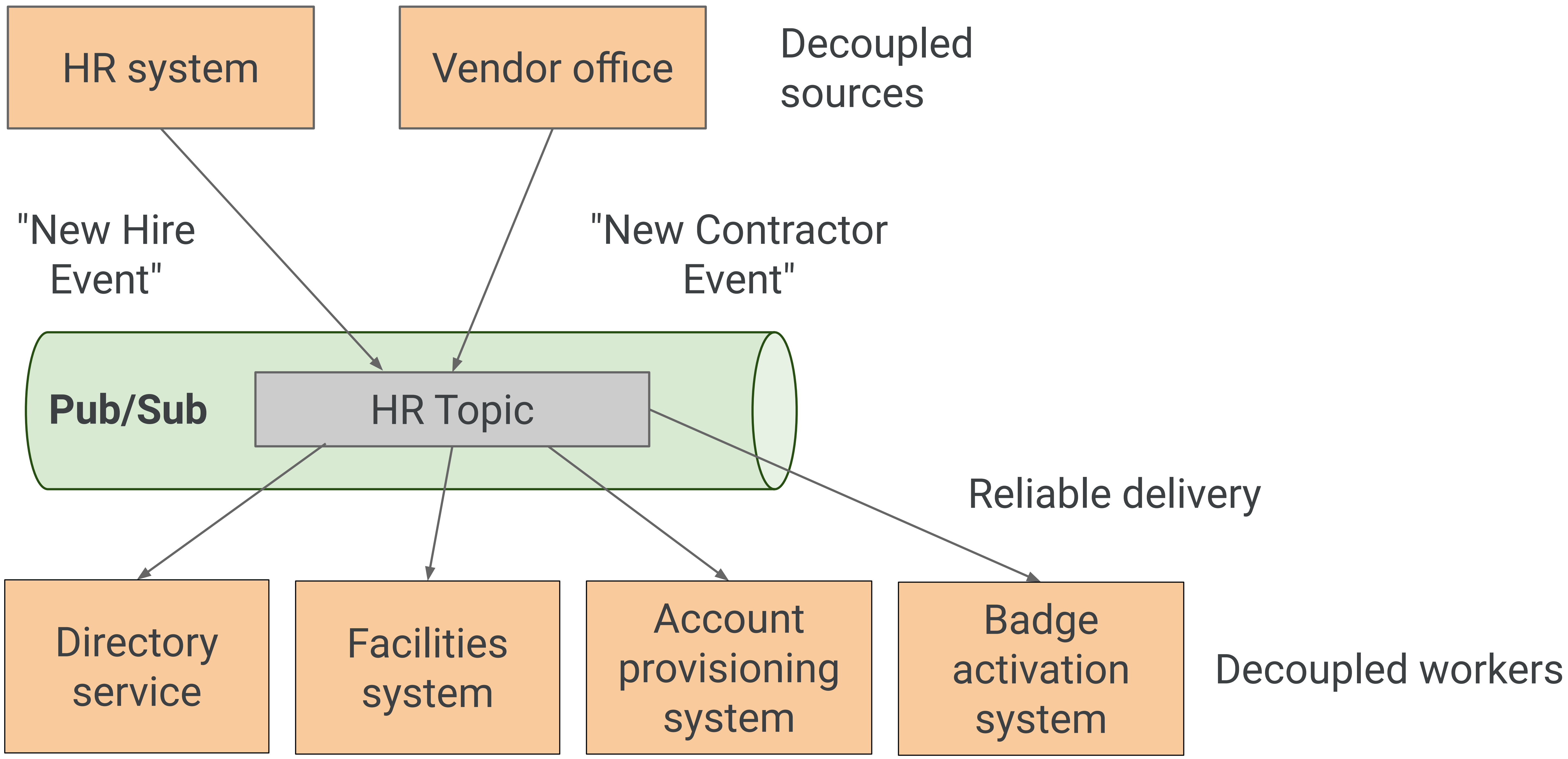
Scenario: HR messaging system



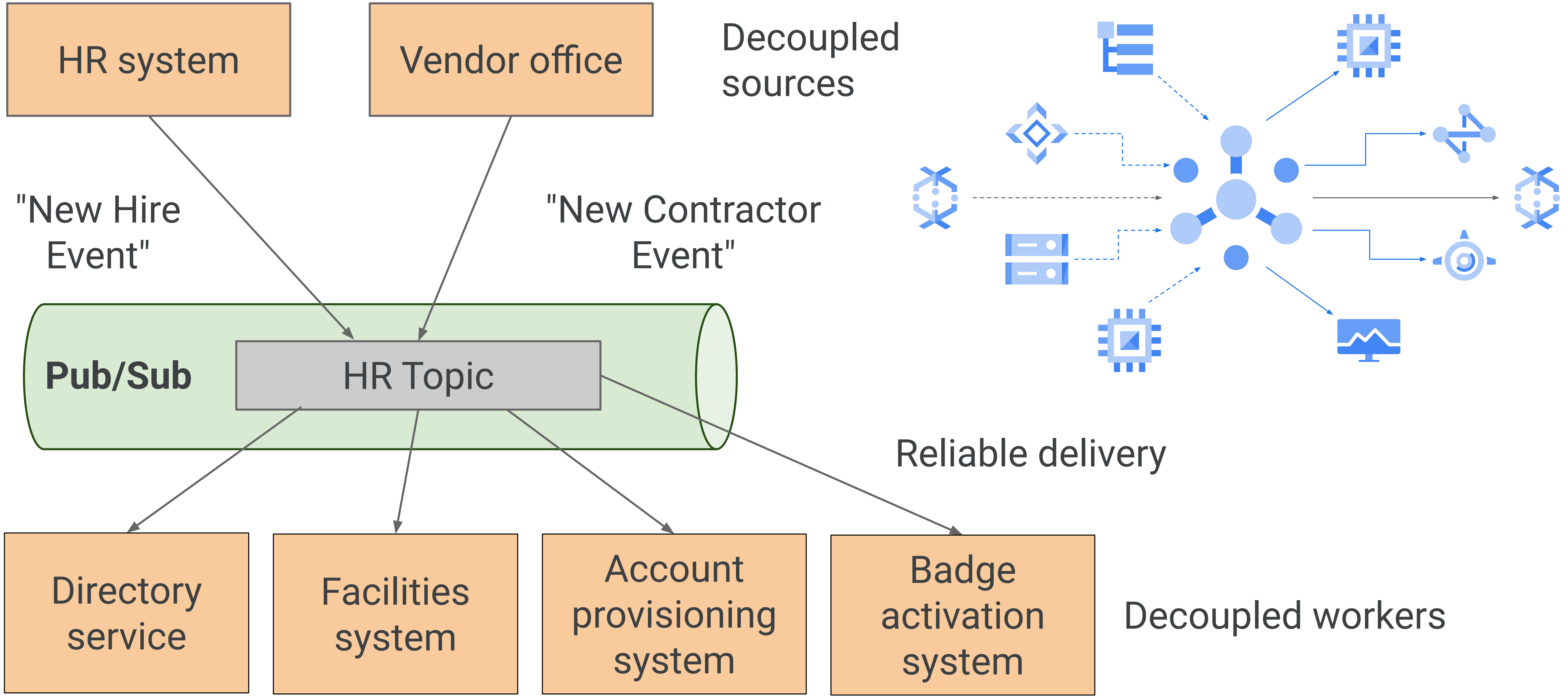
Scenario: HR messaging system



Scenario: HR messaging system

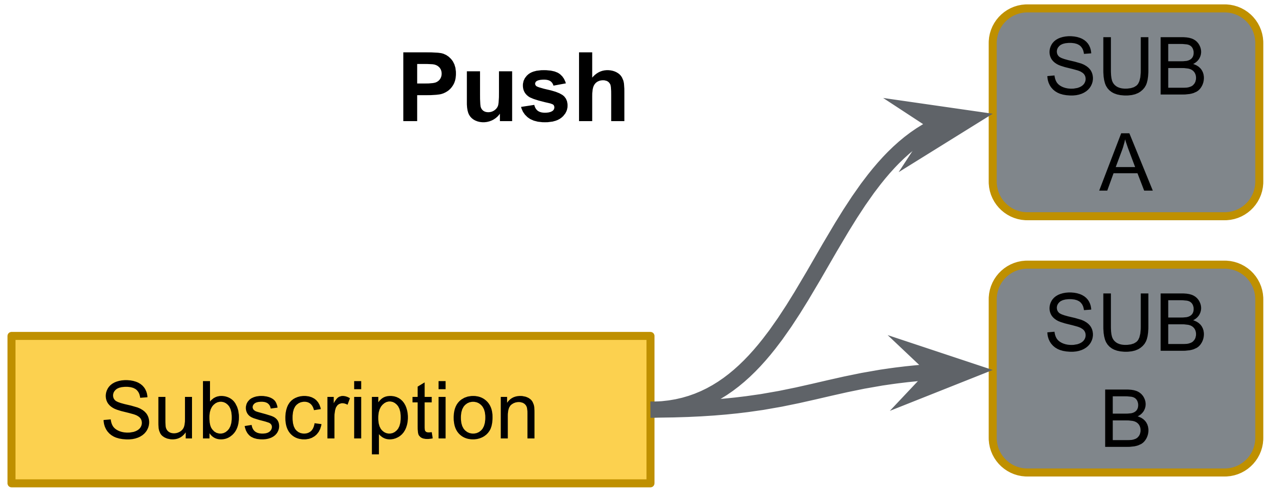


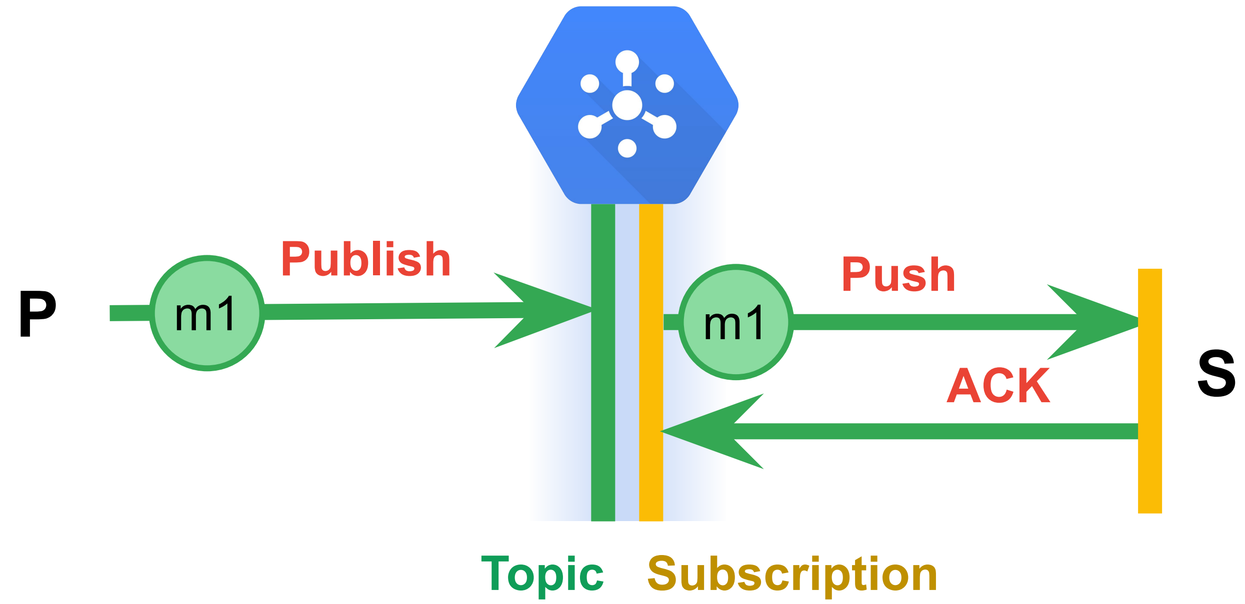
Scenario: HR messaging system



Cloud Pub/Sub provides both Push and Pull delivery

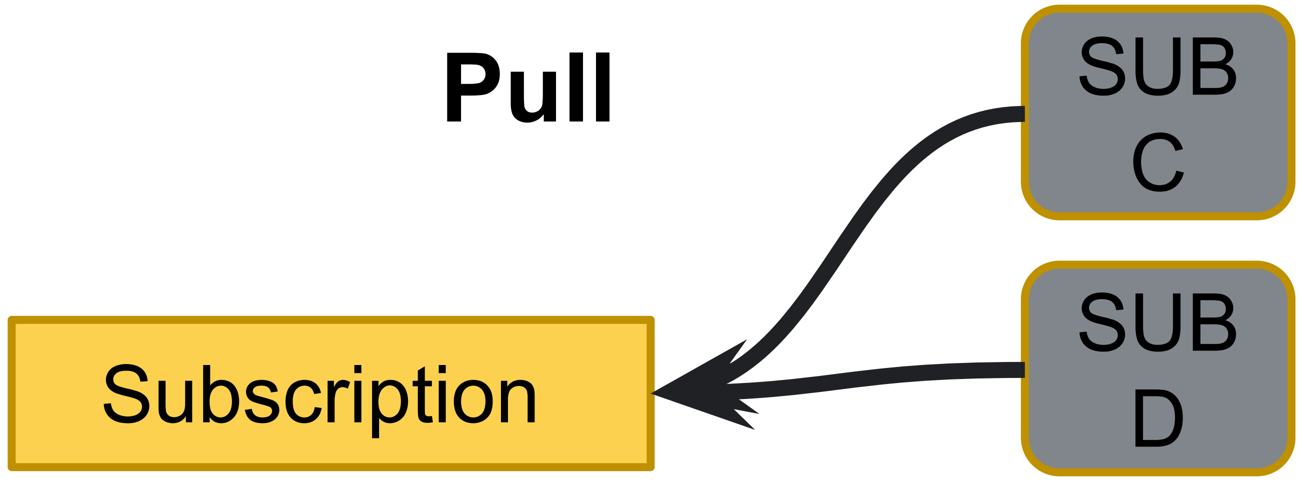
Push

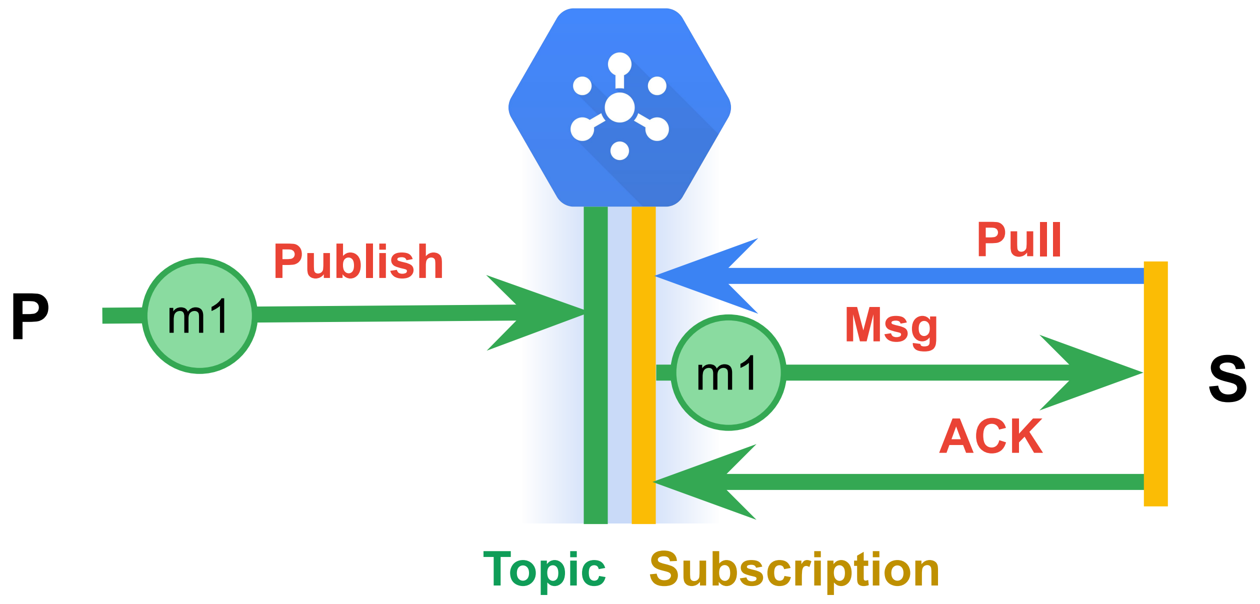




Acknowledgement used for dynamic rate control

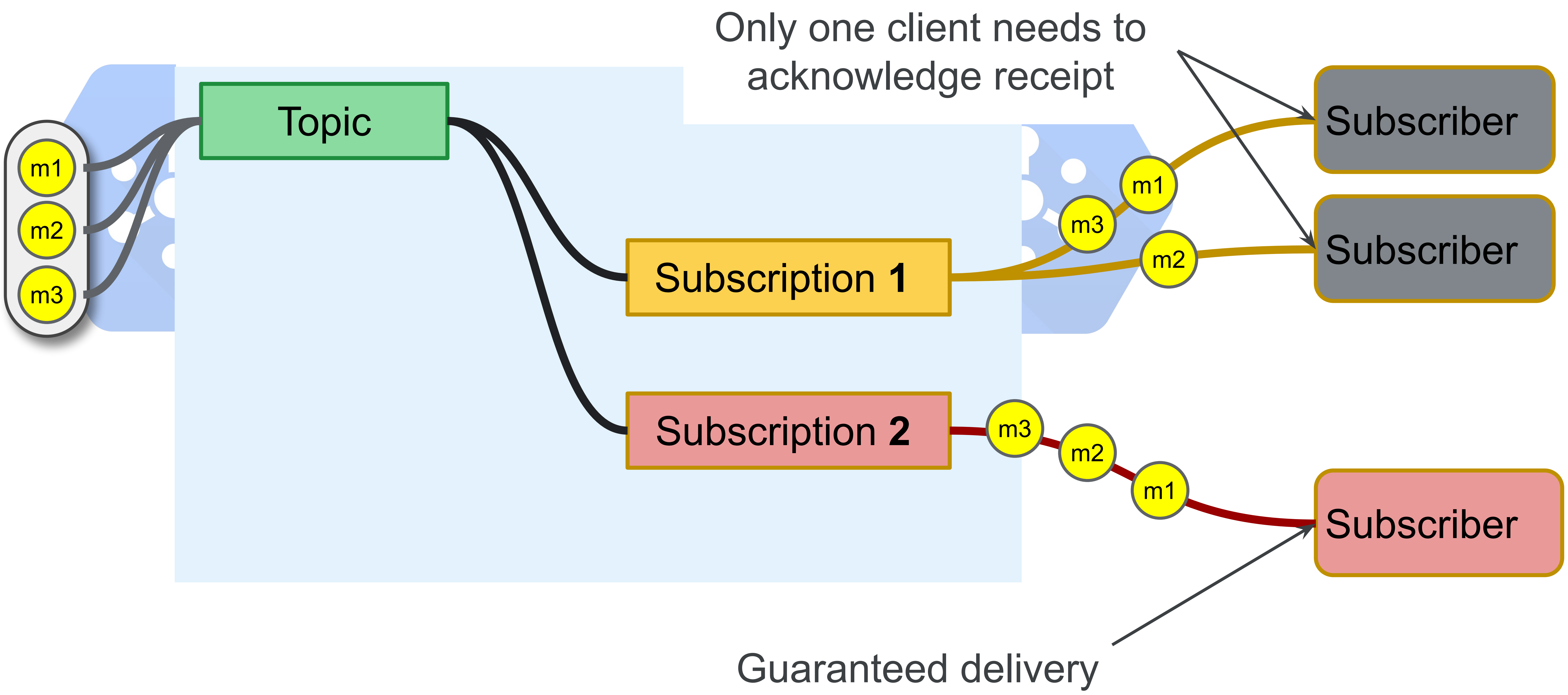
Pull





Messages are stored up to 7 days

Subscribers can work as a team or separately



Agenda

Modern Data Pipeline

Message-oriented Architectures

Introducing Apache Beam

Serverless Data Pipeline

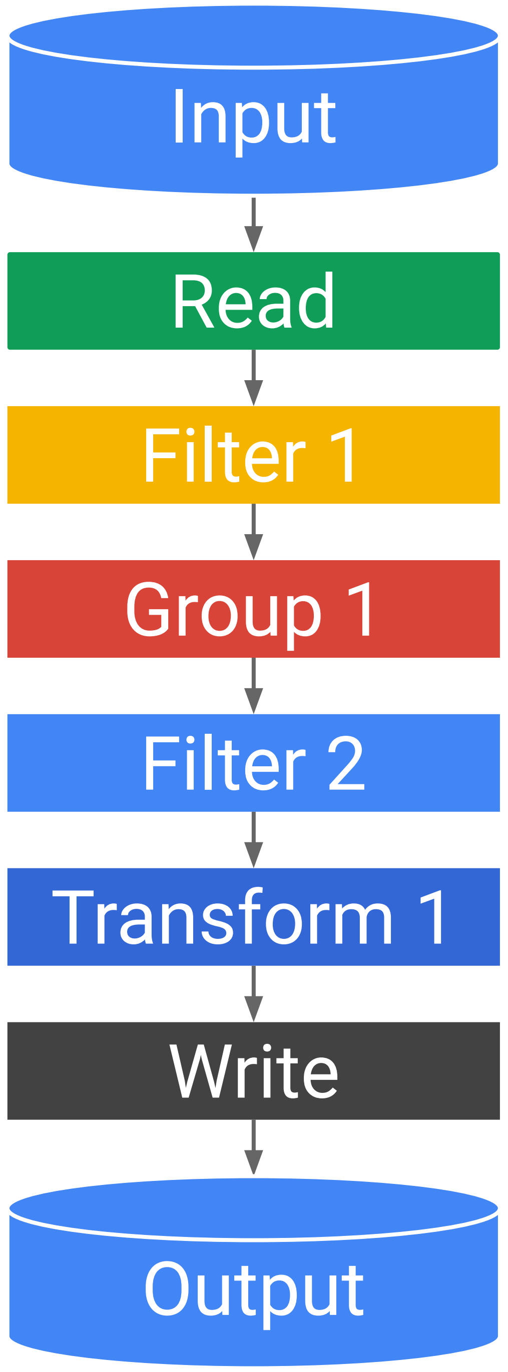
Data Engineers need to solve two distinct problems

Pipeline design with Apache Beam

- Will my code work with both batch and streaming data? ....Yes
- Does the SDK support the transformations I need to do? ....Likely
- Are there existing solutions? ....Choose from templates

beam

Beam offers NoOps data pipelines



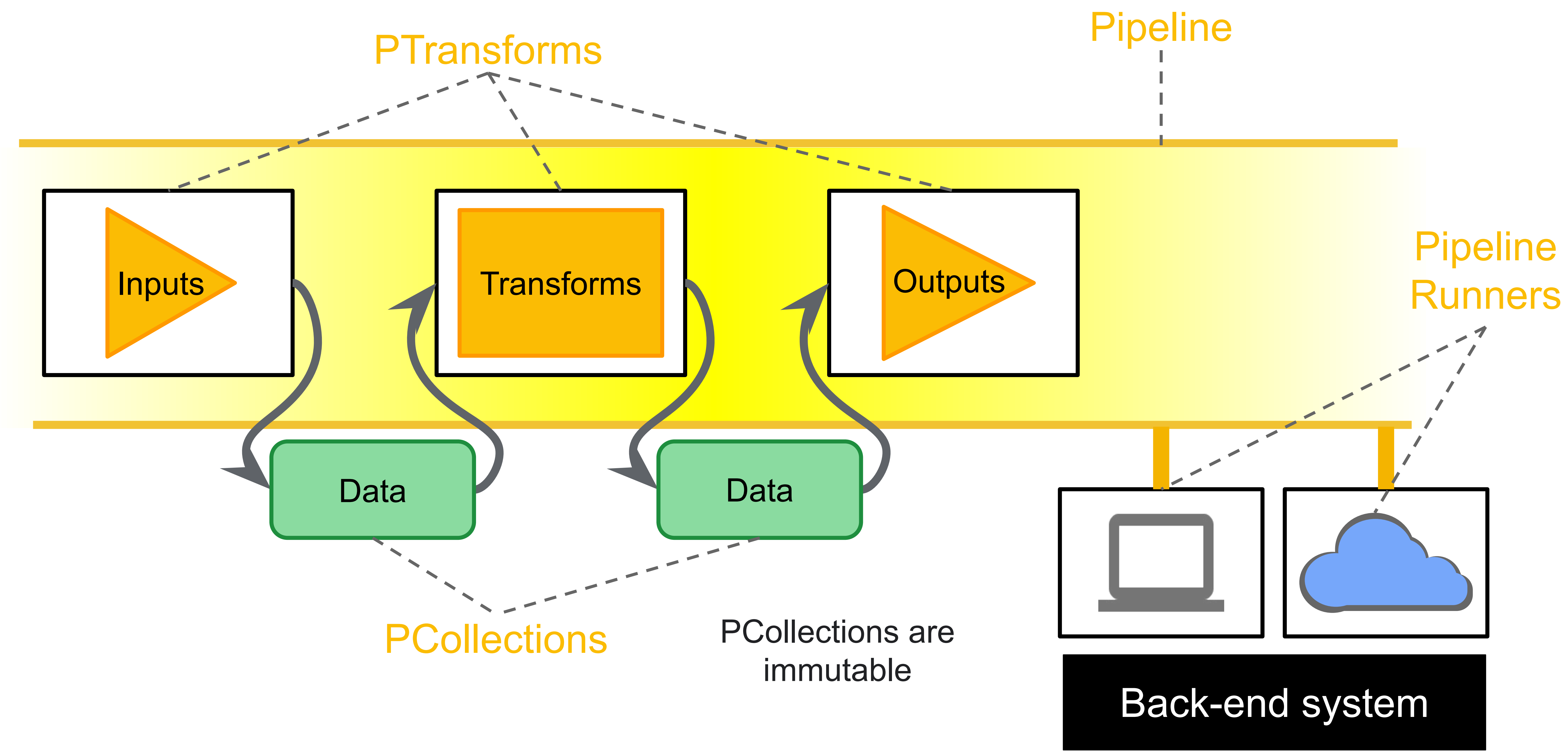

```
Pipeline p = Pipeline.create();
p
.apply(TextIO.Read.from(“gs://…”))
.apply(ParDo.of(new Filter1()))
.apply(new Group1())
.apply(ParDo.of(new Filter2())
.apply(new Transform1())
.apply(TextIO.Write.to(“gs://…”));

p.run();
```

Open-source API (Apache Beam) can be executed on Flink, Spark, etc. also

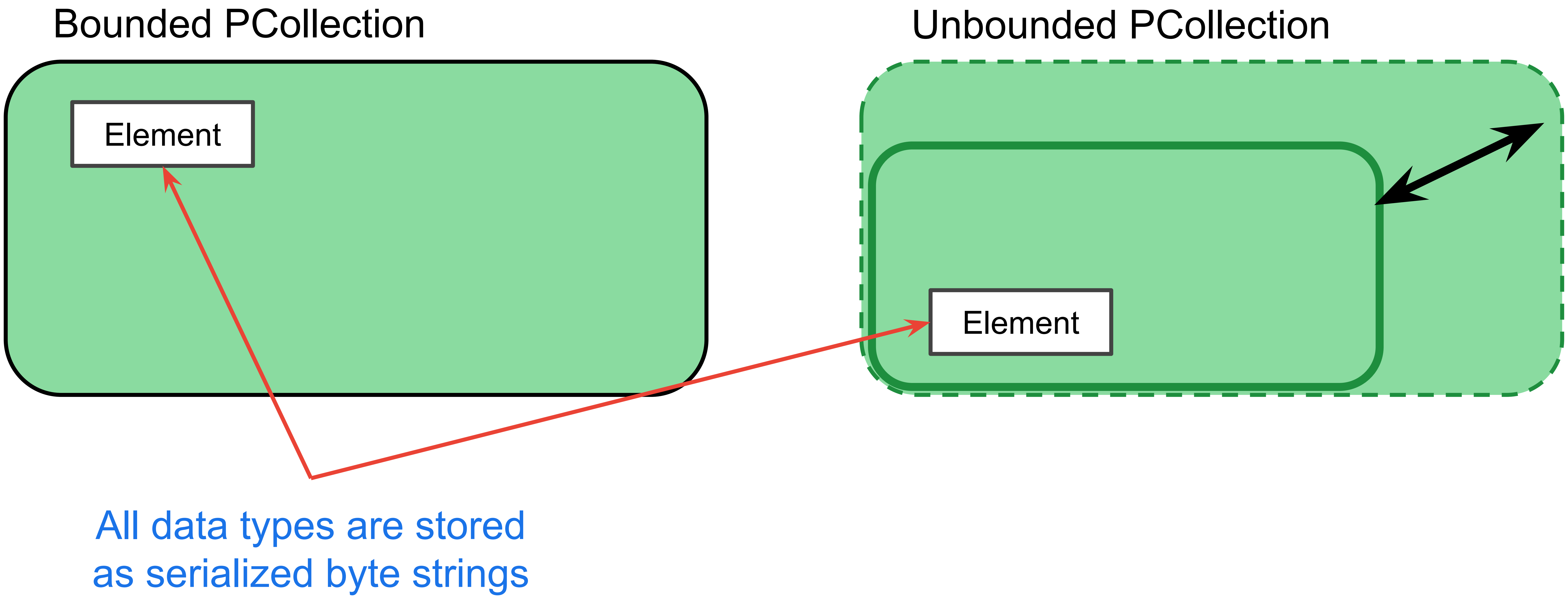
Parallel task (autoscaled by execution framework)

```
class Filter1 extends DoFn<...> {
  public void
  processElement(ProcessContext c) {
      ... = c.element();
      ...
      c.output(...);
  }
```

Apache BEAM = Batch + strEAM



A PCollection represents batch or stream data



Note: Bounded means the data has a fixed size not that the PCollection size is limited. A PCollection can be any size and be distributed across many workers.

Agenda

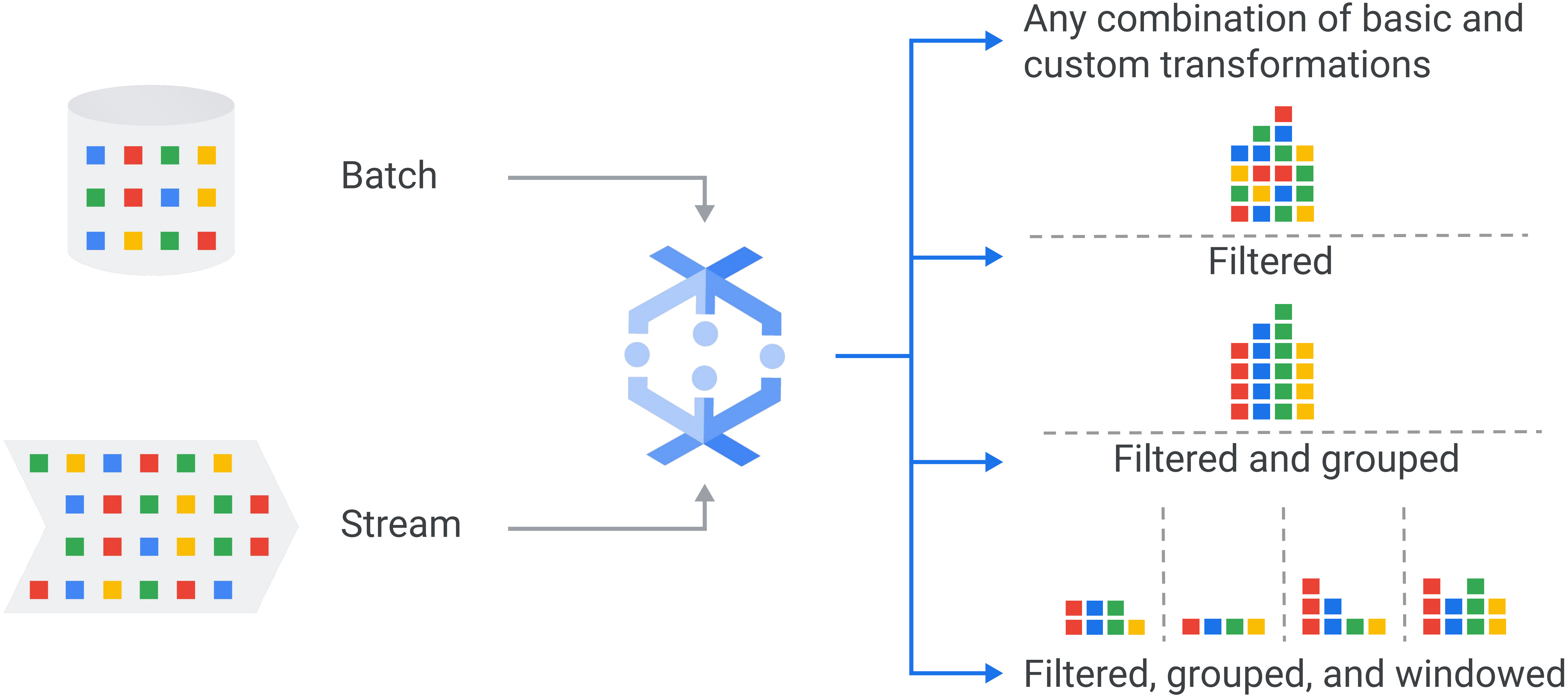
Modern Data Pipeline

Message-oriented Architectures

Introducing Apache Beam

Serverless Data Pipeline

Dataflow does ingest, transform, and load



Why Dataflow

- Serverless, fully managed data processing
- Unified batch and streaming processing + autoscale
- Open source programming model using beam
- Intelligently scales to millions of queries per second

How to construct a simple pipeline


```
PCollection_out = (PCollection_in | PTransform_1
                  | PTransform_2
                  | PTransform_3 )
```

Python

Python overloads the pipe operator

Java

```
PCollection_out = PCollection_in.apply(PTransform_1)
.apply(PTransform_2)
.apply(PTransform_3)
```

Java uses the .apply method

A Pipeline is a directed graph of steps

Python

```
import apache_beam as beam

if __name__ == '__main__':

   with beam.Pipeline(argv=sys.argv) as p:

        (p
           | beam.io.ReadFromText('gs://...')
           | beam.FlatMap(lambda line: count_words(line))
           | beam.io.WriteToText('gs://...')
        )

   # end of with-clause: runs, stops the pipeline
```

Create a pipeline parameterized by command line flags

Read input

Apply transform

Write output

ParDo requires code passed as a **DoFn** object

Python

```
words = ...

class ComputeWordLengthFn(beam.DoFn):
  def process(self, element):
    return [len(element)]

word_lengths = words | beam.ParDo(ComputeWordLengthFn())
```

DoFN

ParDo

The input is a PCollection of strings.

The DoFn to perform on each element in the input PCollection.

The output is a PCollection of integers.

Apply a ParDo to the PCollection "words" to compute lengths for each word.

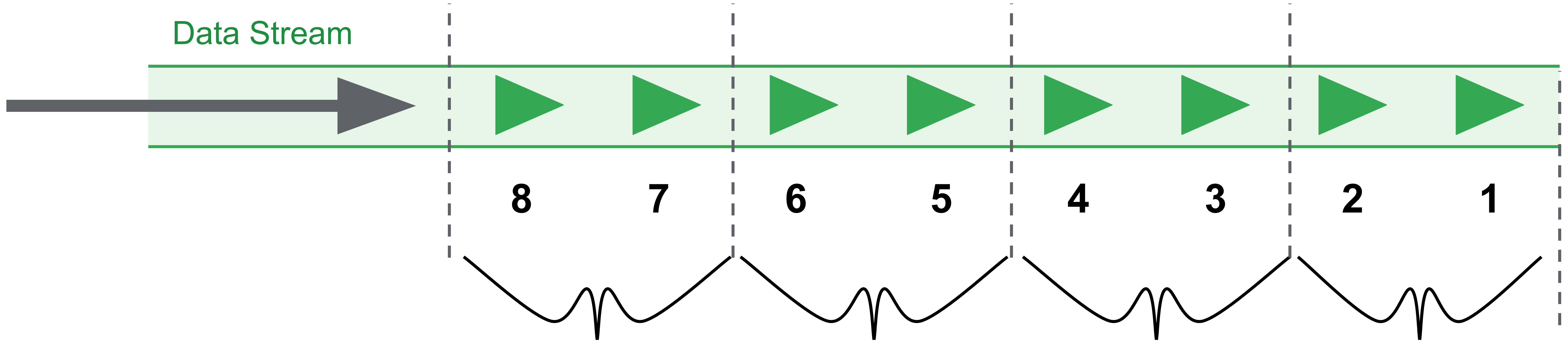
GroupByKey explicitly shuffles key-values pairs

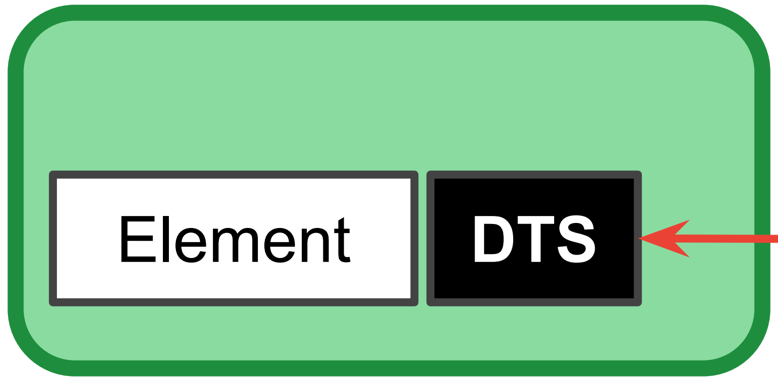
```
cityAndZipcodes = p | beam.Map(lambda fields : (fields[0], fields[1]))

grouped = cityAndZipCodes | beam.GroupByKey()
```

```
Lexington, 40513
Nashville, 37027
Lexington, 40502
Seattle, 98125
Mountain View, 94041
Seattle, 98133
Lexington, 40591
Mountain View, 94085
```

```
Lexington, [40513, 40502, 40592]
Nashville, [37027]
Seattle, [98125, 98133]
Mountain View, [94041, 94085]
```

Divide the stream into a series of finite windows



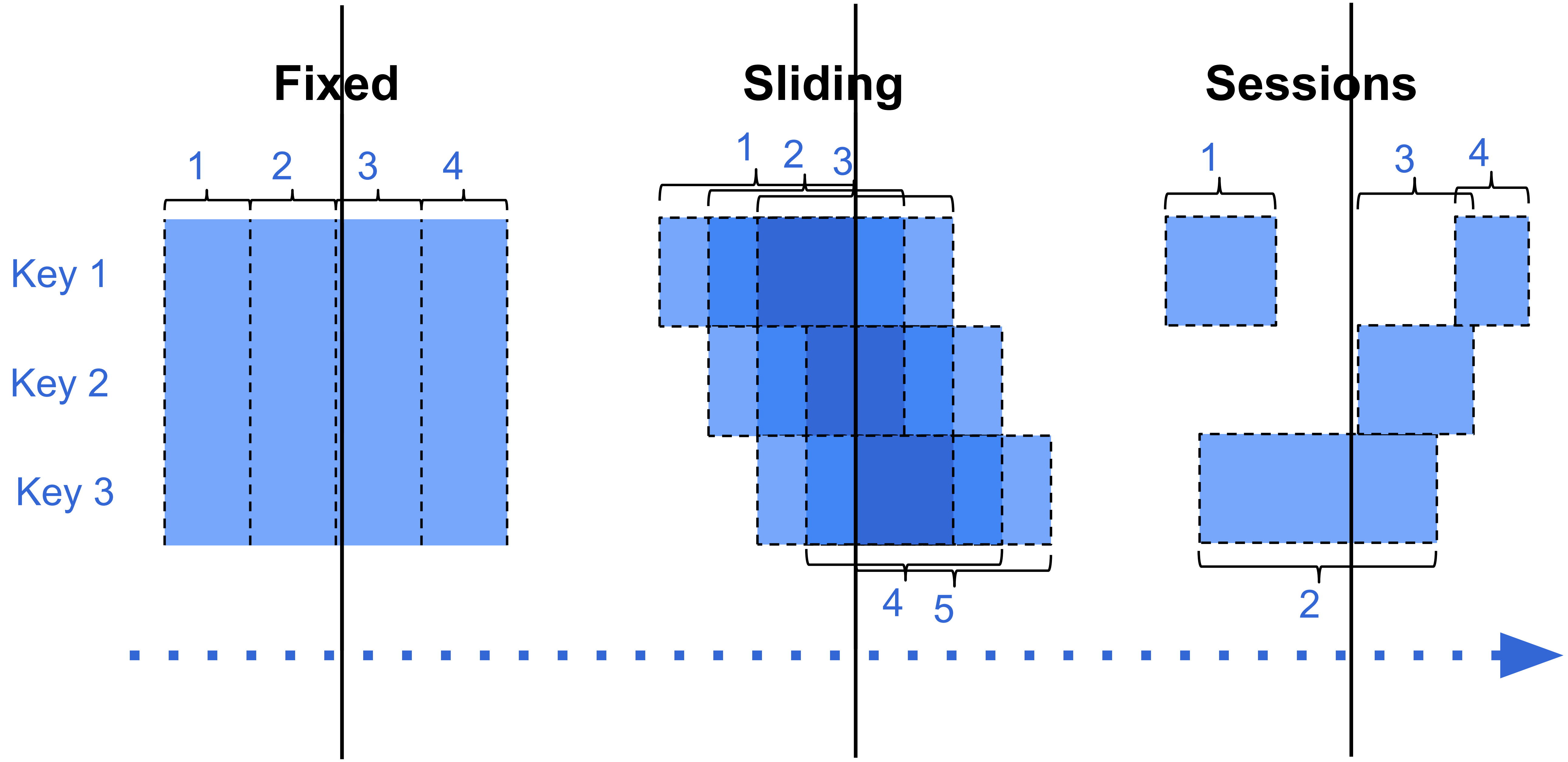


Windows are defined by element date-timestamp (metadata, usually Pub/Sub ingest time)

$$\sum_{i=w}^{n} x_i \qquad \sum_{i=w}^{n} x_i \qquad \sum_{i=w}^{n} x_i$$

Needed For the Exam

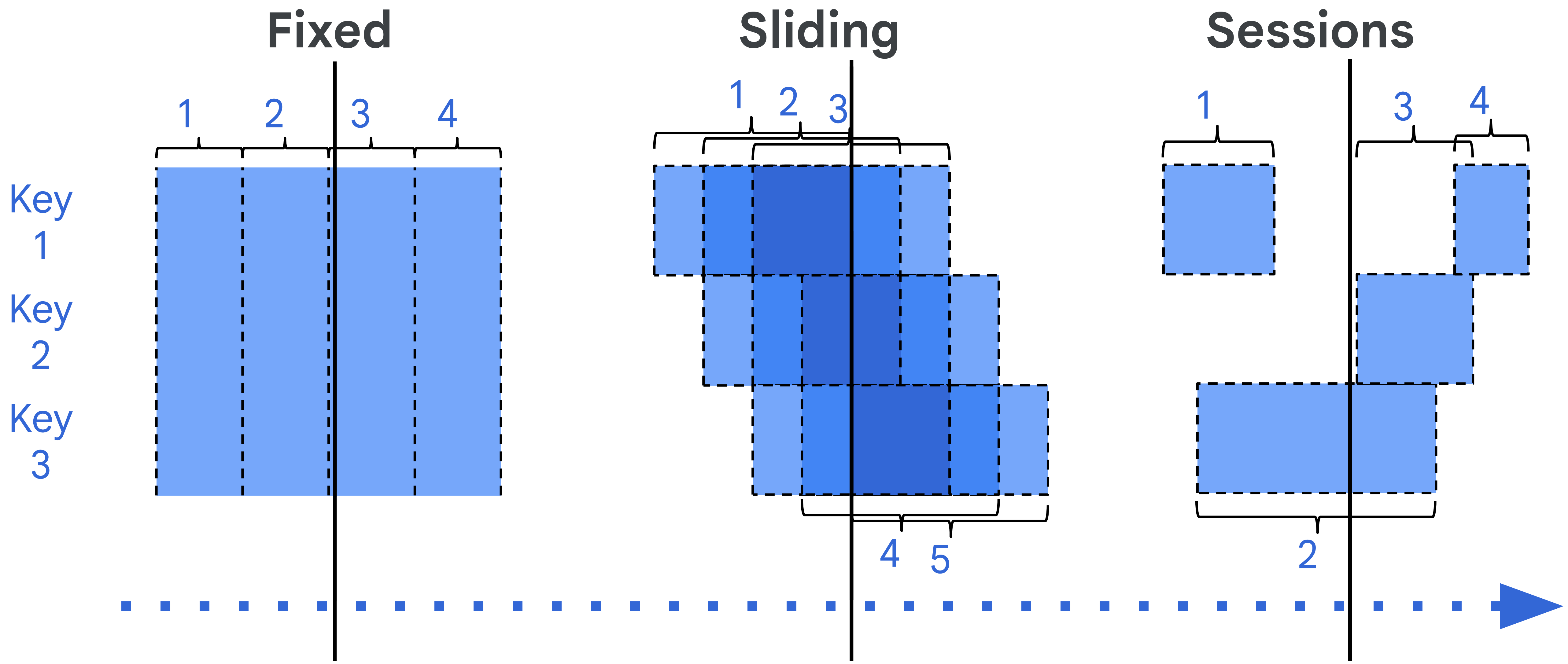
Three kinds of windows fit most circumstances



Windowing divides data into time-based finite chunks

Often required when doing aggregations over unbounded data

Three kinds of windows fit most circumstances



Windowing divides data into time-based finite chunks

Often required when doing aggregations over unbounded data

FYI: For your knowledge

Setting time windows

Fixed-time windows

Python

```
from apache_beam import window
fixed_windowed_items = (
    items | 'window' >> beam.WindowInto(window.FixedWindows(60)))
```

Sliding time windows

Python

```
from apache_beam import window
sliding_windowed_items = (
    items | 'window' >> beam.WindowInto(window.SlidingWindows(30, 5)))
```

Session windows

Python

```
from apache_beam import window
session_windowed_items = (
    items | 'window' >> beam.WindowInto(window.Sessions(10 * 60)))
```

Big data and ML product categories

FYI: For your knowledge

Ingestion & process

- Pub/Sub
- Dataflow
- Dataproc
- Cloud Data Fusion

Storage

- Cloud Storage
- Cloud SQL
- Cloud Spanner
- Cloud Bigtable
- Firestore

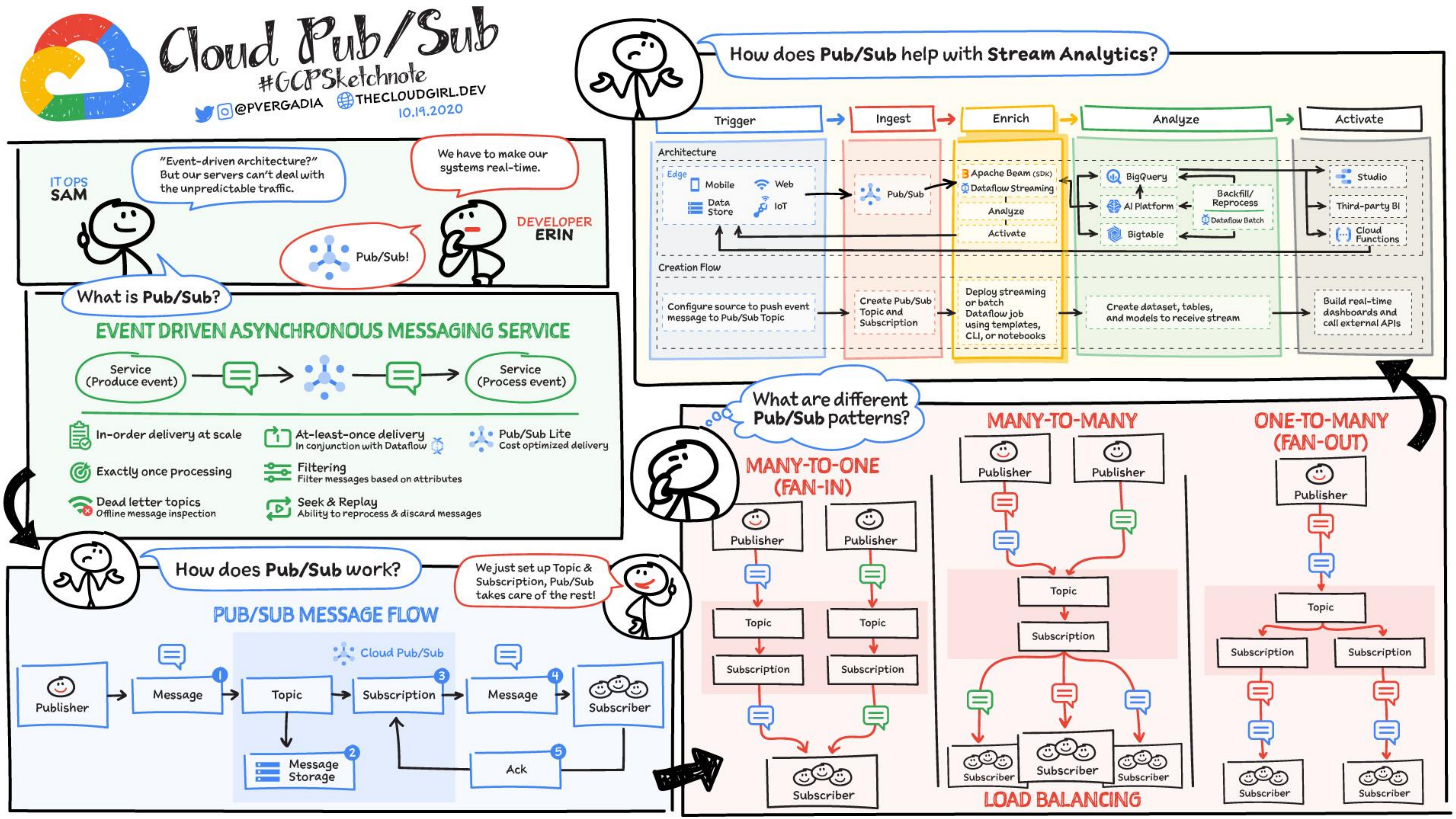
Analytics

- BigQuery
- Google Data Studio
- Looker

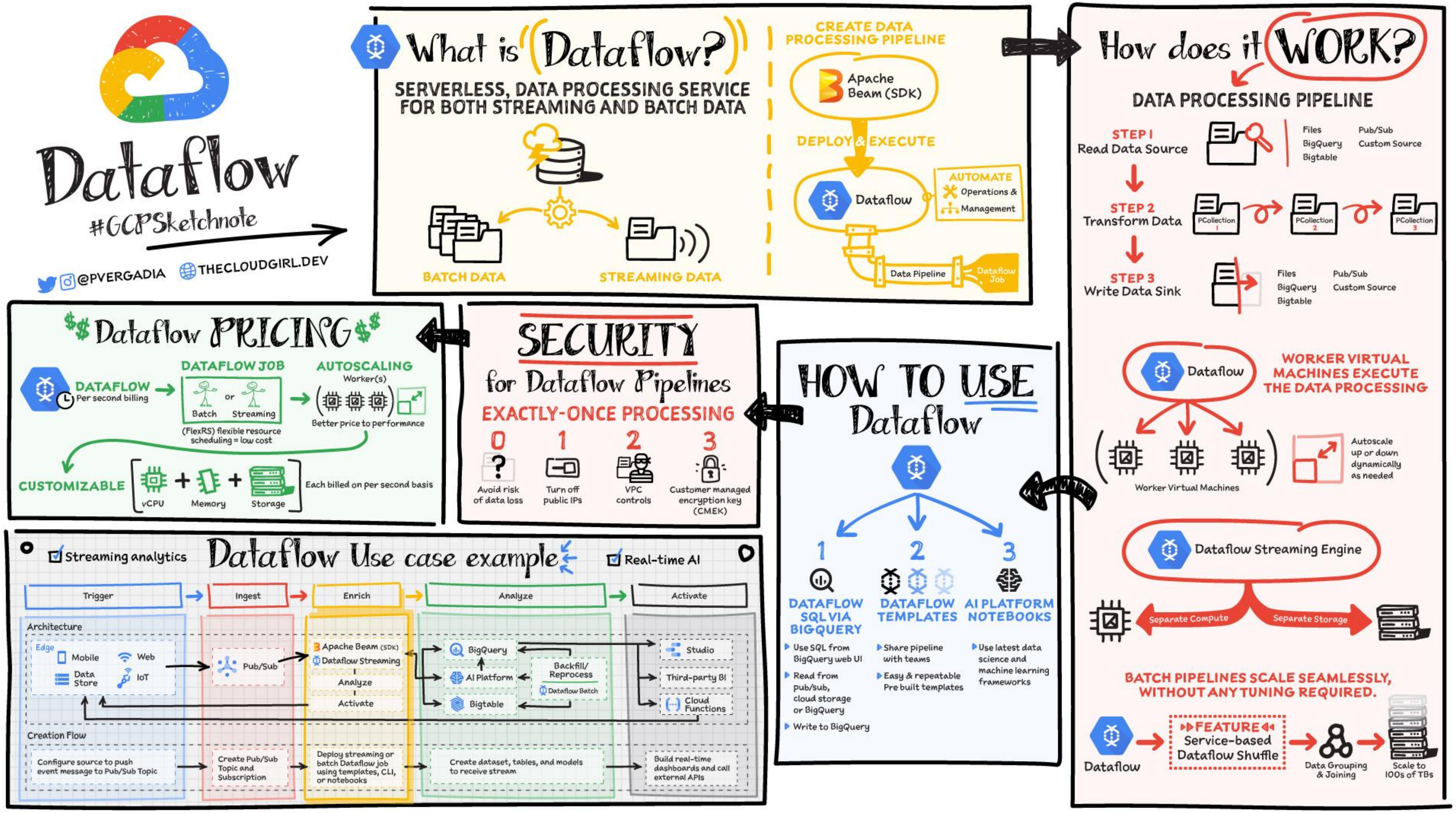
Machine learning

- Vertex AI
- AutoML
- Vertex Workbench
- TensorFlow
- Document AI
- Contact Center AI
- Retail Product Discovery
- Healthcare Data Engine

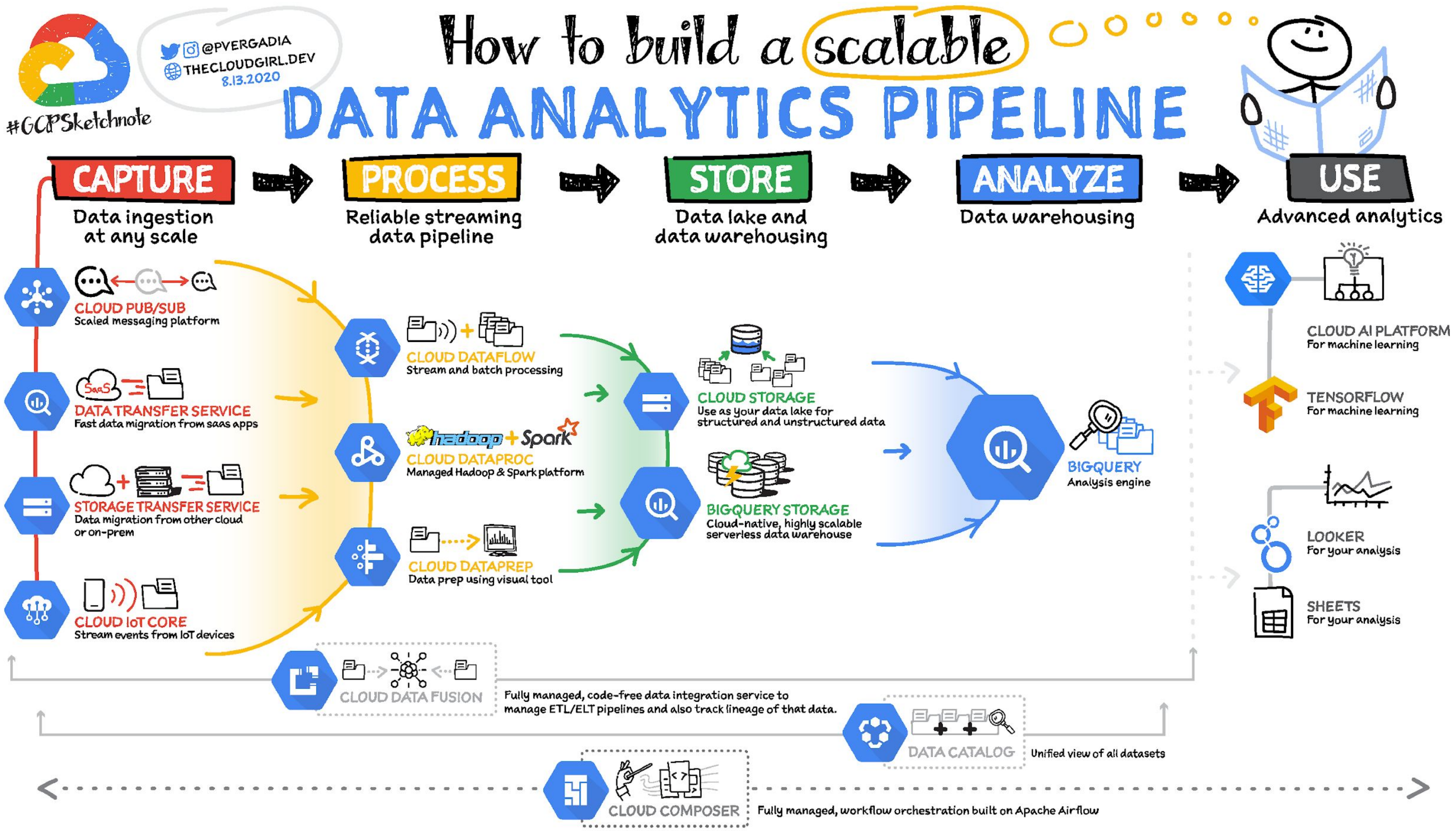
Data-to-AI workflow



Cloud Pub/Sub
#GCPSketchnote
@PVERGADIA
THECLOUDGIRL.DEV
10.19.2020
IT OPS SAM
"Event-driven architecture?" But our servers can't deal with the unpredictable traffic.
We have to make our systems real-time.
DEVELOPER ERIN
Pub/Sub!
What is Pub/Sub?
EVENT DRIVEN ASYNCHRONOUS MESSAGING SERVICE
Service (Produce event)
Service (Process event)
In-order delivery at scale
At-least-once delivery
In conjunction with Dataflow
Pub/Sub Lite
Cost optimized delivery
Exactly once processing
Filtering
Filter messages based on attributes
Dead letter topics
Offline message inspection
Seek & Replay
Ability to reprocess & discard messages
How does Pub/Sub work?
We just set up Topic & Subscription, Pub/Sub takes care of the rest!
PUB/SUB MESSAGE FLOW
Cloud Pub/Sub
Publisher
Message
1
Topic
Subscription
3
Message
4
Subscriber
Message Storage
2
Ack
5
How does Pub/Sub help with Stream Analytics?
Trigger
Ingest
Enrich
Analyze
Activate
Architecture
Edge
Mobile
Web
Data Store
IoT
Pub/Sub
Apache Beam (SDK)
Dataflow Streaming
Analyze
Activate
BigQuery
AI Platform
Bigtable
Backfill/ Reprocess
Dataflow Batch
Studio
Third-party BI
Cloud Functions
Creation Flow
Configure source to push event message to Pub/Sub Topic
Create Pub/Sub Topic and Subscription
Deploy streaming or batch Dataflow job using templates, CLI, or notebooks
Create dataset, tables, and models to receive stream
Build real-time dashboards and call external APIs
What are different Pub/Sub patterns?
MANY-TO-ONE (FAN-IN)
Publisher
Publisher
Topic
Topic
Subscription
Subscription
Subscriber
MANY-TO-MANY
Publisher
Publisher
Topic
Subscription
Subscriber
Subscriber
Subscriber
LOAD BALANCING
ONE-TO-MANY (FAN-OUT)
Publisher
Topic
Subscription
Subscription
Subscriber
Subscriber



Dataflow
#GCPSketchnote
@PVERGADIA
THECLOUDGIRL.DEV
What is Dataflow?
SERVERLESS, DATA PROCESSING SERVICE FOR BOTH STREAMING AND BATCH DATA
BATCH DATA
STREAMING DATA
CREATE DATA PROCESSING PIPELINE
Apache Beam (SDK)
DEPLOY & EXECUTE
Dataflow
AUTOMATE
Operations & Management
Data Pipeline
Dataflow Job
How does it WORK?
DATA PROCESSING PIPELINE
STEP 1
Read Data Source
Files
BigQuery
Bigtable
Pub/Sub
Custom Source
STEP 2
Transform Data
PCollection 1
PCollection 2
PCollection 3
STEP 3
Write Data Sink
Files
BigQuery
Bigtable
Pub/Sub
Custom Source
Dataflow
WORKER VIRTUAL MACHINES EXECUTE THE DATA PROCESSING
Worker Virtual Machines
Autoscale up or down dynamically as needed
Dataflow Streaming Engine
Separate Compute
Separate Storage
BATCH PIPELINES SCALE SEAMLESSLY, WITHOUT ANY TUNING REQUIRED.
Dataflow
FEATURE
Service-based Dataflow Shuffle
Data Grouping & Joining
Scale to 100s of TBs
$$ Dataflow PRICING $$
DATAFLOW
Per second billing
DATAFLOW JOB
Batch
or
Streaming
(FlexRS) flexible resource scheduling = low cost
AUTOSCALING
Worker(s)
Better price to performance
CUSTOMIZABLE
vCPU
Memory
Storage
Each billed on per second basis
SECURITY
for Dataflow Pipelines
EXACTLY-ONCE PROCESSING
0
Avoid risk of data loss
1
Turn off public IPs
2
VPC controls
3
Customer managed encryption key (CMEK)
HOW TO USE Dataflow
1
DATAFLOW SQL VIA BIGQUERY
Use SQL from BigQuery web UI
Read from pub/sub, cloud storage or BigQuery
Write to BigQuery
2
DATAFLOW TEMPLATES
Share pipeline with teams
Easy & repeatable Pre built templates
3
AI PLATFORM NOTEBOOKS
Use latest data science and machine learning frameworks
Streaming analytics
Dataflow Use case example
Real-time AI
Trigger
Ingest
Enrich
Analyze
Activate
Architecture
Edge
Mobile
Web
Data Store
IoT
Pub/Sub
Apache Beam (SDK)
Dataflow Streaming
Analyze
Activate
BigQuery
AI Platform
Bigtable
Backfill/ Reprocess
Dataflow Batch
Studio
Third-party BI
Cloud Functions
Creation Flow
Configure source to push event message to Pub/Sub Topic
Create Pub/Sub Topic and Subscription
Deploy streaming or batch Dataflow job using templates, CLI, or notebooks
Create dataset, tables, and models to receive stream
Build real-time dashboards and call external APIs

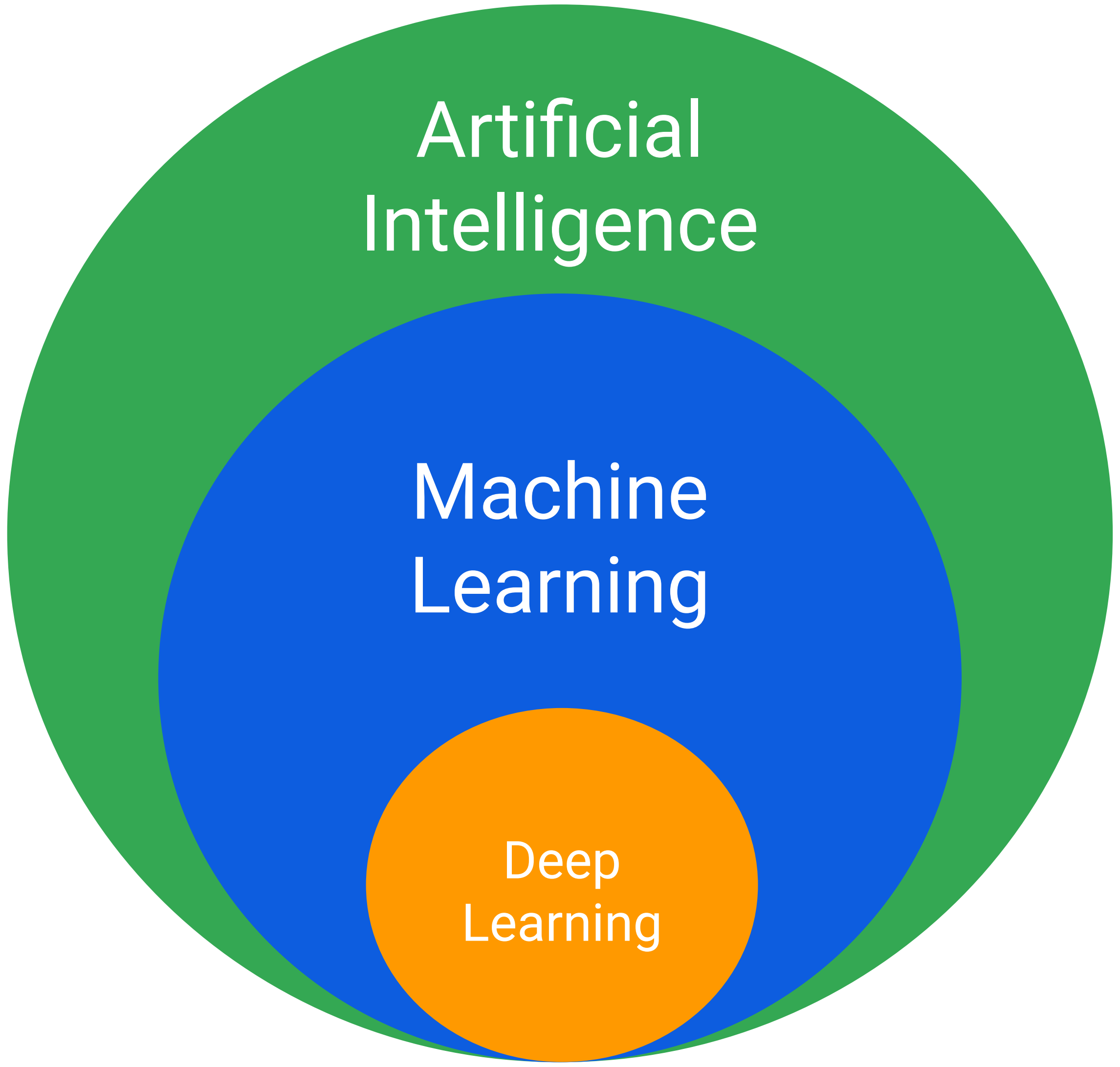


#GCPSketchnote
@PVERGADIA
THECLOUDGIRL.DEV
8.13.2020
How to build a scalable
DATA ANALYTICS PIPELINE
CAPTURE
Data ingestion at any scale
PROCESS
Reliable streaming data pipeline
STORE
Data lake and data warehousing
ANALYZE
Data warehousing
USE
Advanced analytics
CLOUD PUB/SUB
Scaled messaging platform
SaaS
DATA TRANSFER SERVICE
Fast data migration from saas apps
STORAGE TRANSFER SERVICE
Data migration from other cloud or on-prem
CLOUD IoT CORE
Stream events from IoT devices
CLOUD DATAFLOW
Stream and batch processing
hadoop + Spark
CLOUD DATAPROC
Managed Hadoop & Spark platform
CLOUD DATAPREP
Data prep using visual tool
CLOUD STORAGE
Use as your data lake for structured and unstructured data
BIGQUERY STORAGE
Cloud-native, highly scalable serverless data warehouse
BIGQUERY
Analysis engine
CLOUD AI PLATFORM
For machine learning
TENSORFLOW
For machine learning
LOOKER
For your analysis
SHEETS
For your analysis
CLOUD DATA FUSION
Fully managed, code-free data integration service to manage ETL/ELT pipelines and also track lineage of that data.
DATA CATALOG
Unified view of all datasets
CLOUD COMPOSER
Fully managed, workflow orchestration built on Apache Airflow

How Google does Machine Learning

Instructor: Ben Ahmed

Why are Machine Learning and Deep Learning so exciting?

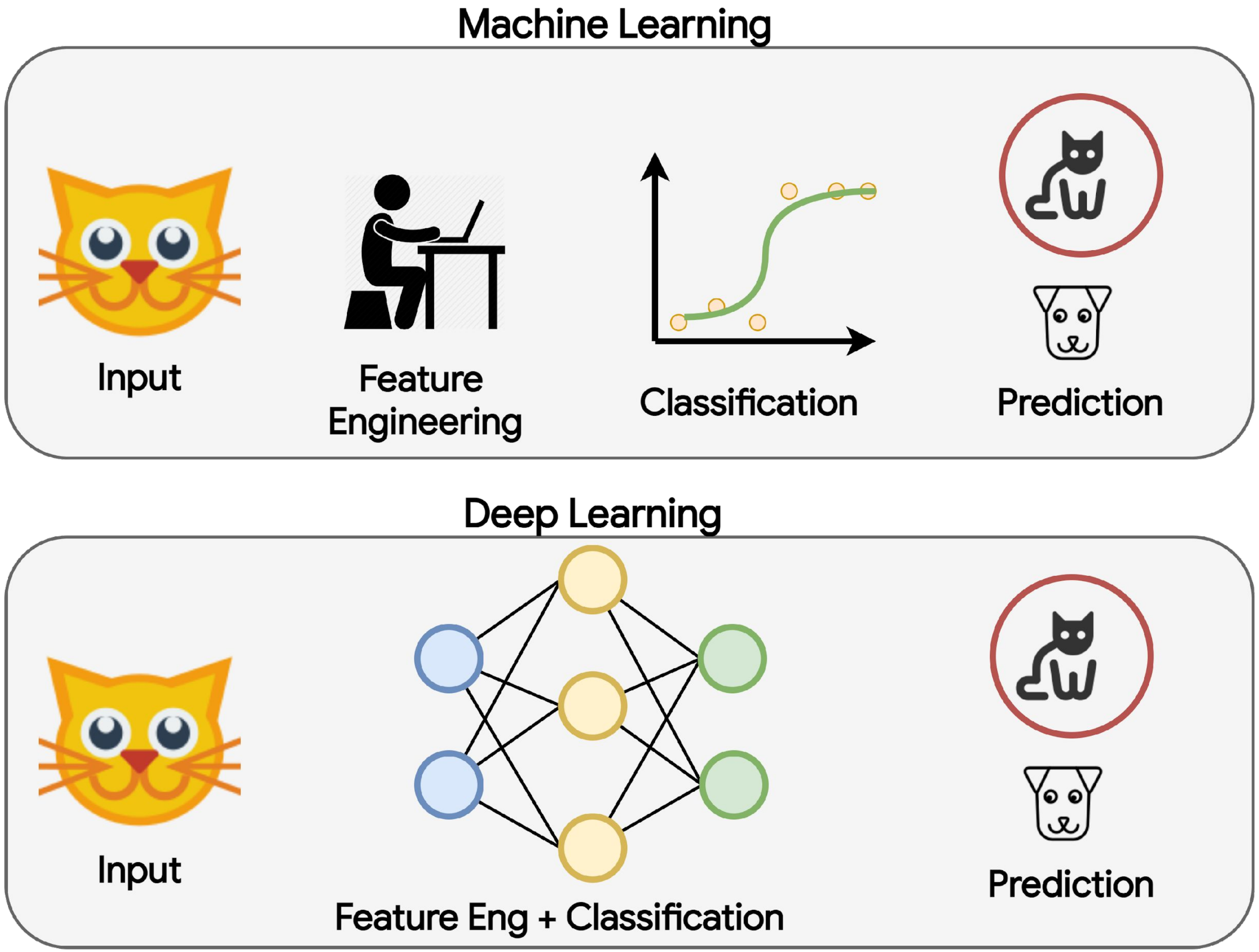


Class of problems we can solve when **computers think/act like humans**

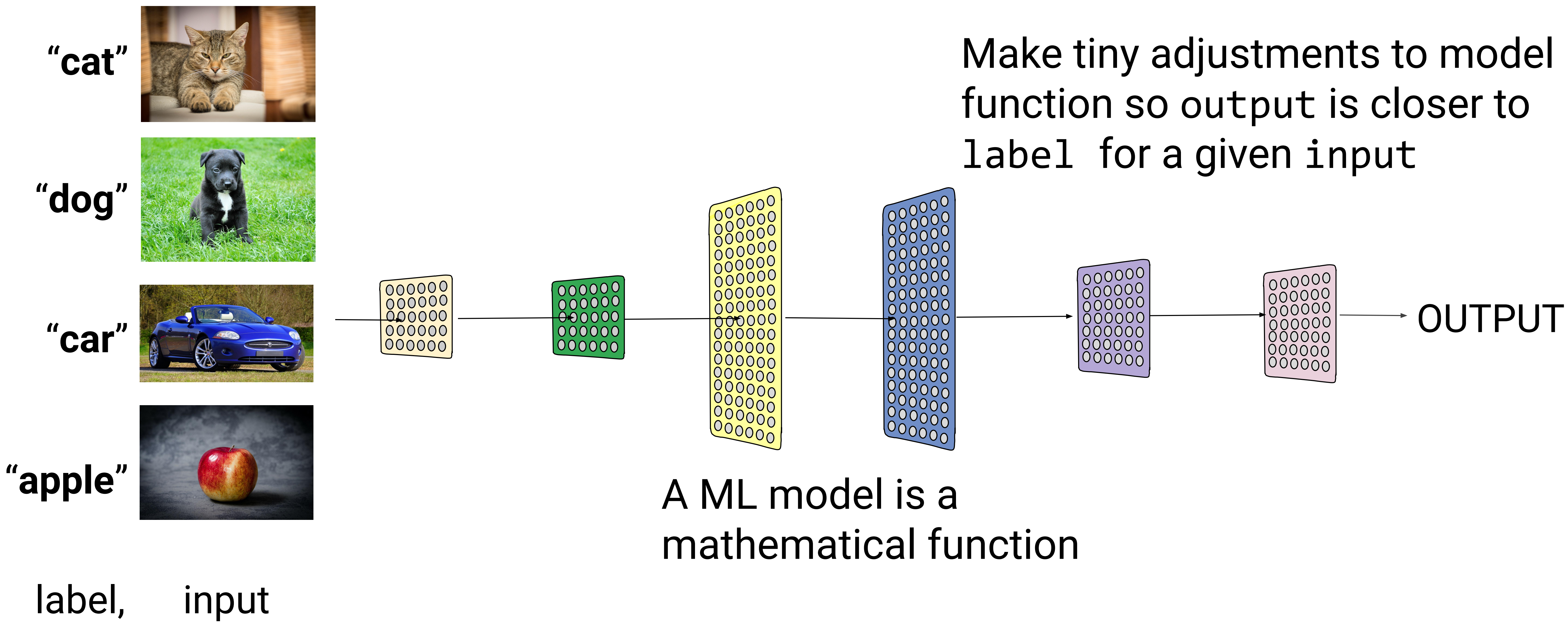
Scalably solve those problems using **data** examples (**not custom code**)

Even when that data consists of **unstructured data** like images, speech, video, natural language text, etc.

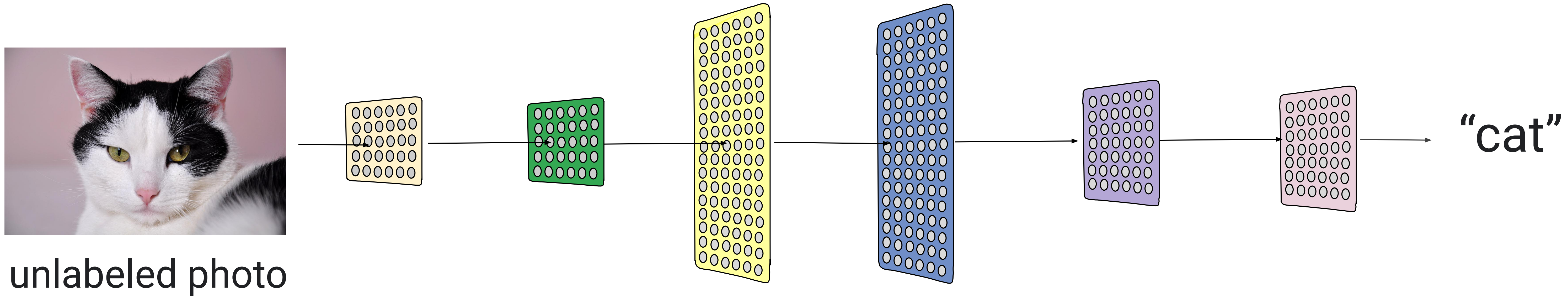
Machine Learning VS Deep Learning



Stage 1: Train an ML model with examples

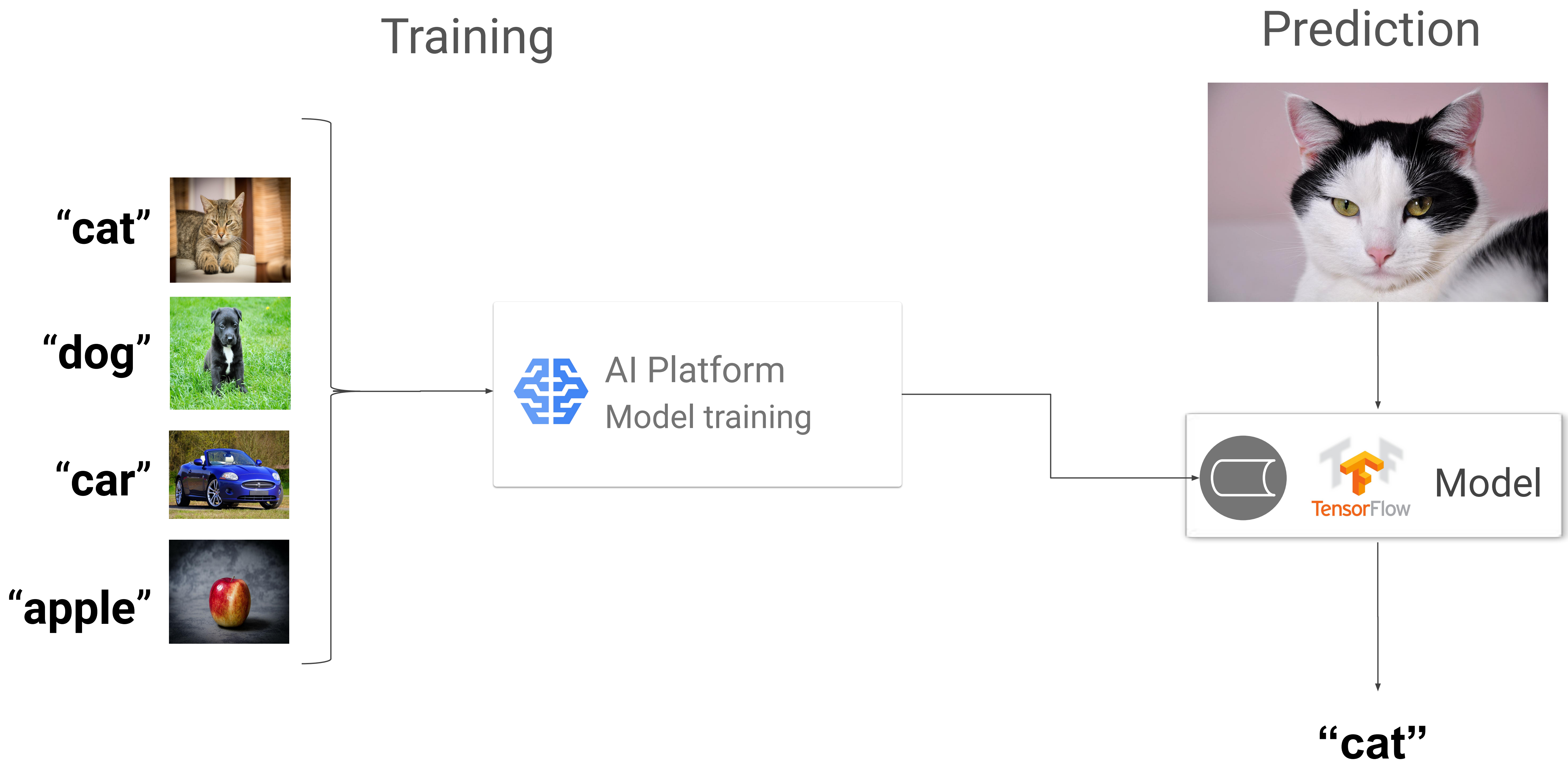


Stage 2: Predict with a trained model



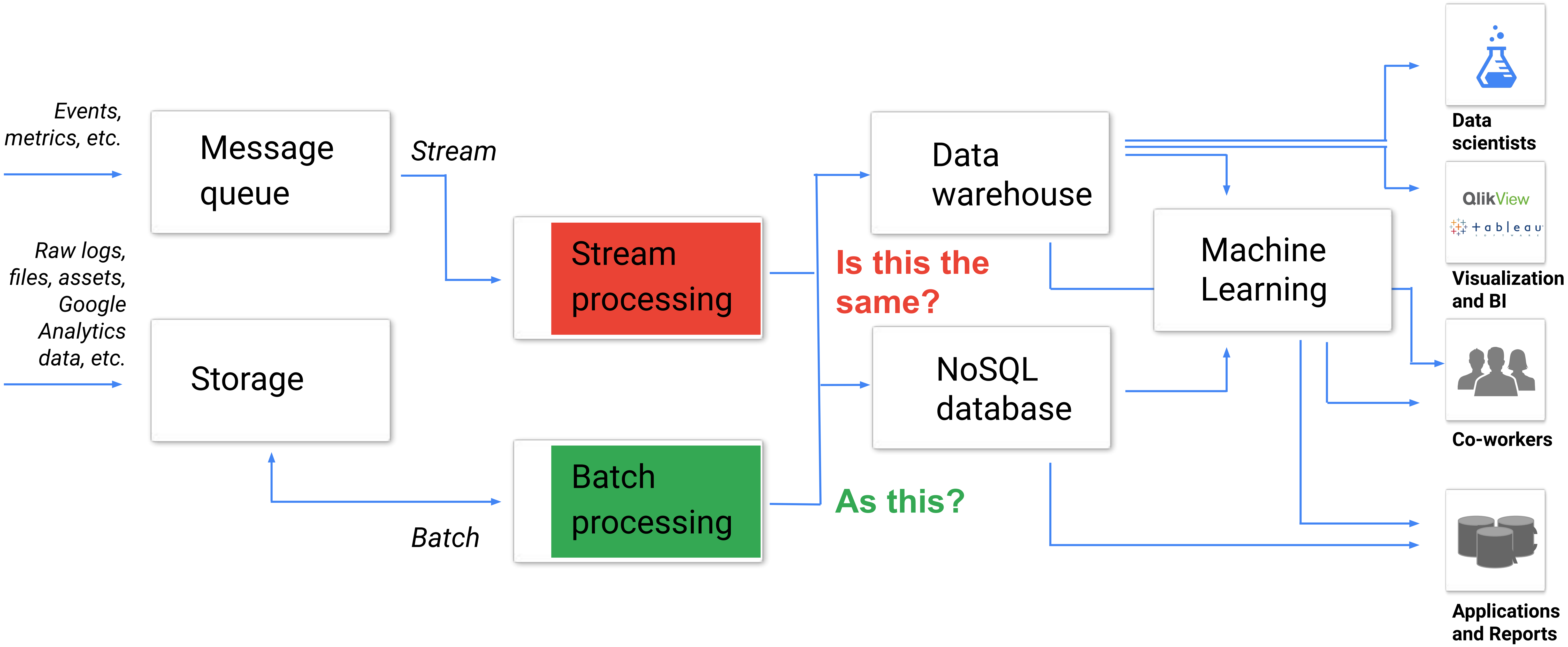
unlabeled photo

Data scientists must focus on both the training and prediction stages of ML

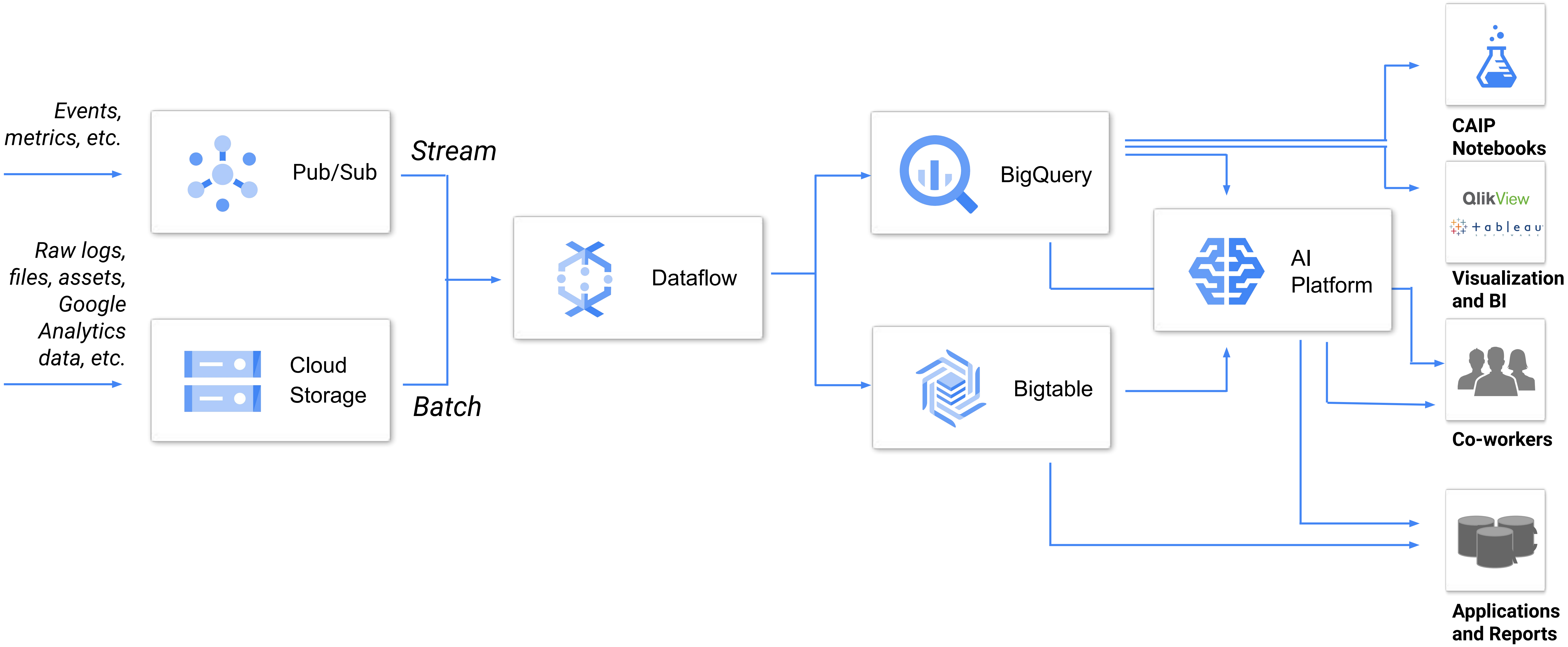


Google infuses Machine Learning into almost all its products

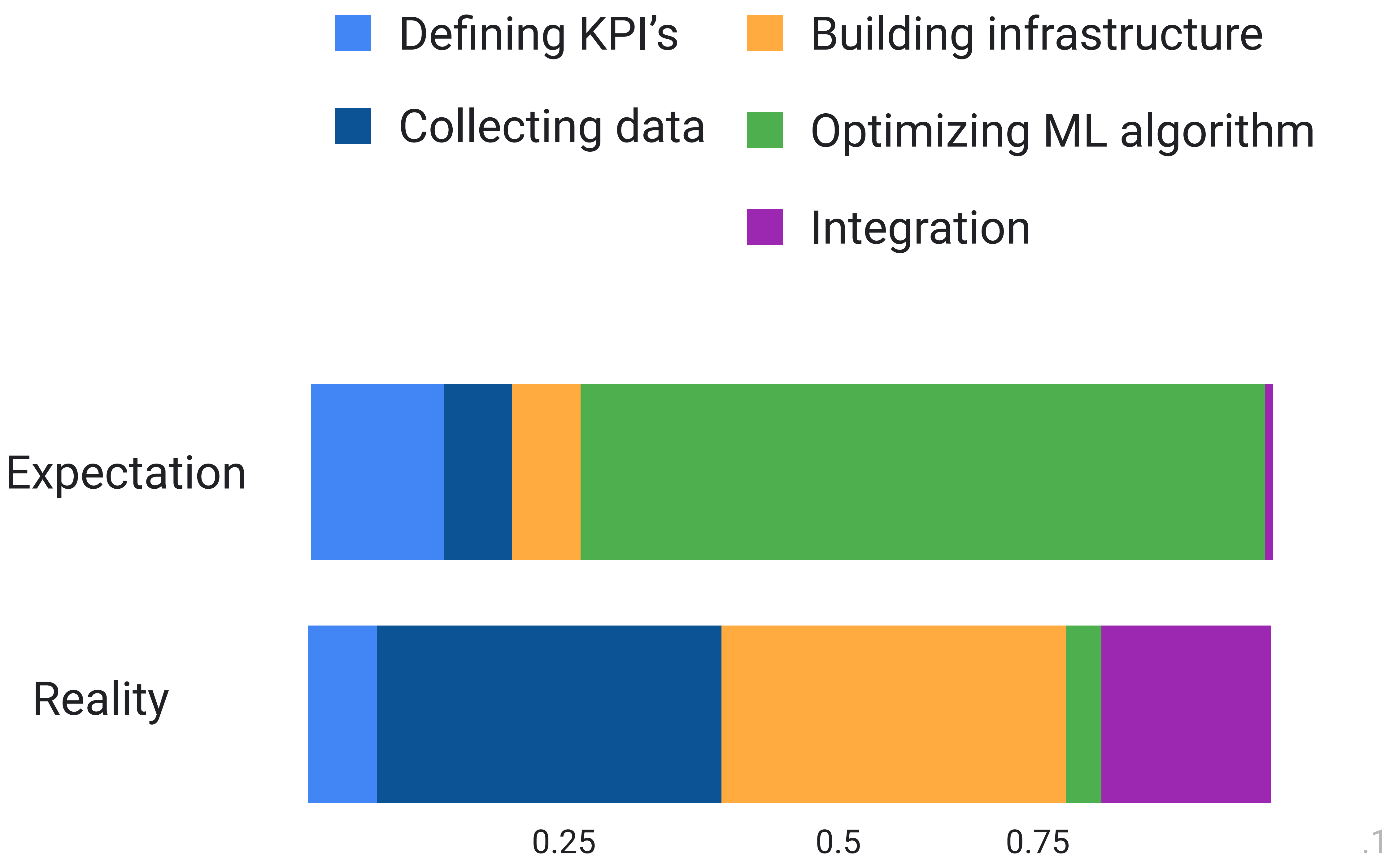
For machine learning, you need to build a streaming pipeline in addition to a batch pipeline



Sophistication around real-time data is key



ML effort allocation



Avoid these top 10 ML pitfalls

- Defining KPI's
- Collecting data
- Integration
- Infrastructure
- Optimizing ML

1. ML requires just as much software infrastructure
2. No data collected yet
3. Assume the data is ready for use
4. Keep humans in the loop
5. Product launch focused on the ML algorithm
6. ML optimizing for the wrong thing
7. Is your ML improving things in the real world
8. Using a pre-trained ML algorithm vs building your own
9. ML algorithms are trained more than once
10. Trying to design your own perception or NLP algorithm

Get your hands dirty by practicing with technical skills

Machine Learning Basics: Algorithms

Part1

Instructor: Ben Ahmed

Section Agenda

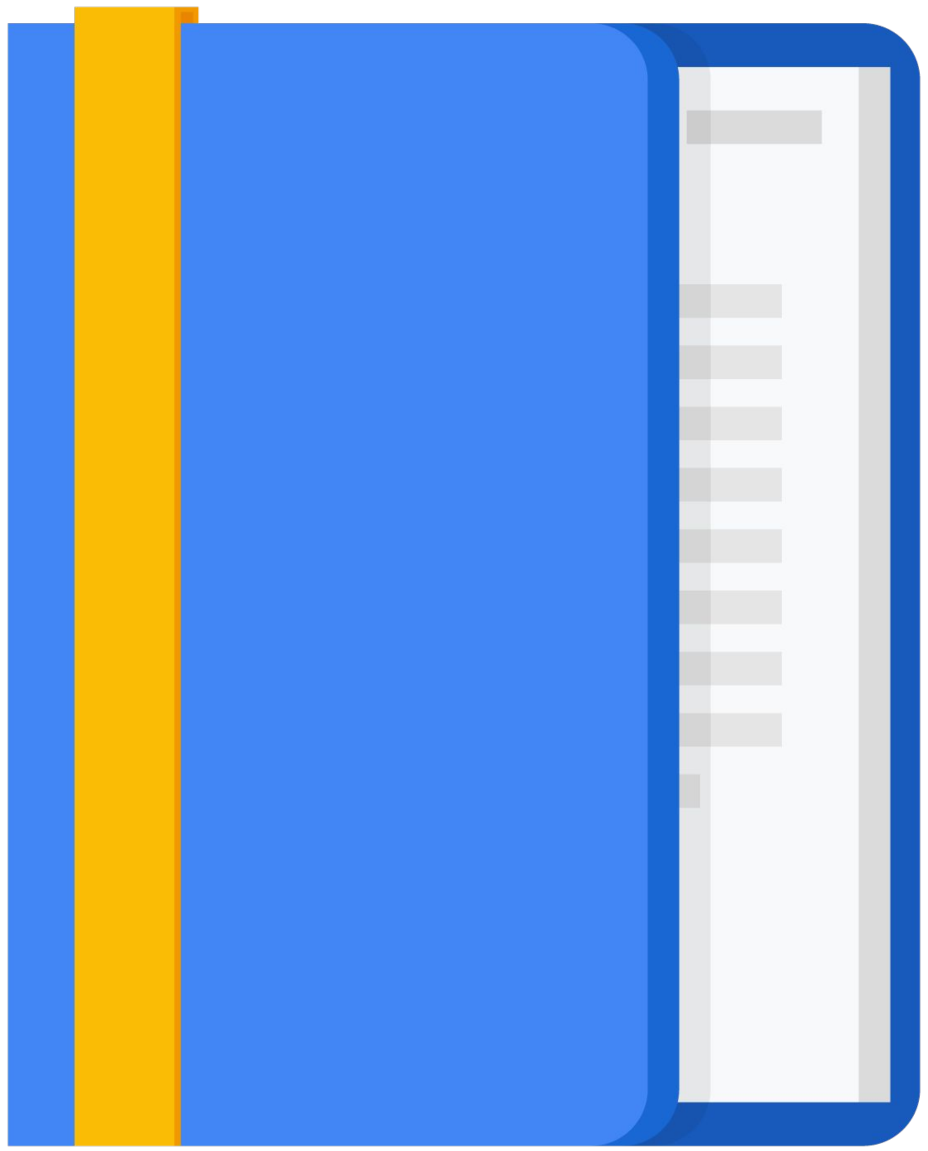
Introduction to ML Algorithms

Unsupervised Learning Algorithms

Supervised Learning Algorithms

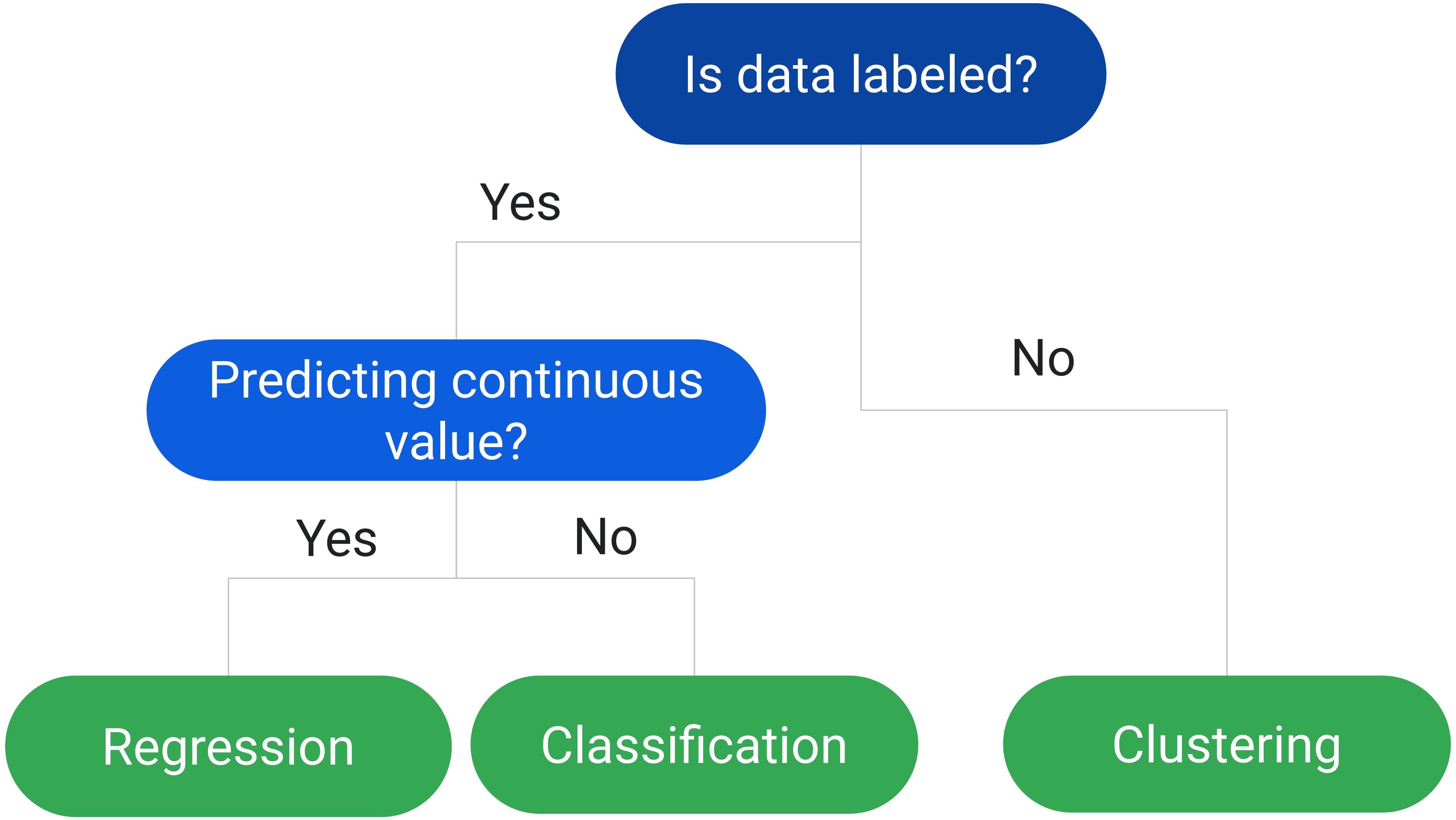
Loss functions

Gradient Descent



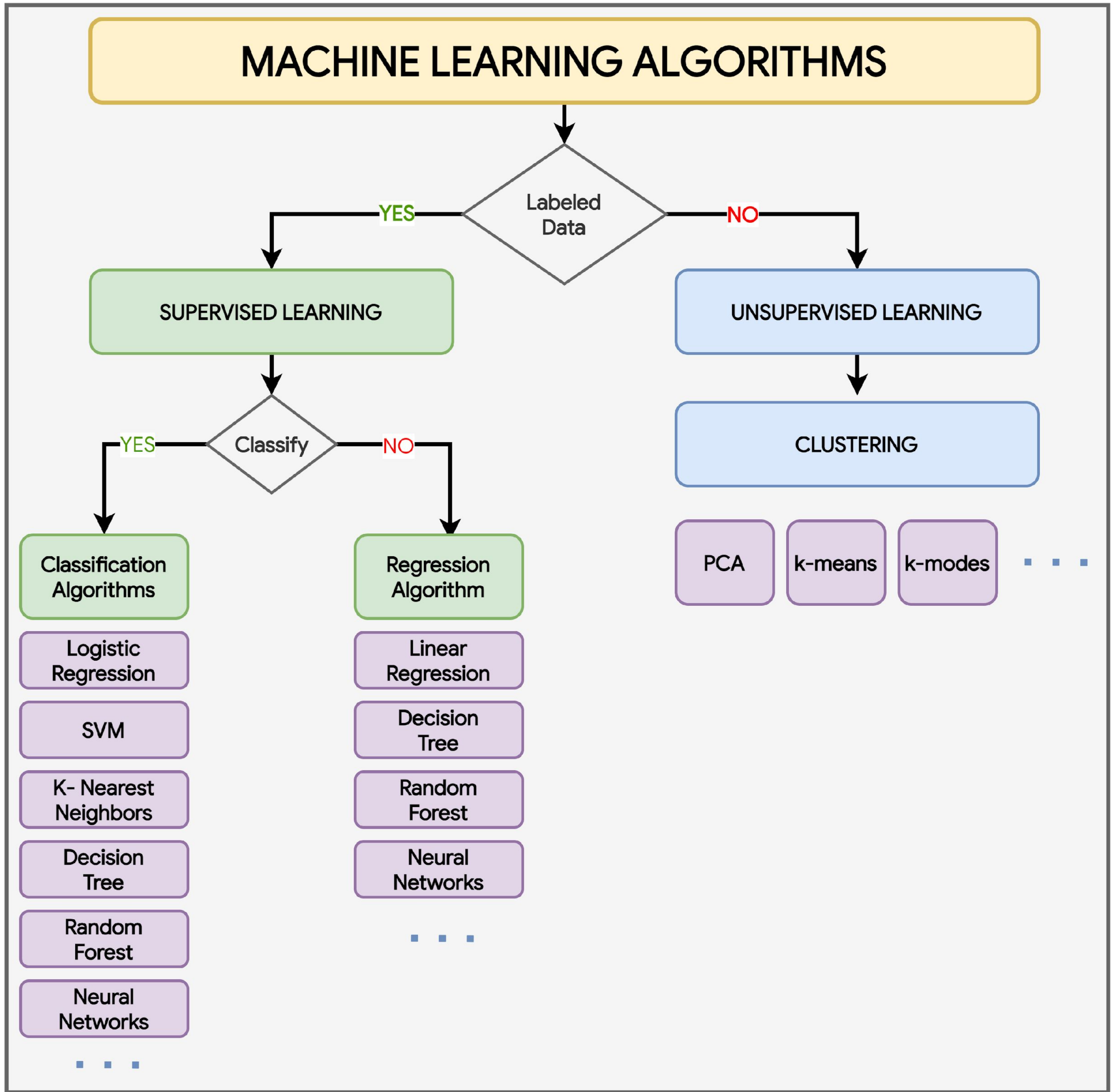
Needed For the Exam

The type of ML problem depends on whether or not you have labeled data and what you are interested in predicting



Needed For the Exam

ML Algorithms…Which one should I pick



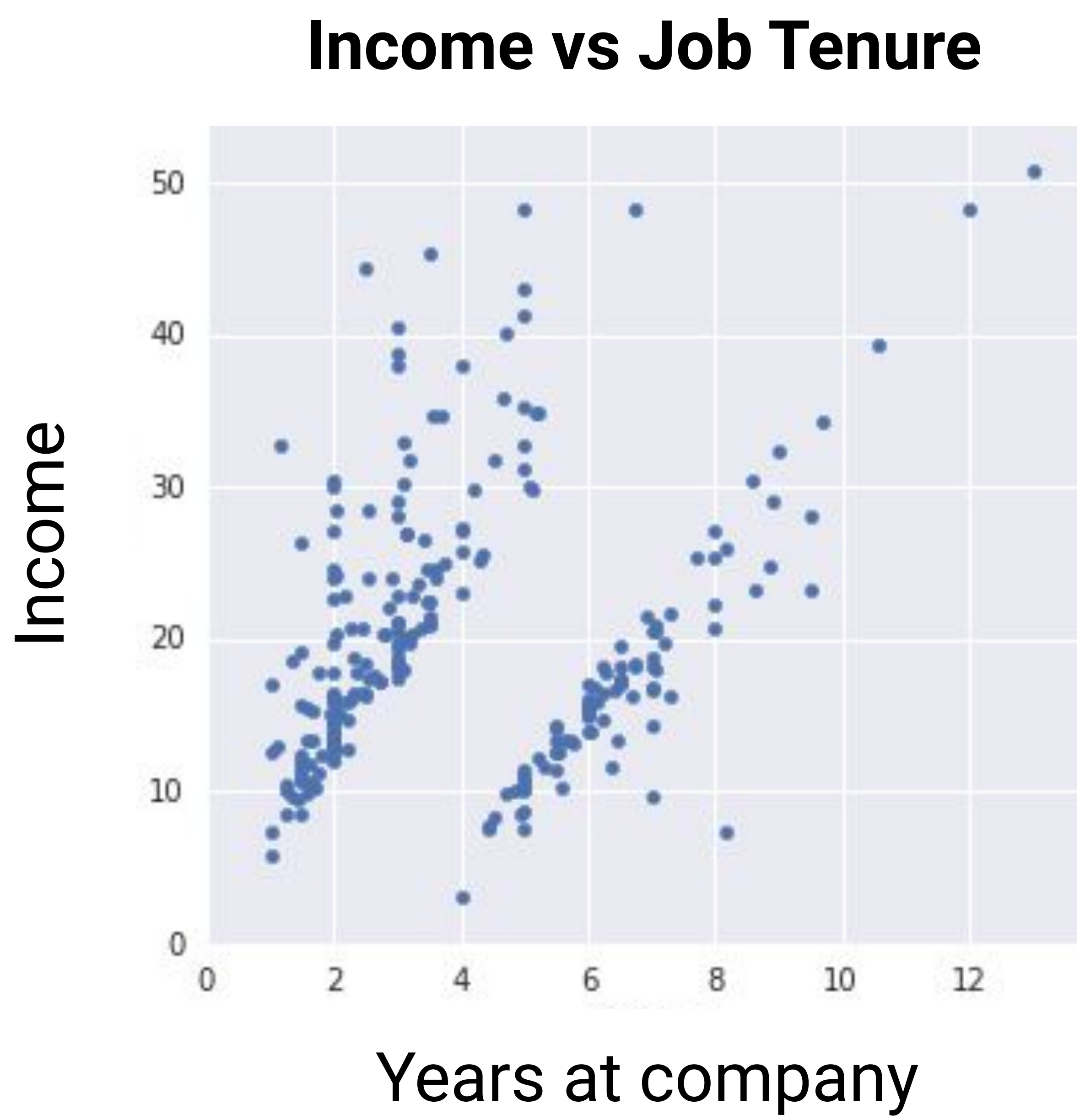
FYI: For your knowledge

Unsupervised and supervised learning are the two types of ML algorithms

Example Model: Clustering

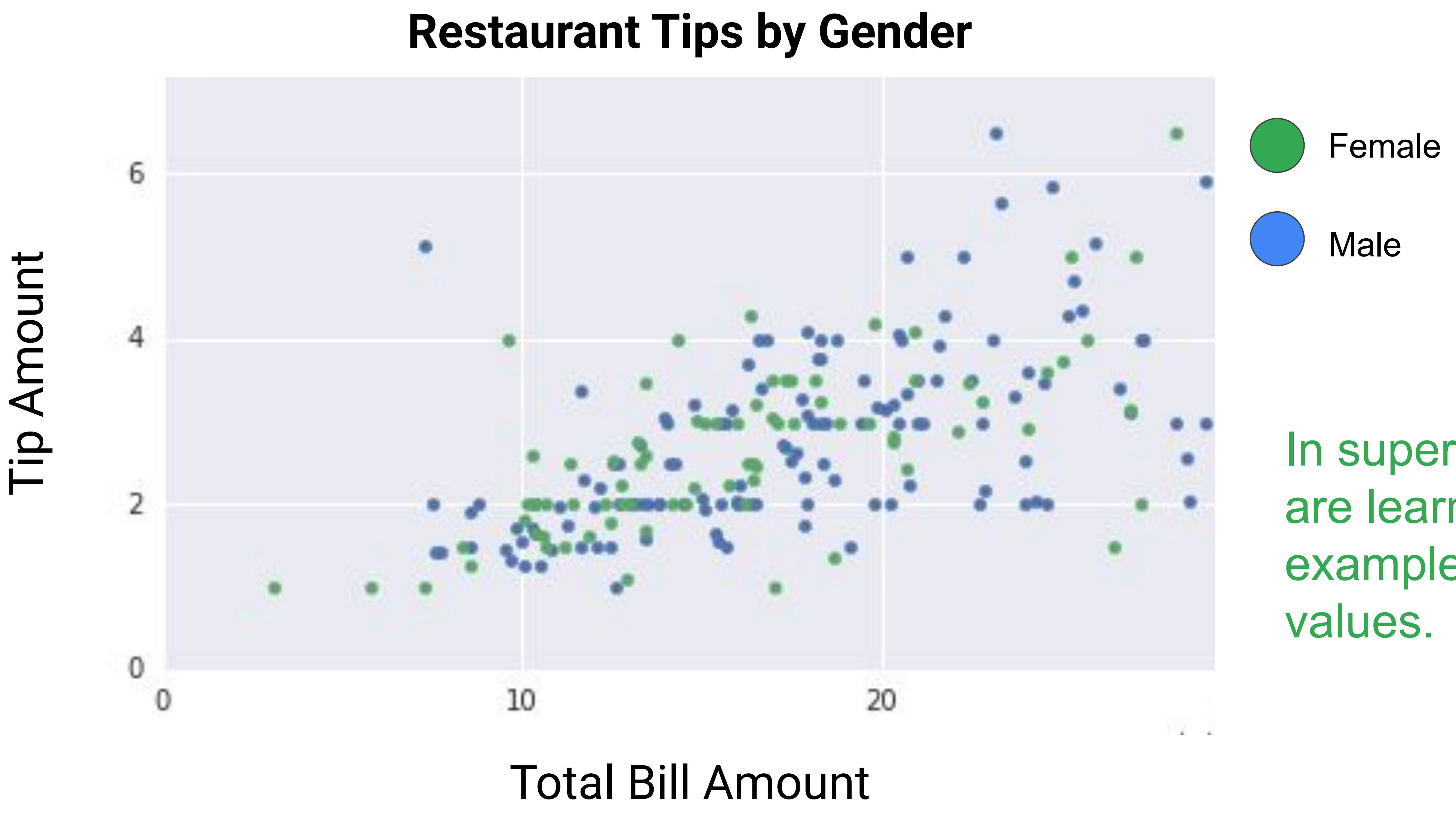
Is this employee on the “fast-track” or not?

In unsupervised learning, data is not labeled.



FYI: For your knowledge

Supervised learning implies the data is already labeled



In supervised learning we are learning from past examples to predict future values.

Needed For the Exam

Regression and classification are supervised ML model types

	total_bill	tip	sex	smoker	day	time
1	total_bill	tip	sex	smoker	day	time
2	16.99	1.01	Female	No	Sun	Dinner
3	10.34	1.66	Male	No	Sun	Dinner
4	21.01	3.5	Male	No	Sun	Dinner
5	23.68	3.31	Male	No	Sun	Dinner
6	24.59	3.61	Female	No	Sun	Dinner
7	25.29	4.71	Male	No	Sun	Dinner
8	8.77	2	Male	No	Sun	Dinner
9	26.88	3.12	Male	No	Sun	Dinner

Option 1
Regression Model
Predict the tip amount

Option 2
Classification Model
Predict the sex of the customer

Needed For the Exam

Quiz: Supervised learning

Imagine you are in banking and you are creating an ML model for detecting if transactions are fraudulent or not. Is this classification or regression and why?

A. Regression, categorical label
B. Regression, continuous label
C. Classification, categorical label
D. Classification, continuous label

Needed For the Exam

Quiz: Supervised learning

Imagine you are in banking and you are creating an ML model for detecting if transactions are fraudulent or not. Is this classification or regression and why?

A. Regression, categorical label
B. Regression, continuous label
C. Classification, categorical label
D. Classification, continuous label

Section Agenda

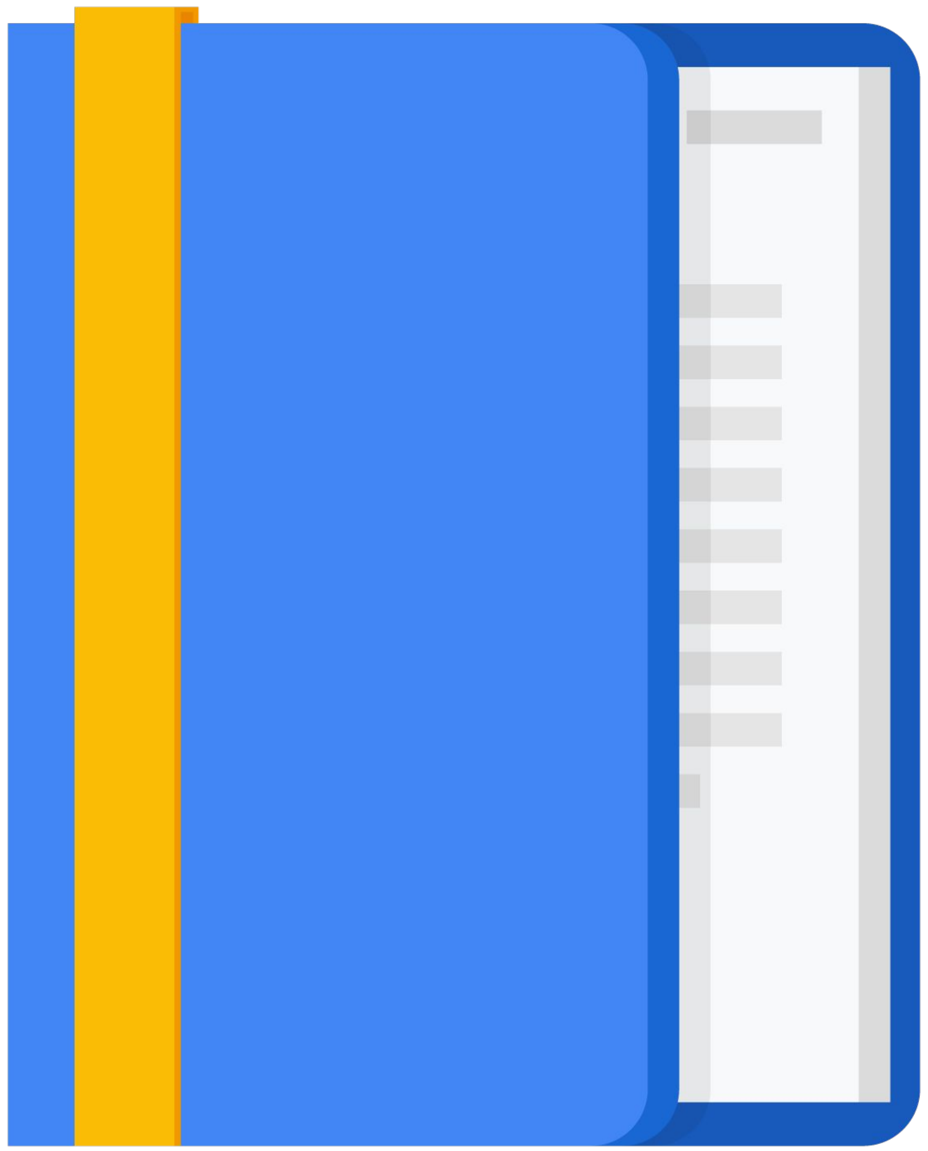
Introduction to ML Algorithms

Unsupervised Learning Algorithms

Supervised Learning Algorithms

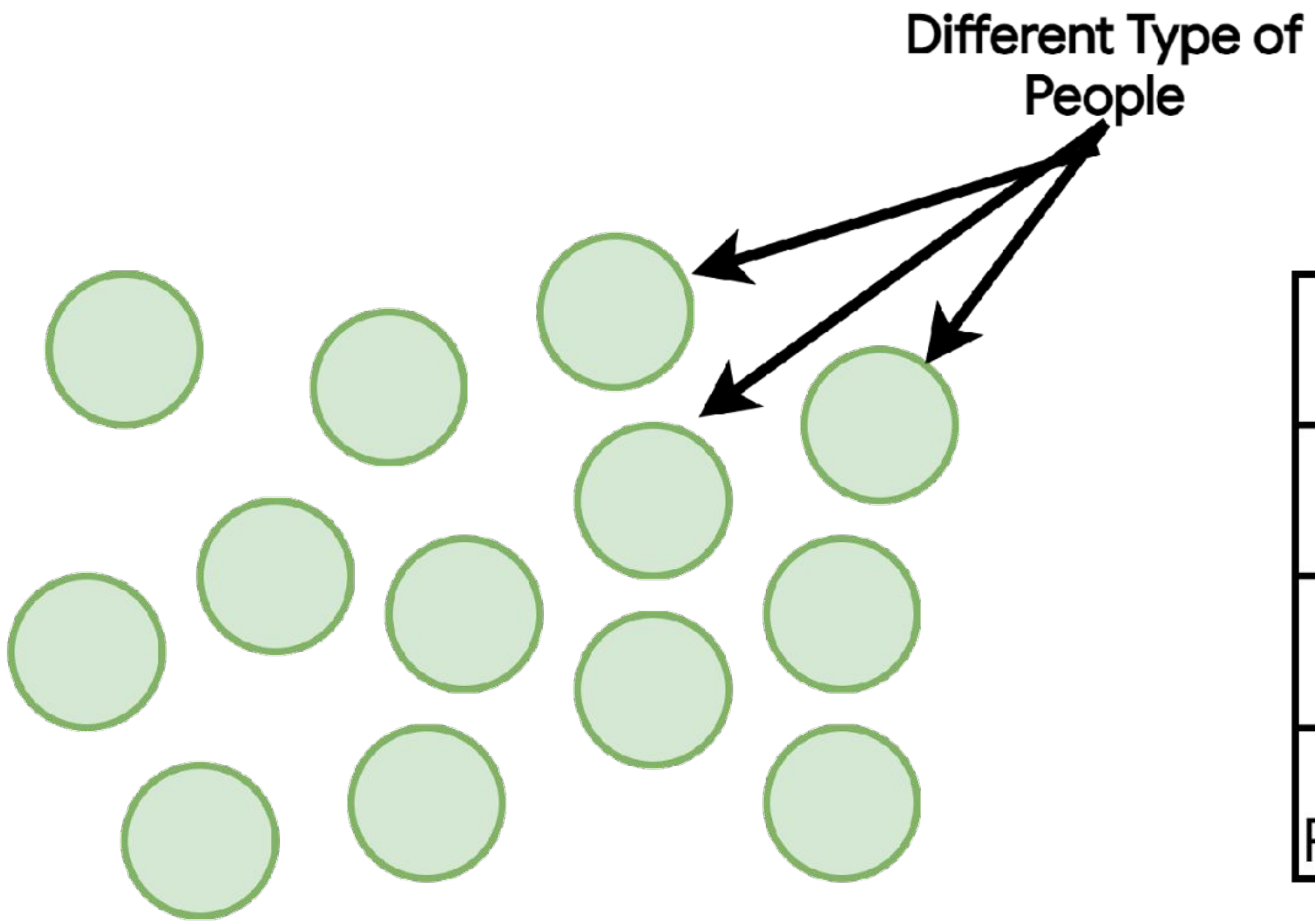
Loss functions

Gradient Descent



Needed For the Exam

Clustering with PCA



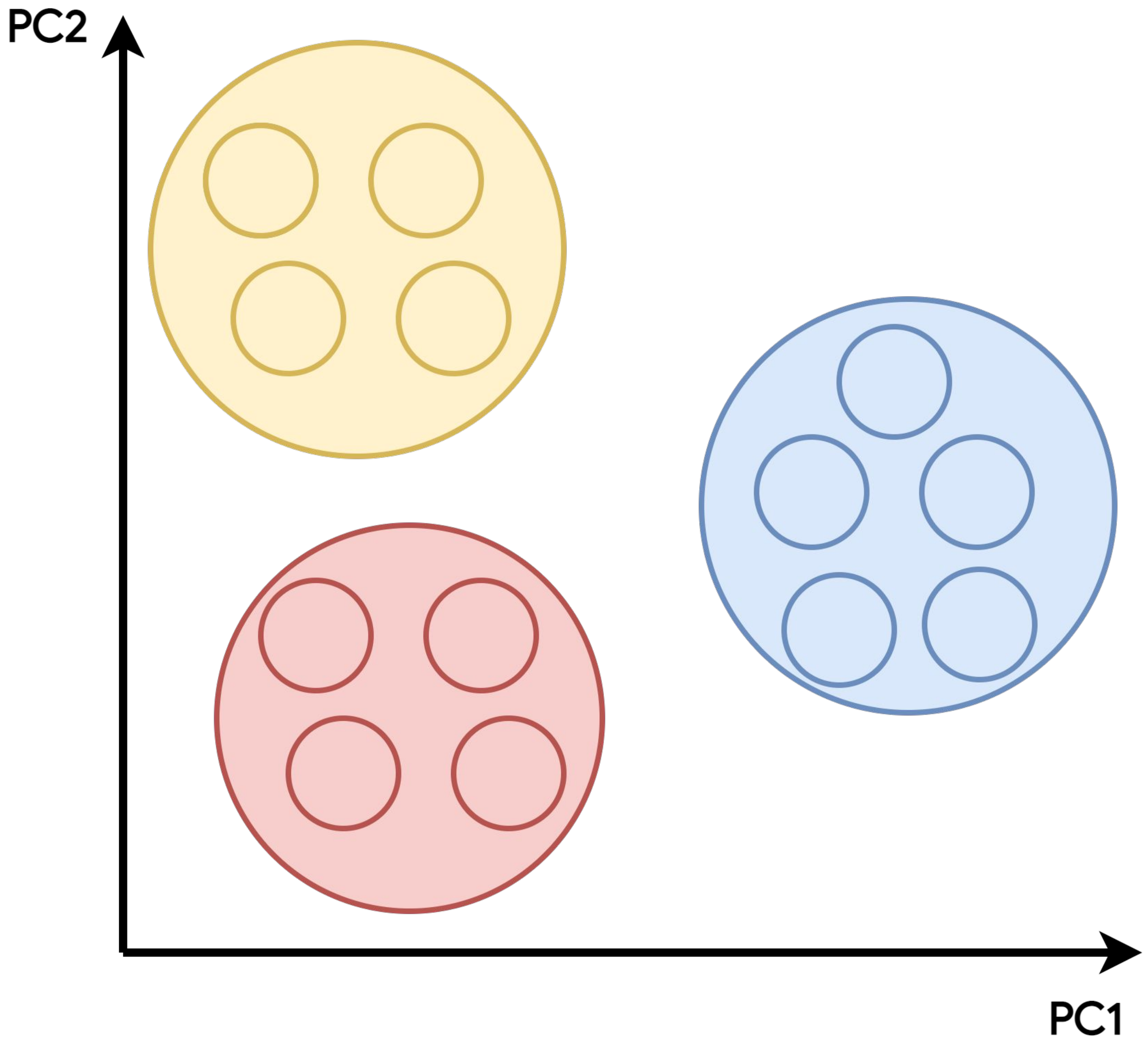
	Type1	Type2	Type3	Type4	Type5	....
Height	1.73	1.84	1.54	1.68	1.88	
Weight	78	65	69	75	90	
Blood Pressure	120	110	90	130	150	

Needed For the Exam

Clustering with PCA... Correlation matters

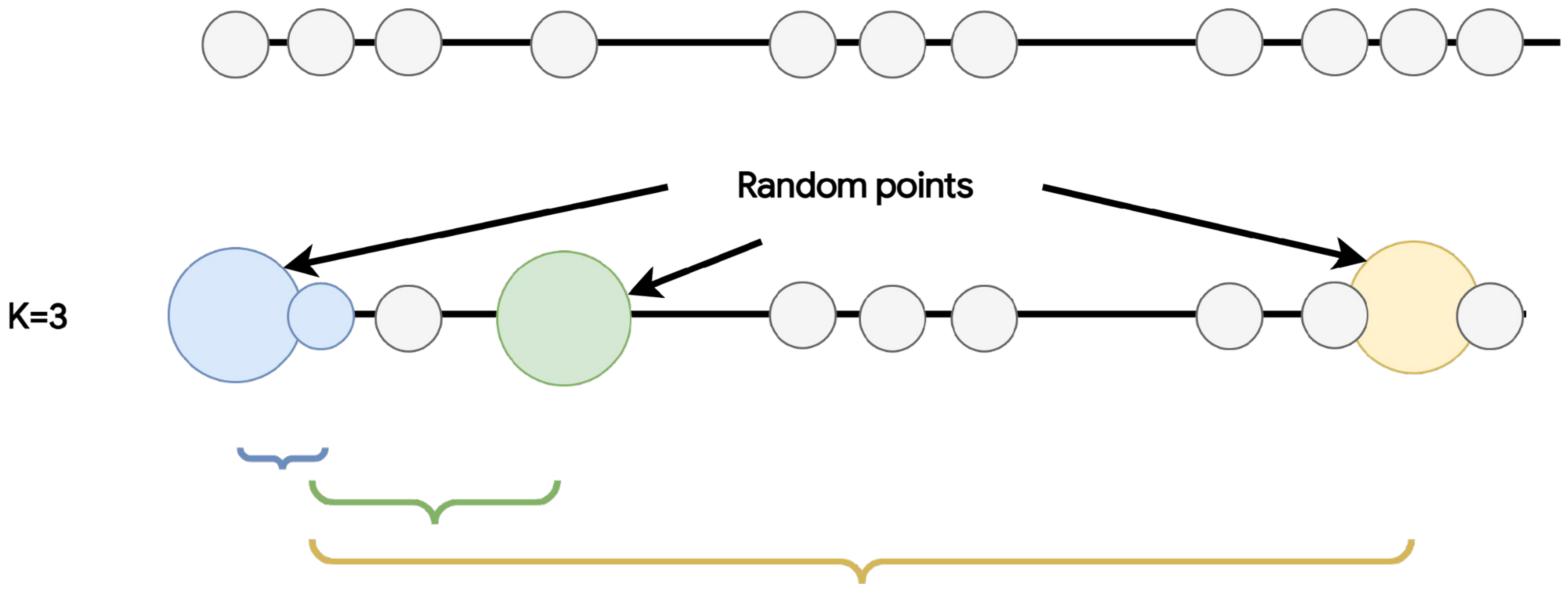
	Type1	Type2	Type3	Type4	Type5
Height	1.73	1.84	1.54	1.68	1.88
Weight	78	65	69	75	90
Blood Pressure	120	110	90	130	150

- Features that are highly correlated cluster together



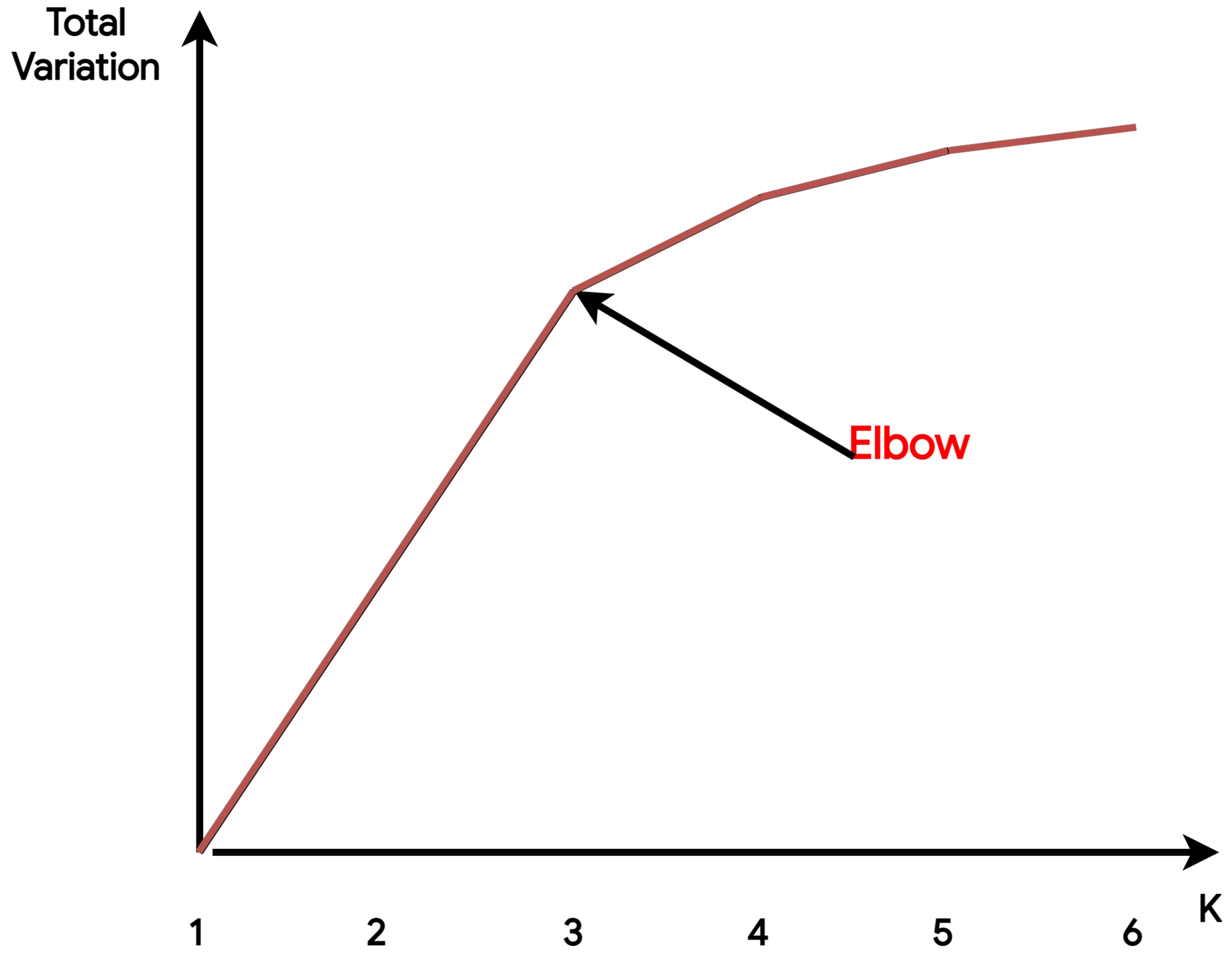
Needed For the Exam

Clustering with K-means… Pick some points



FYI: For your knowledge

Clustering with K-means (4).. How to choose K



Section Agenda

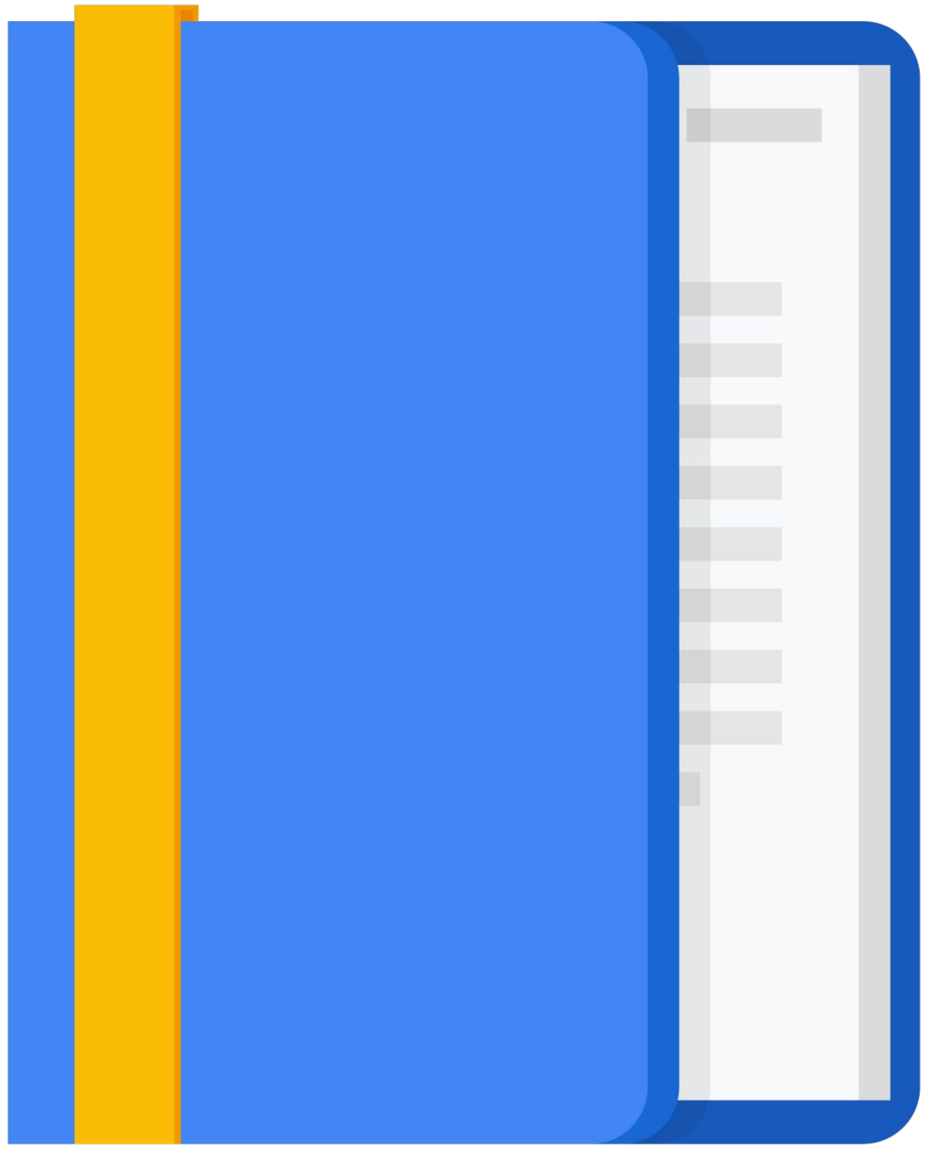
Introduction to ML Algorithms

Unsupervised Learning Algorithms

Supervised Learning Algorithms

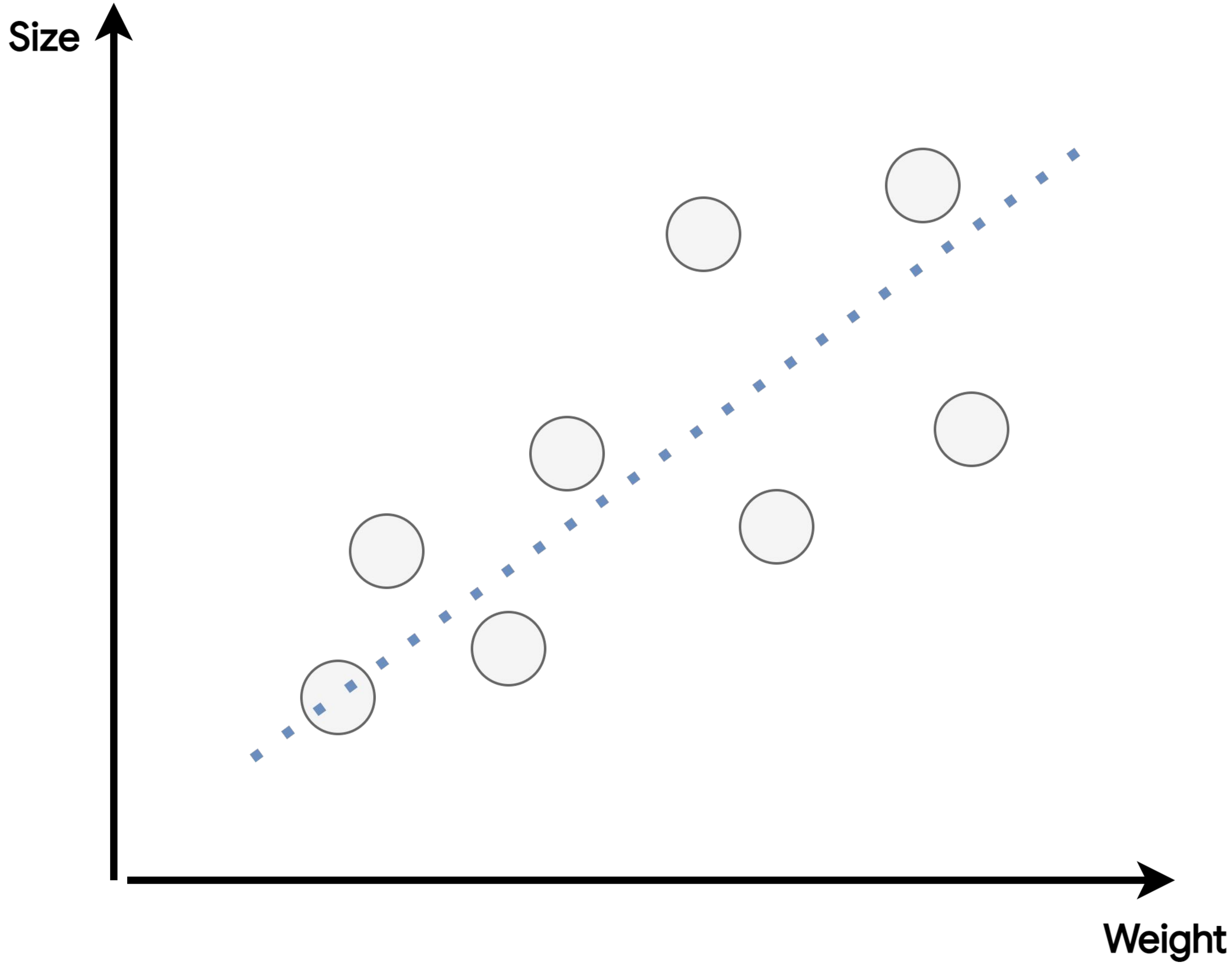
Loss functions

Gradient Descent



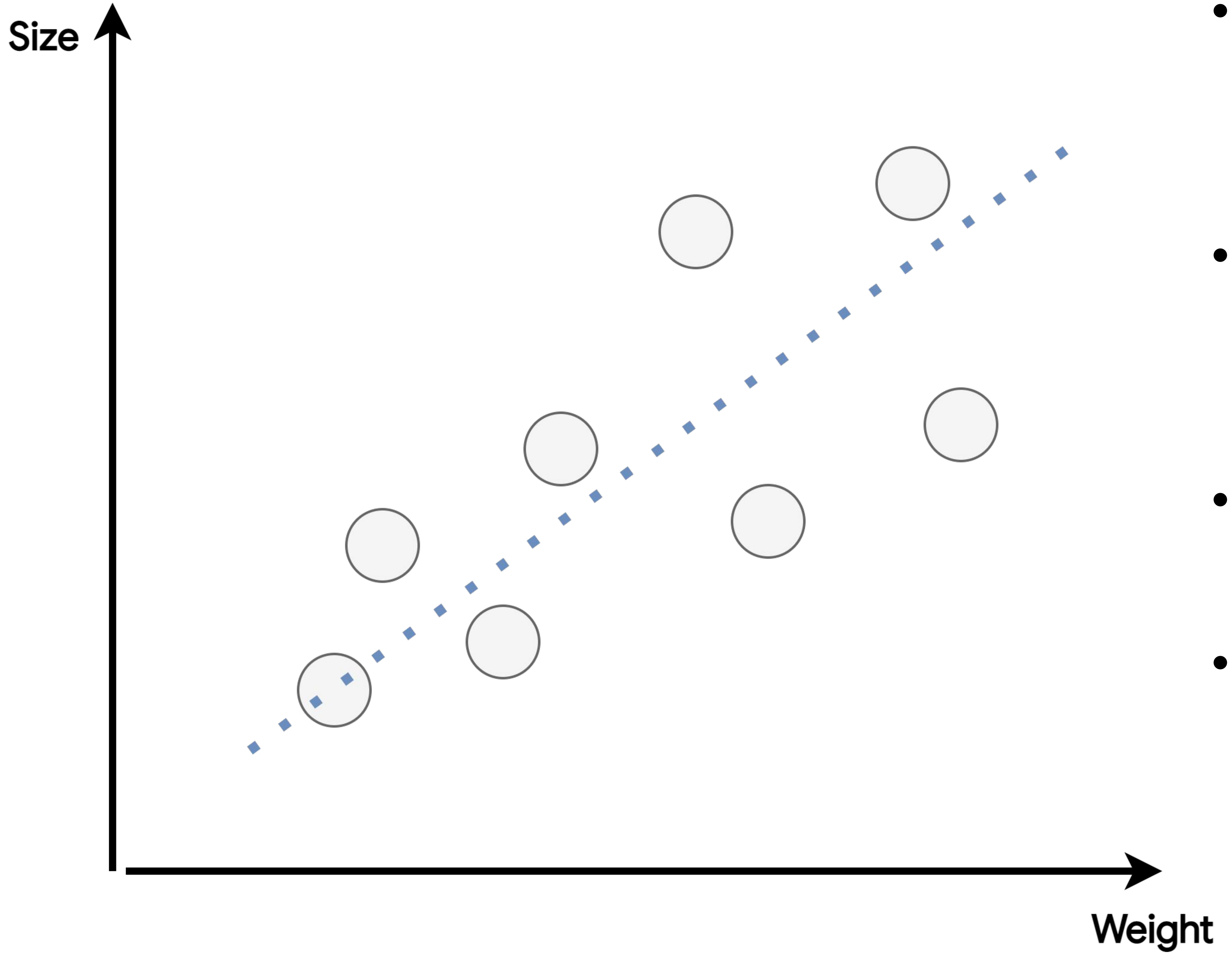
Needed For the Exam

Linear Regression… features



Needed For the Exam

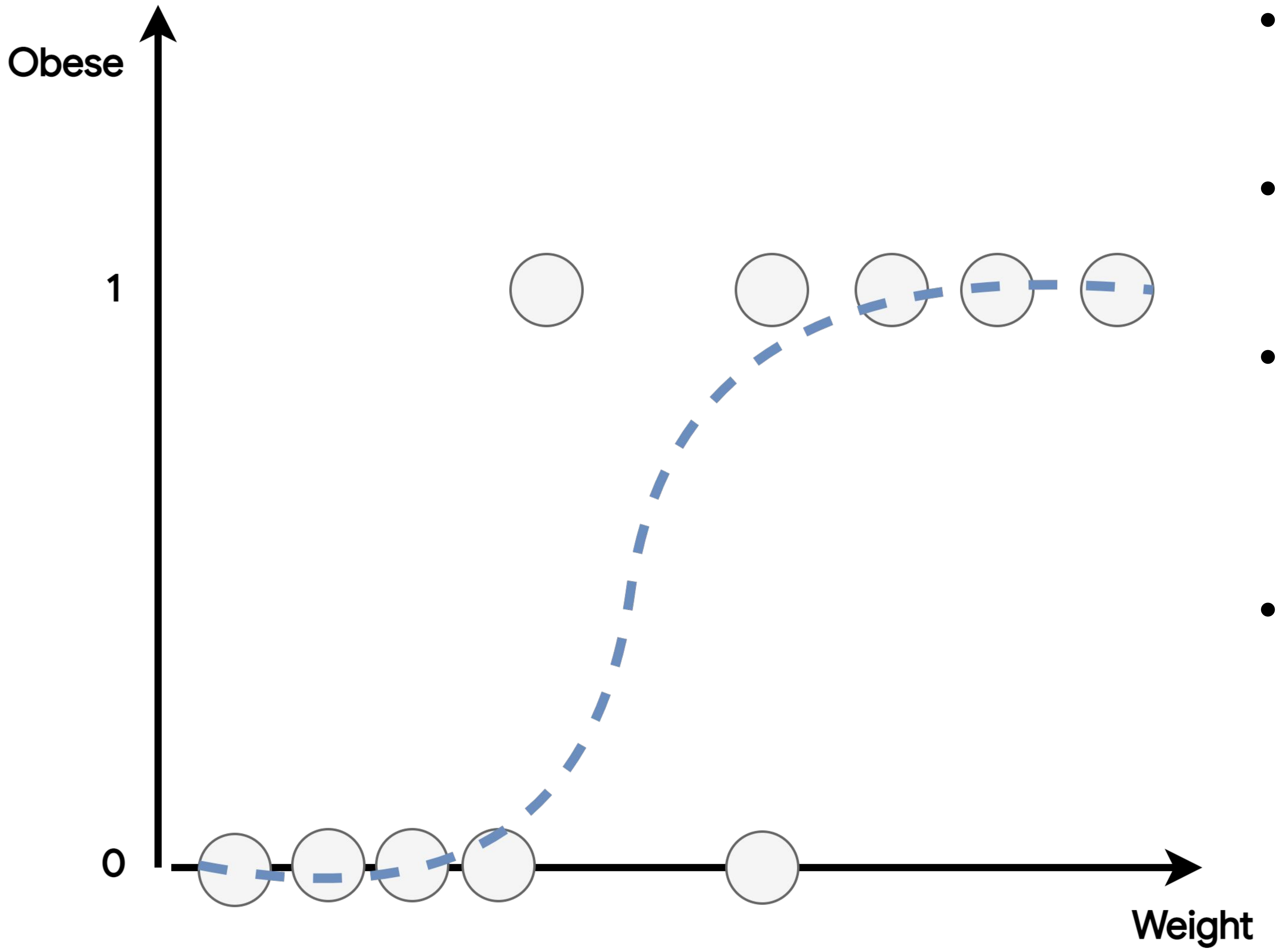
Linear Regression… features



- Determine feature correlation using R Square
- Compute p-value to get how much R Square is statistically significant
- Prediction
- Can work with multi-dimension

Needed For the Exam

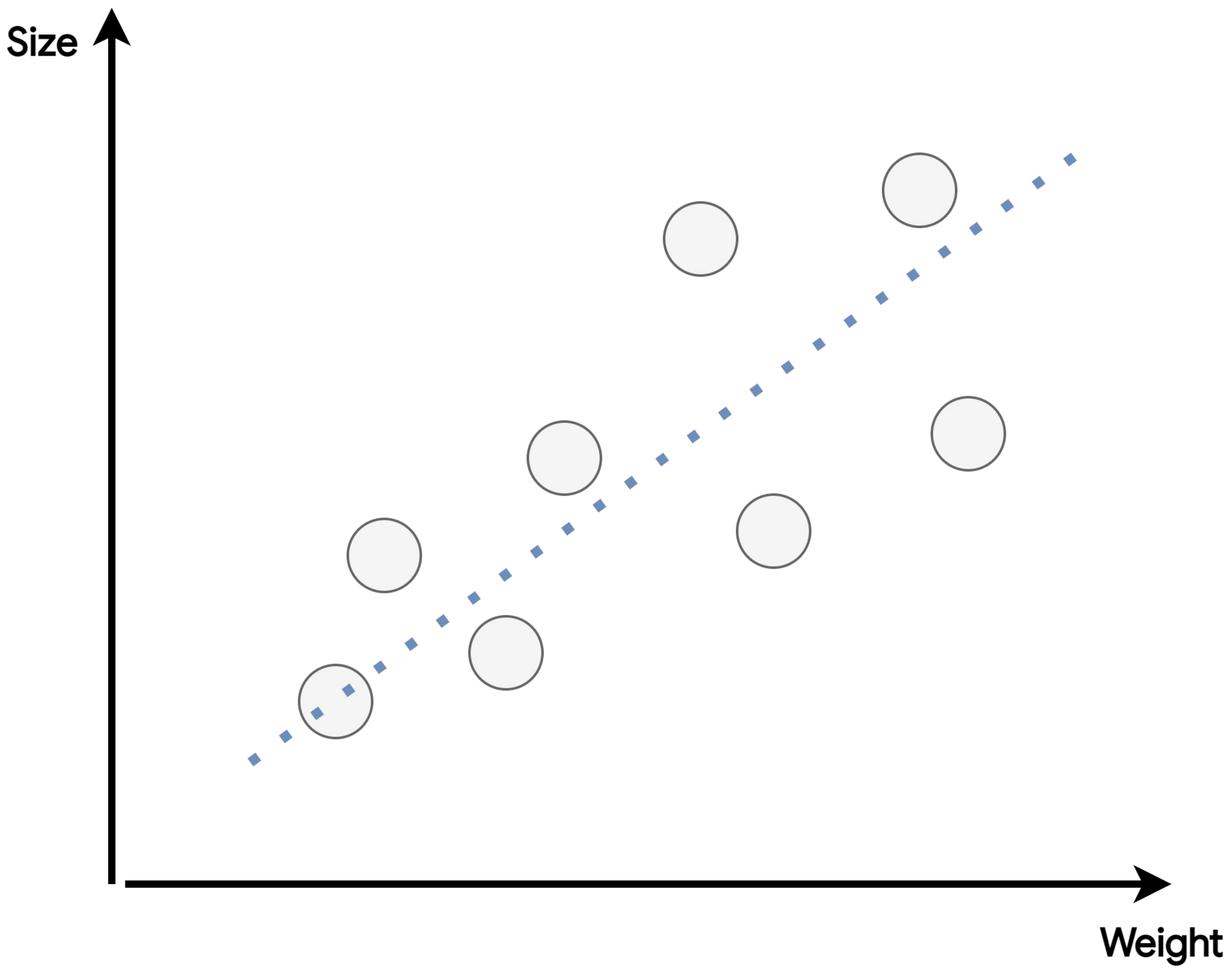
Logistic Regression… some features

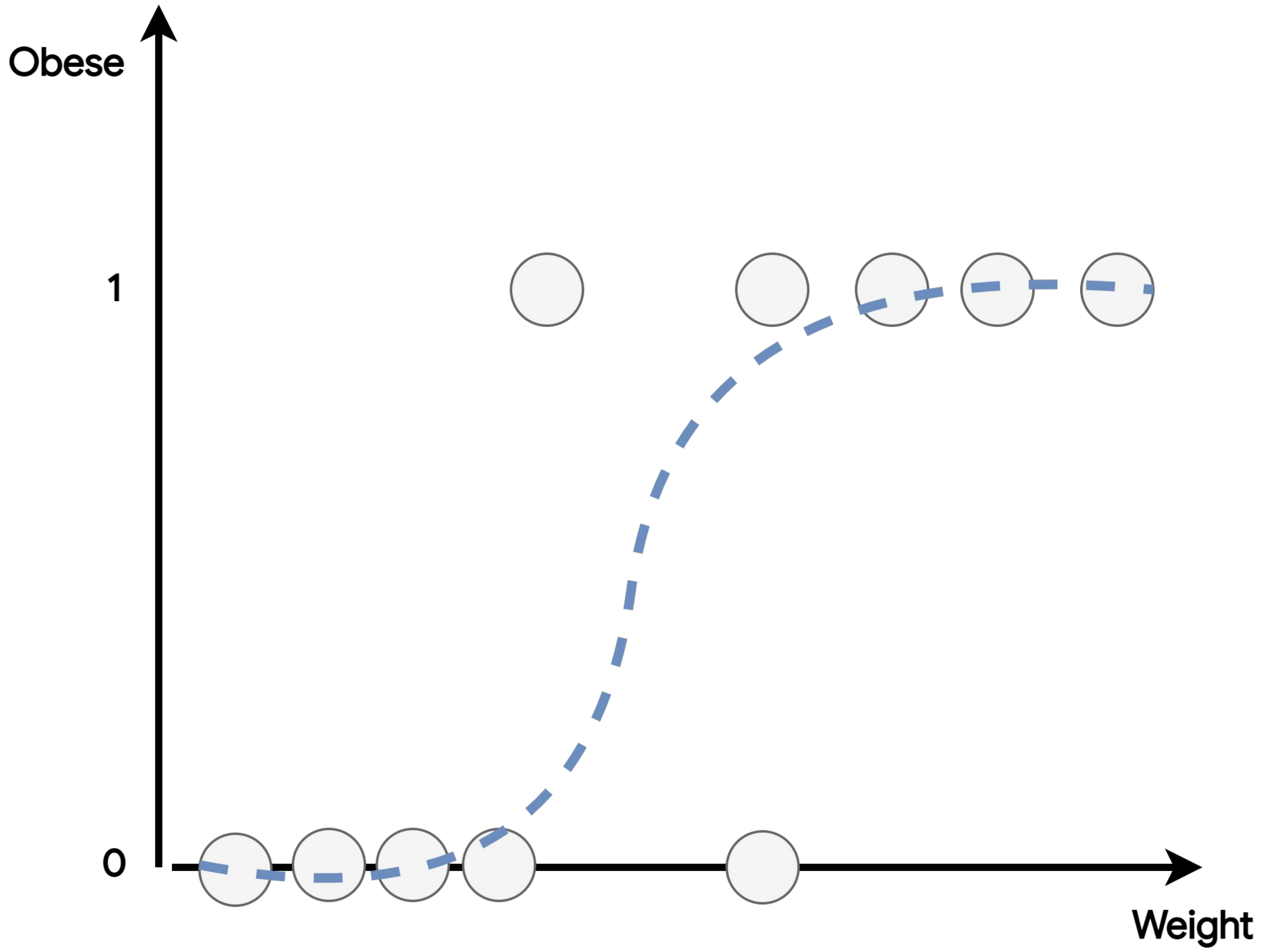


- Predict the probability
- Classify samples
- Can use different type of data (continuous and discrete)
- Can work with multi-dimension

Needed For the Exam

Linear Regression <versus > Logistic Regression

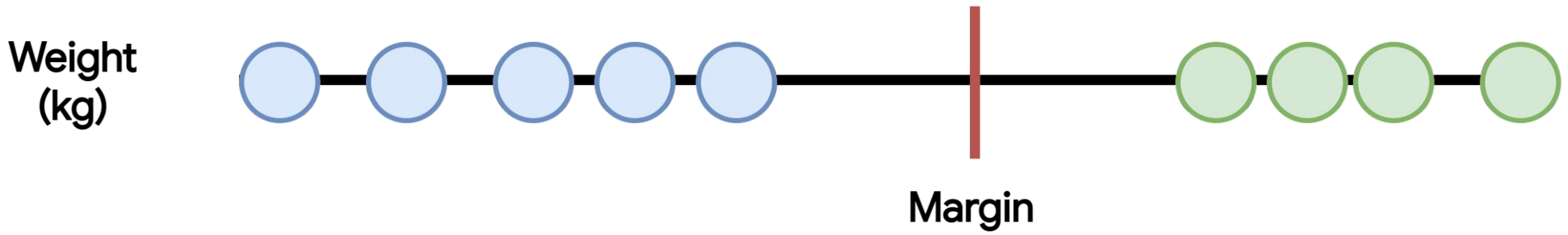




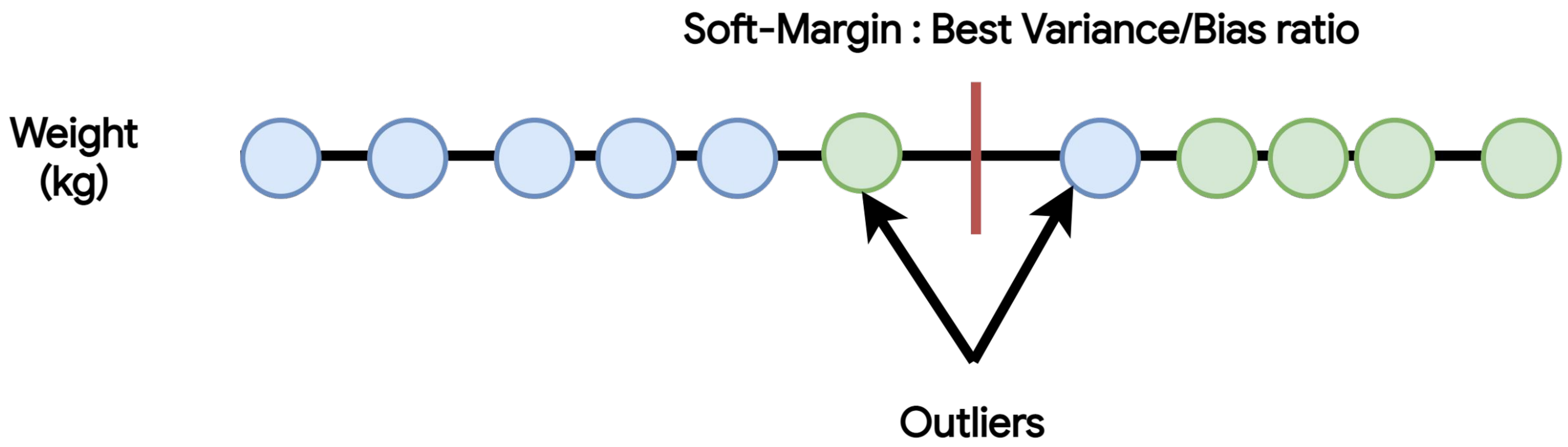
Needed For the Exam

Support Vector Machines… Support Vector Classifier

- We want to classify people based on the weight. We can start with using the Margin

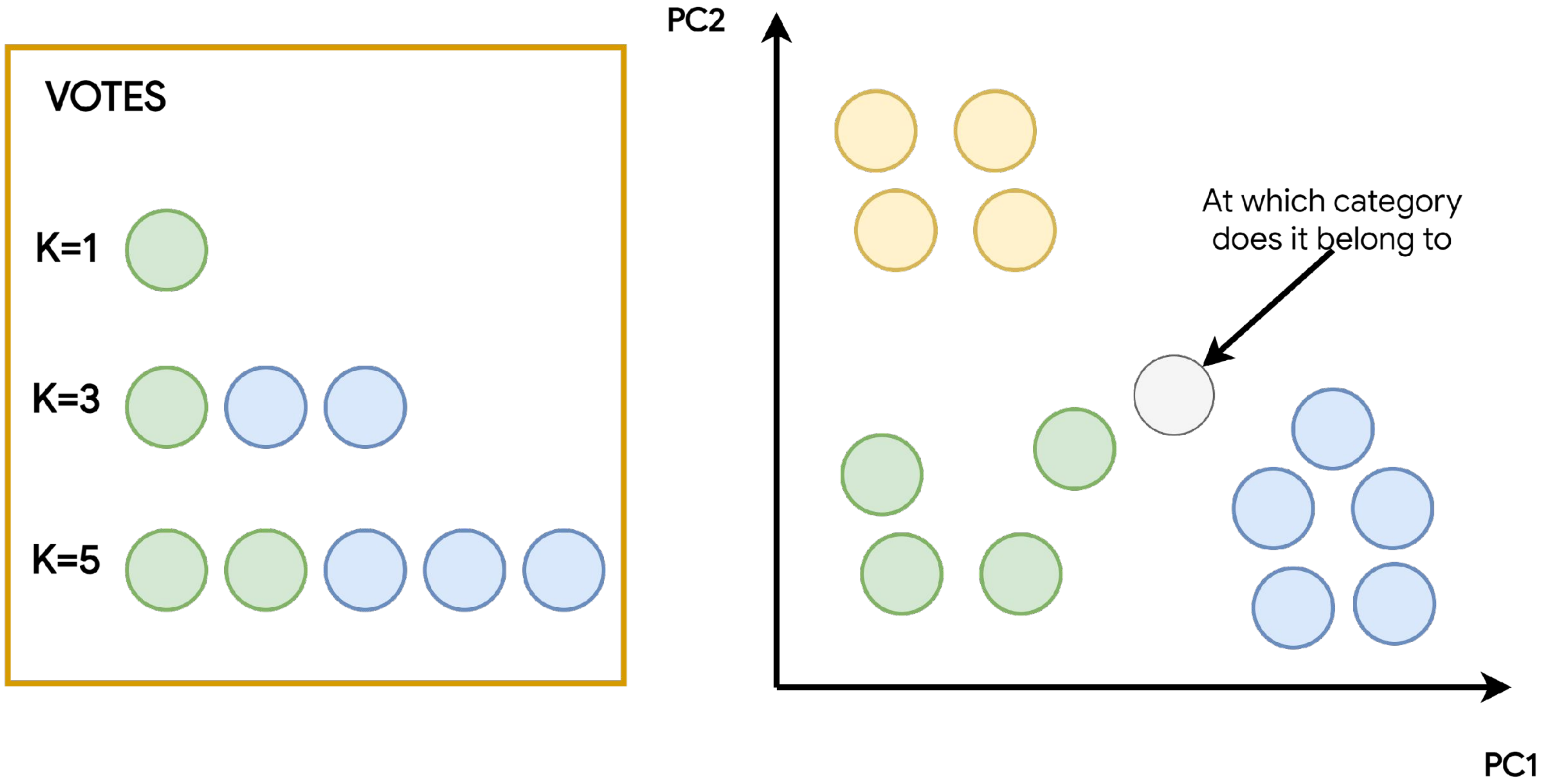


- However, Margin cannot handle outliers. We can use Cross Validation to find the Soft-Margin



Needed For the Exam

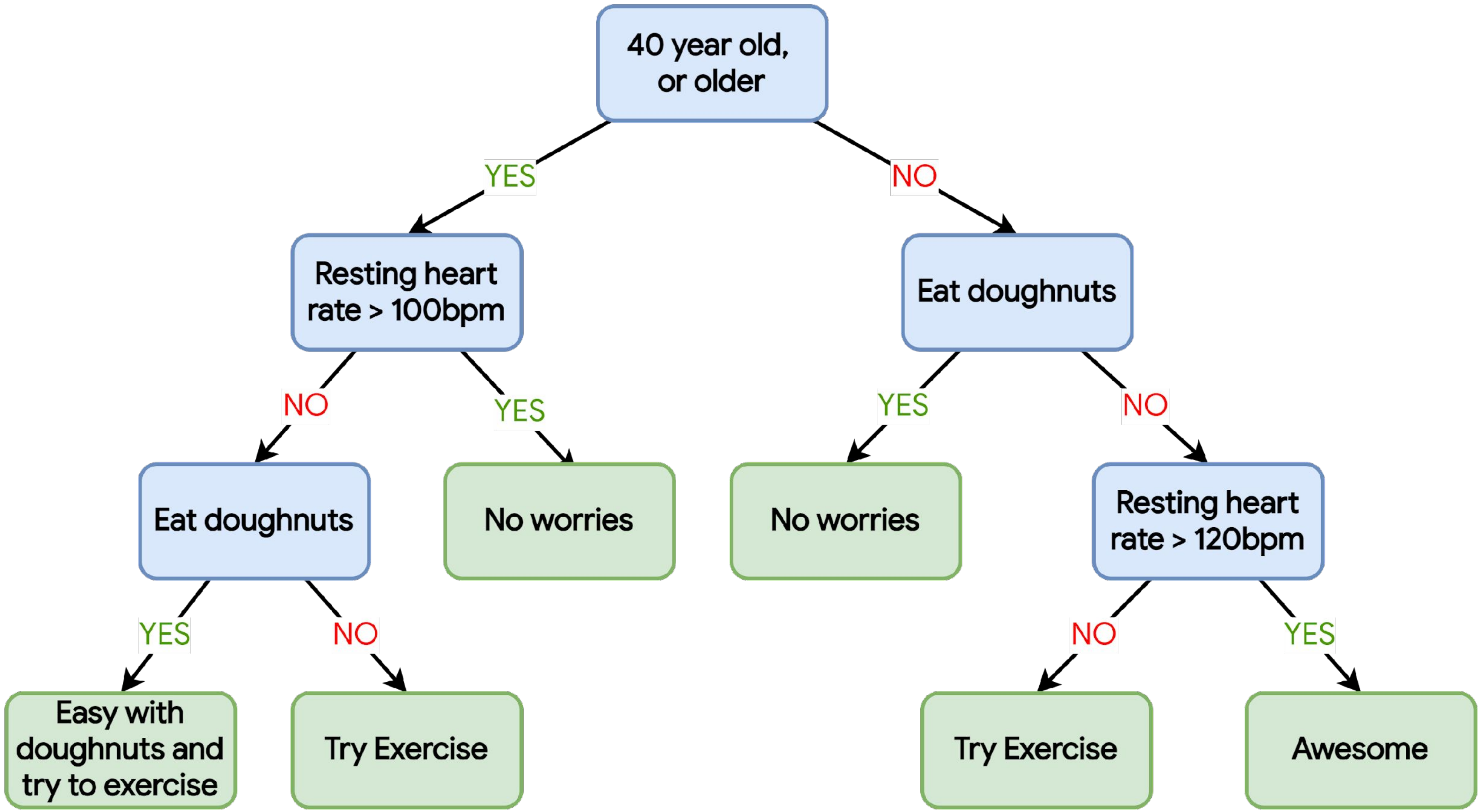
Classify with K-Nearest Neighbors… KNN



- K is an hyperparameter. Use Cross Validation to find out the best value

Needed For the Exam

Decision Tree... Quick introduction

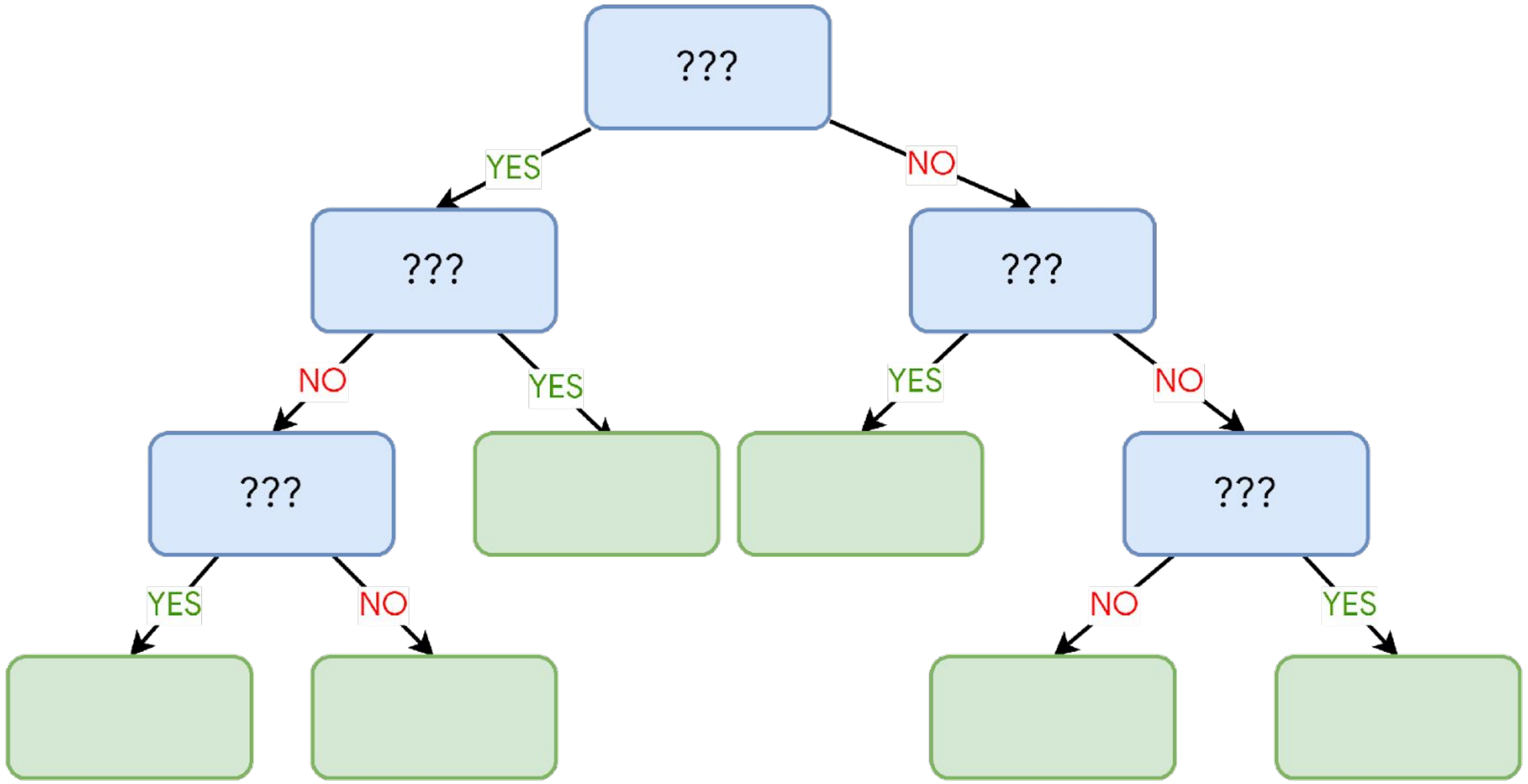


FYI: For your knowledge

Decision Tree… Building process

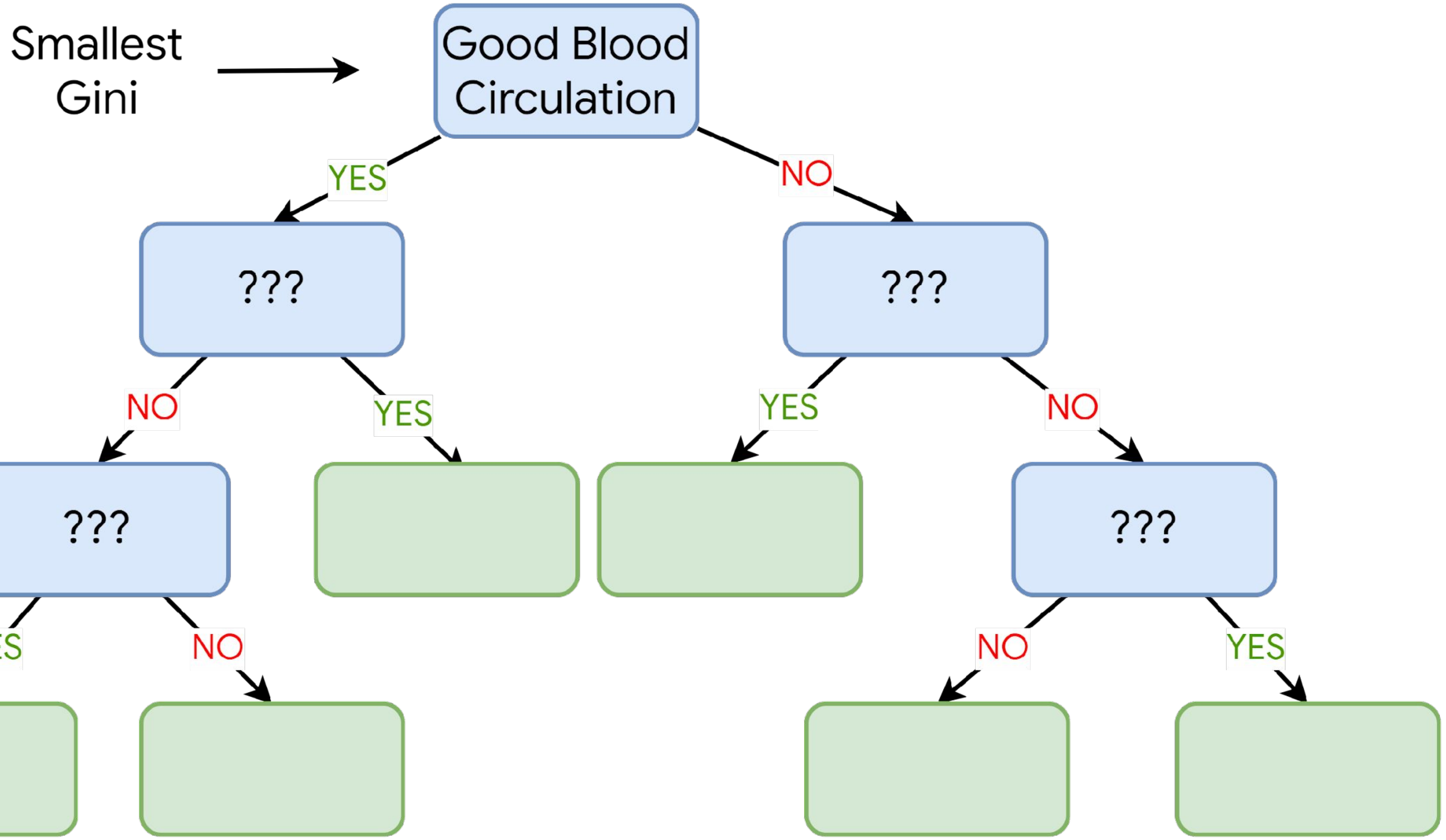
Chest Pain	Good Blood Circulation	Blocker Arteries	Heart Disease
no	no	yes	yes
yes	no	yes	yes
yes	yes	no	no

……..



FYI: For your knowledge

Decision Tree… Choose the feature with less impurity



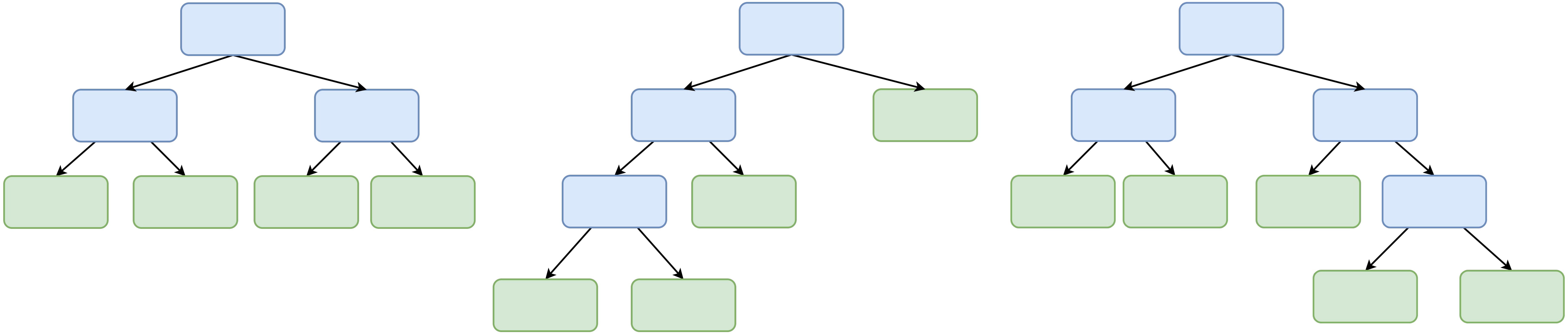
- A feature with less impurity cuts the population the best !
- Proceed this way and stop when the cut returns a higher impurity than before.

Needed For the Exam

Decision Tree are great and simple. However, they are not good in classifying new samples. That's why we use Random Forest

Needed For the Exam

Random Forest… Many Voting Decision Trees



FYI: For your knowledge

Random Forest… 1) Build bootstrap and Out-of-Bag dataset

Original Dataset

Chest Pain	Good Blood Circ	Blocked Arteries	Weight	Heart Disease
NO	NO	NO	125	NO
YES	YES	YES	180	YES
YES	YES	NO	210	NO
YES	NO	YES	167	YES

Out-of-Bag Entry

Bootstrapped Dataset

Chest Pain	Good Blood Circ	Blocked Arteries	Weight	Heart Disease
NO	NO	NO	125	NO
YES	YES	YES	180	YES
YES	NO	YES	167	NO
YES	NO	YES	167	YES

Duplicated Entries
are allowed

FYI: For your knowledge

Random Forest… 2) Build a decision tree

- Use a subset of variables at each step
- Select a random variables from the subset
- Build a decision tree
- Repeat. Make a new bootstrapped DS…

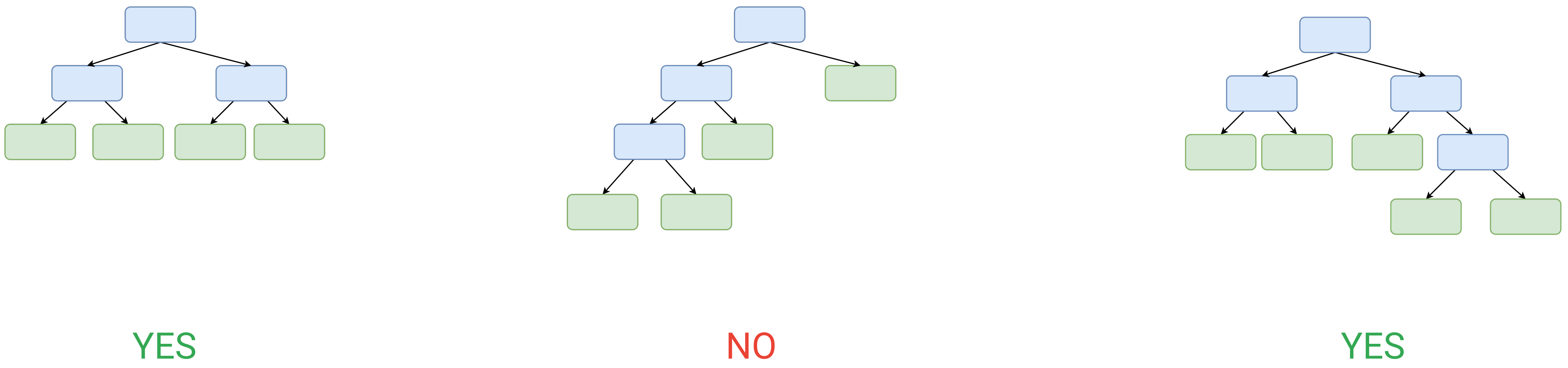
Bootstrapped Dataset

Chest Pain	Good Blood Circ	Blocked Arteries	Weight	Heart Disease
NO	NO	NO	125	NO
YES	YES	YES	180	YES
YES	NO	YES	167	NO
YES	NO	YES	167	YES

FYI: For your knowledge

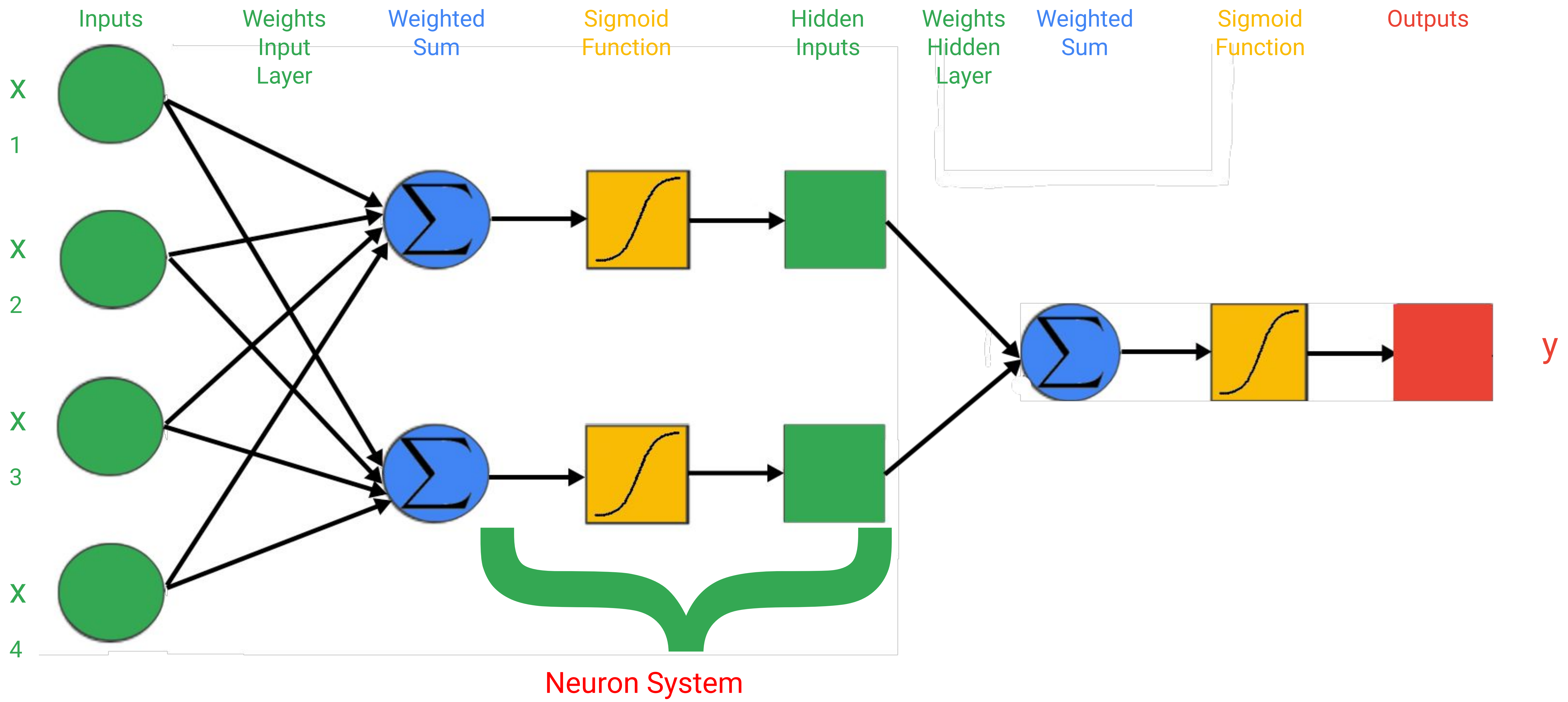
Random Forest… 3) Vote

Chest Pain	Good Blood	Blocked Arteries	Weight	Heart Disease
YES	NO	NO	173	?



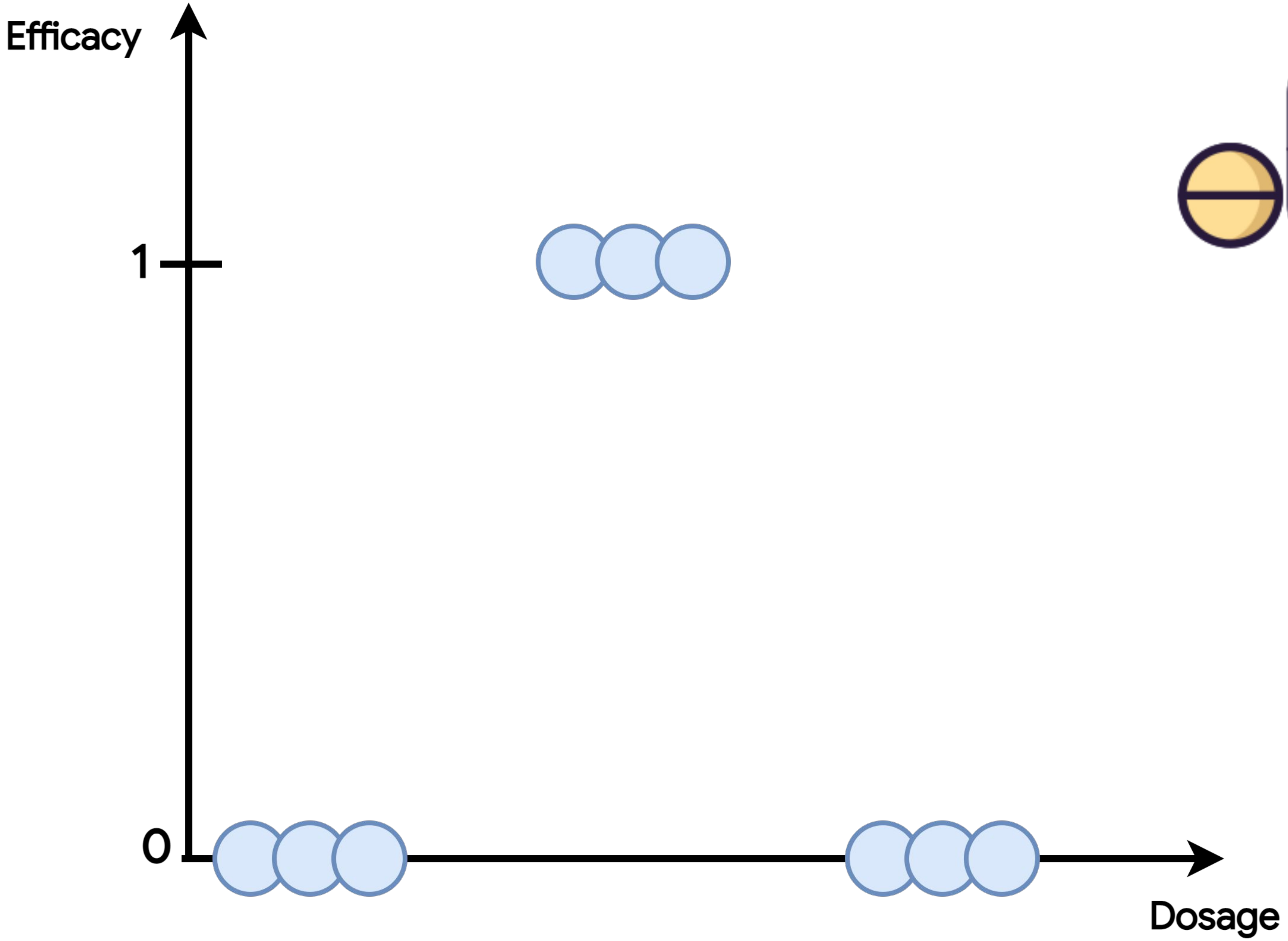
FYI: For your knowledge

Neural networks: Multi-layer perceptron



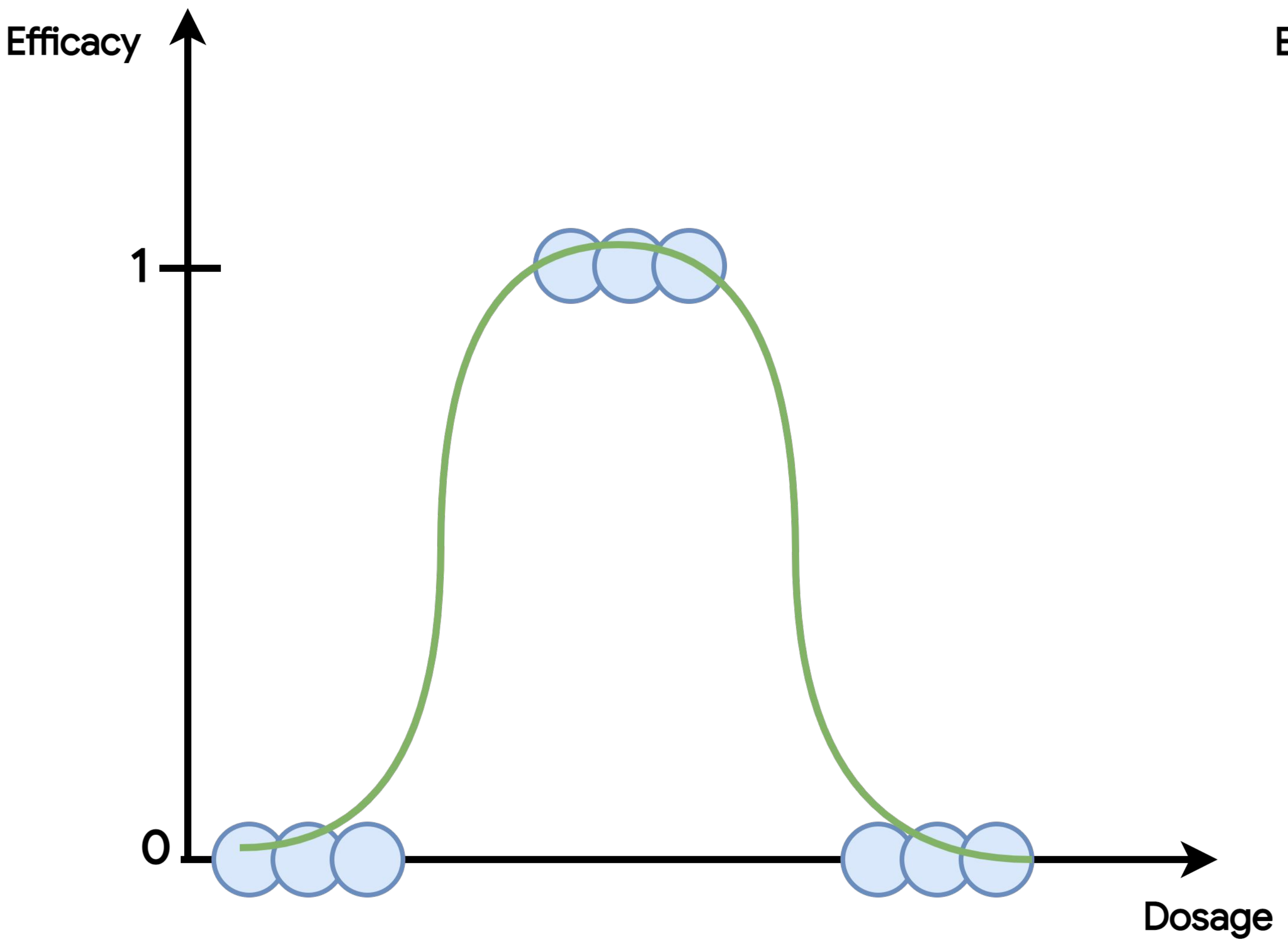
FYI: For your knowledge

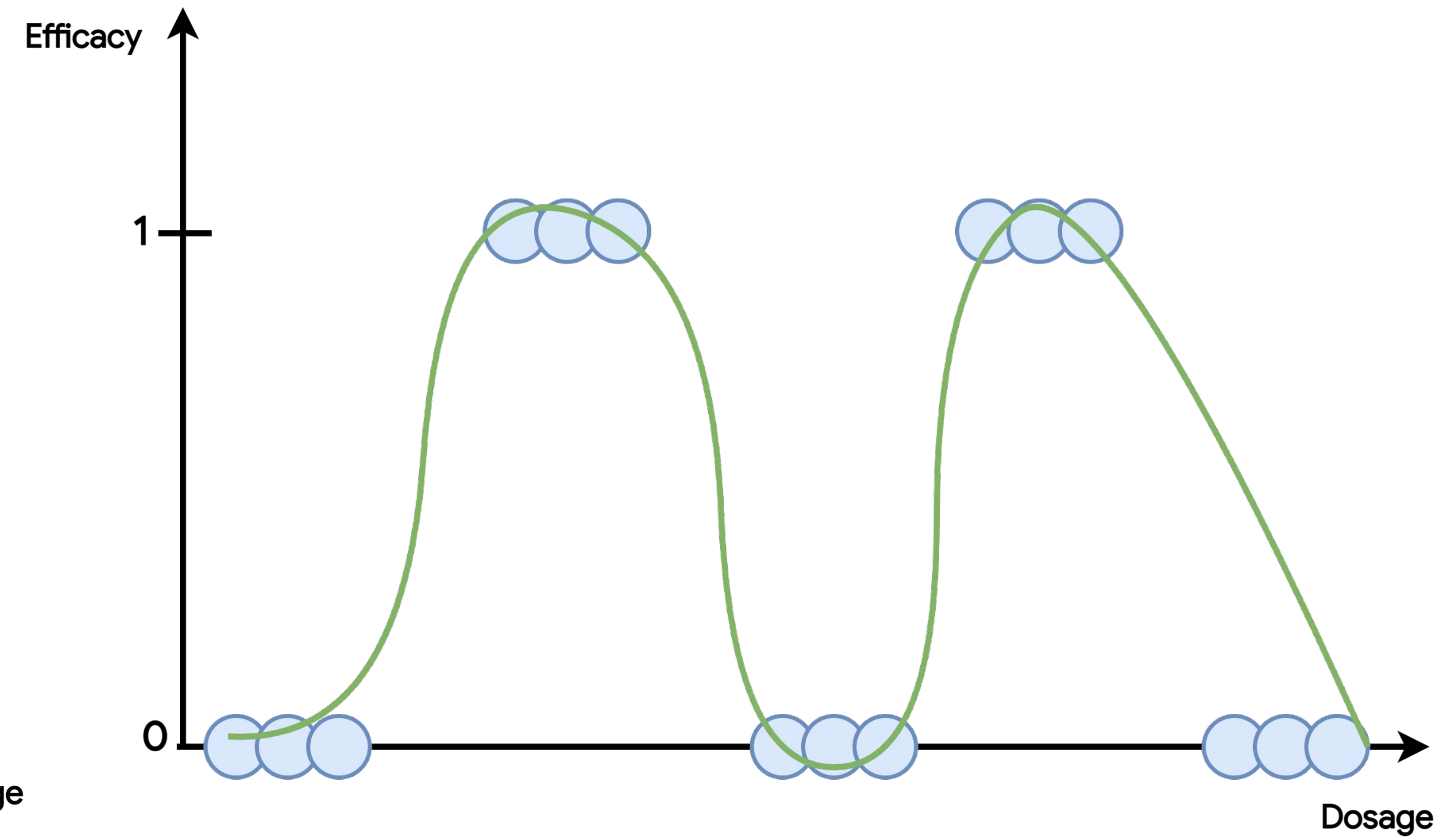
Neural networks… why do we need them?



FYI: For your knowledge

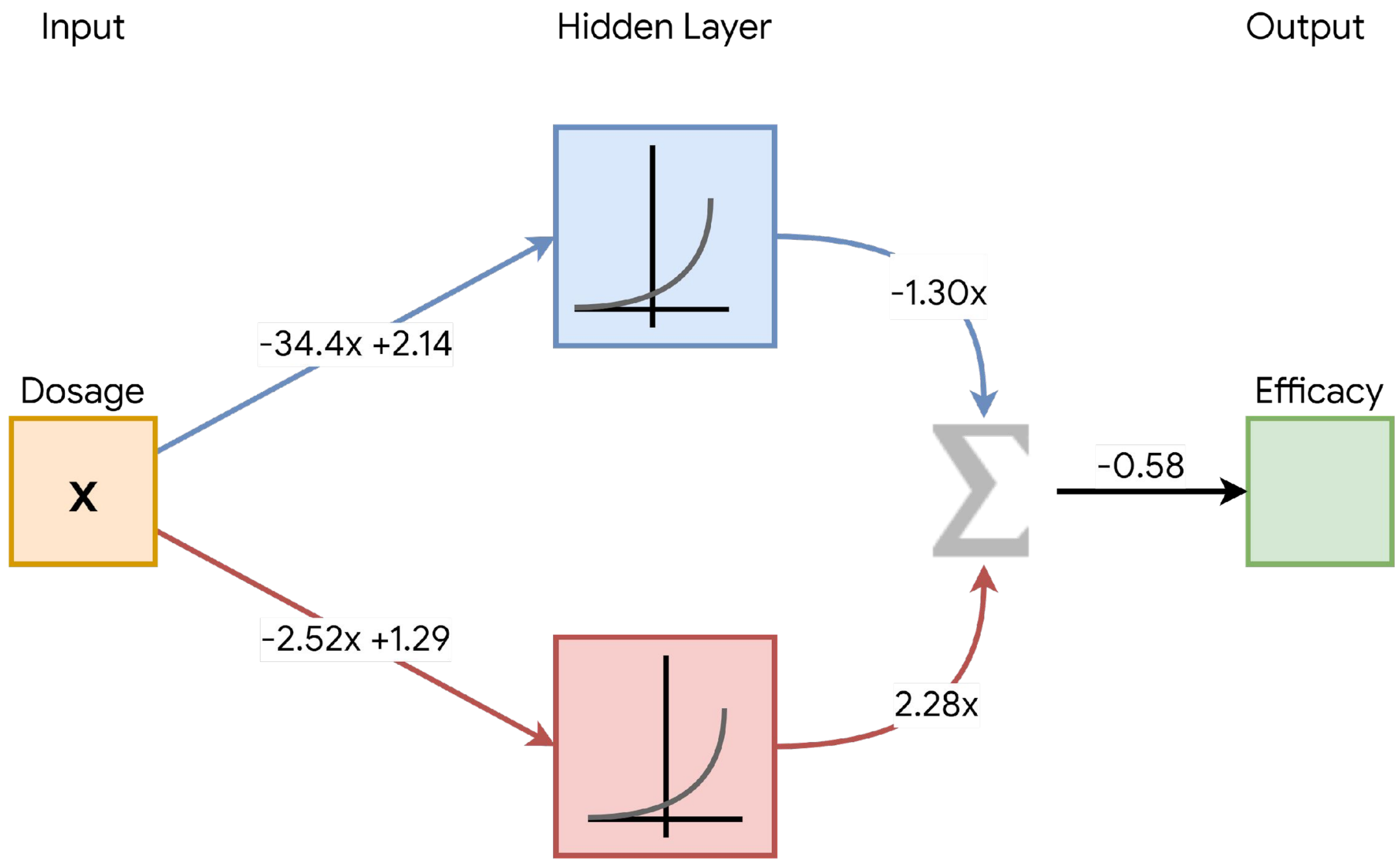
Neural networks… can solve non-linear problem





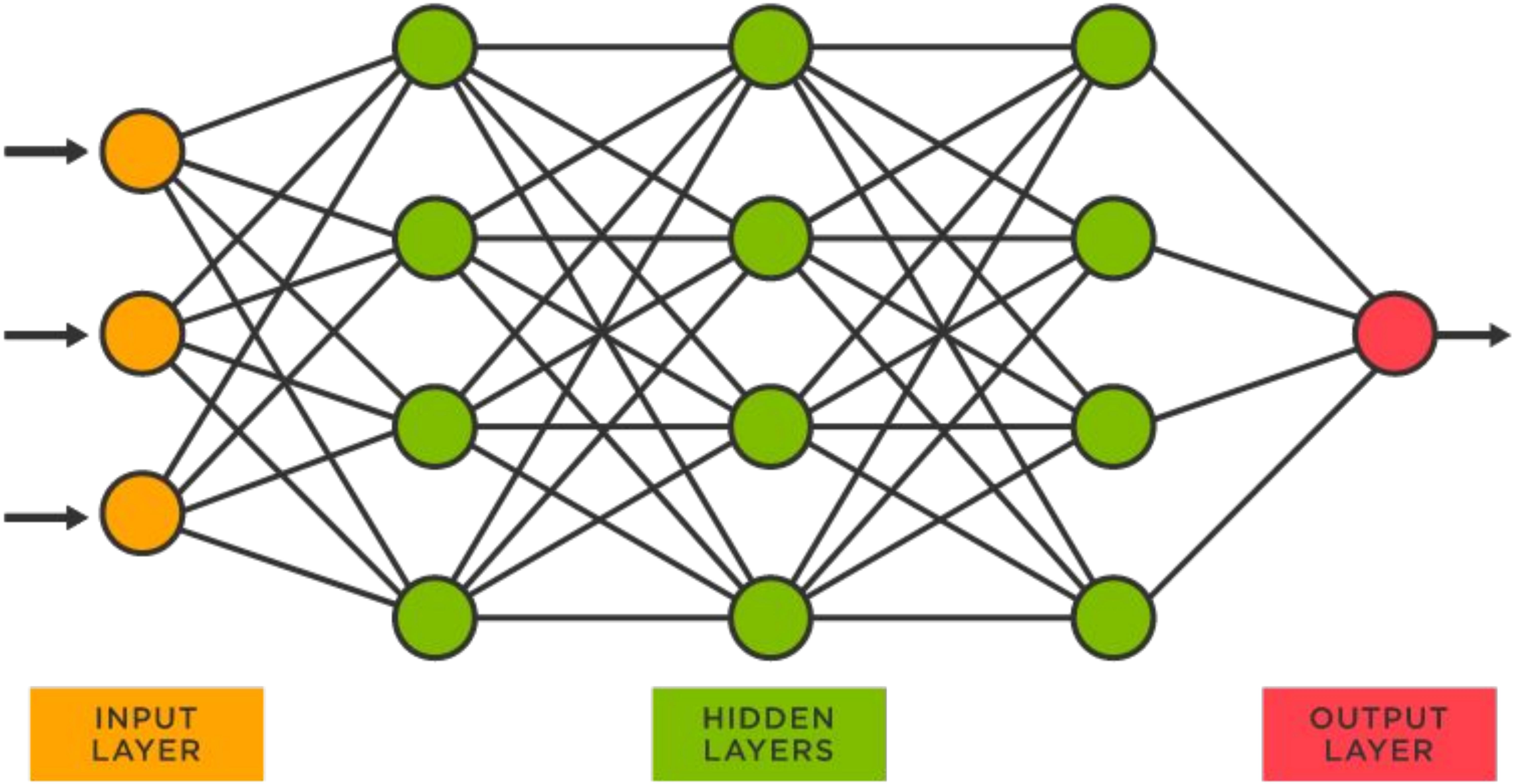
FYI: For your knowledge

Neural networks… how it works



Needed For the Exam

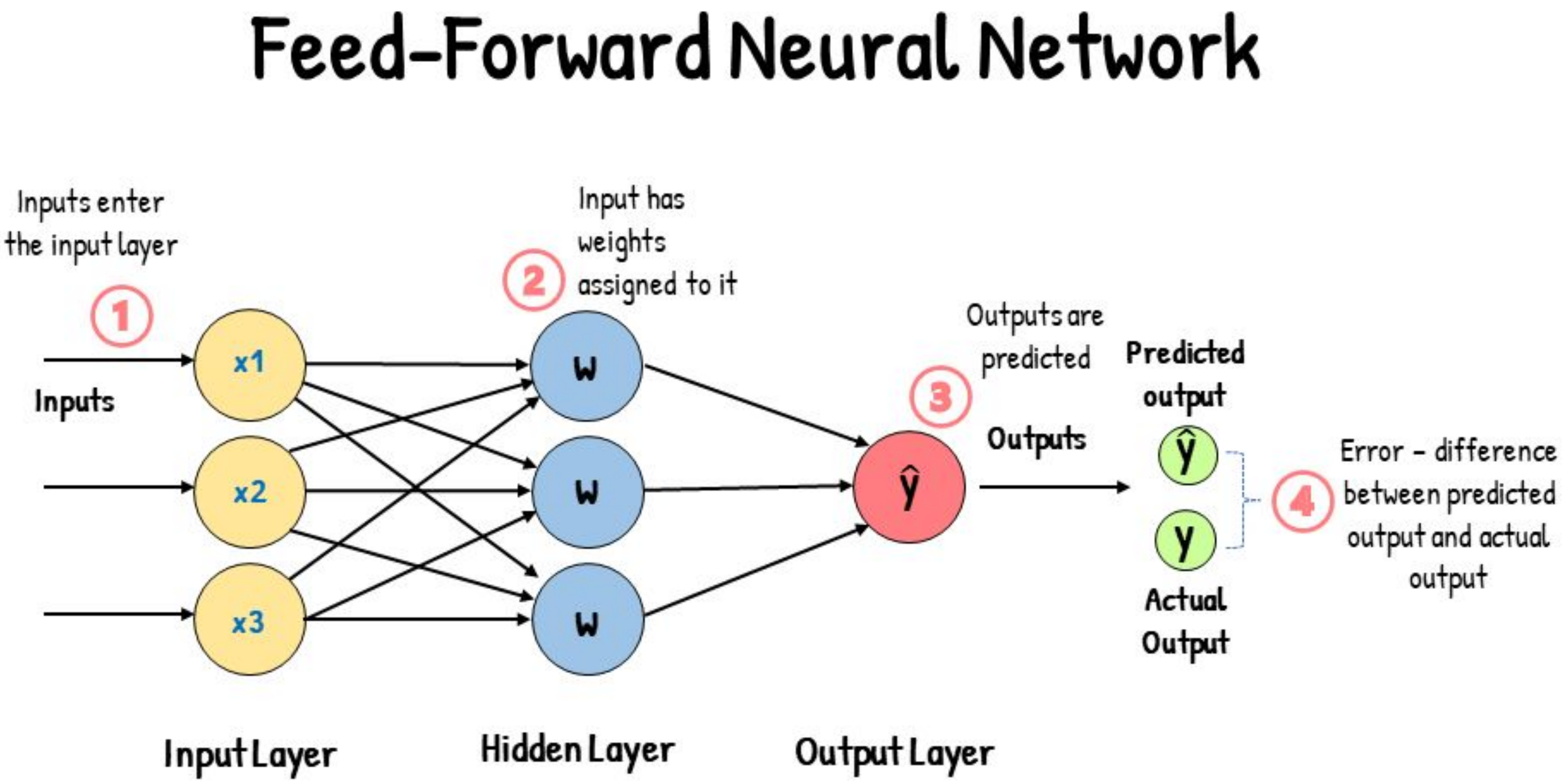
Neural networks… more layers, more computation



BACKPROPAGATION

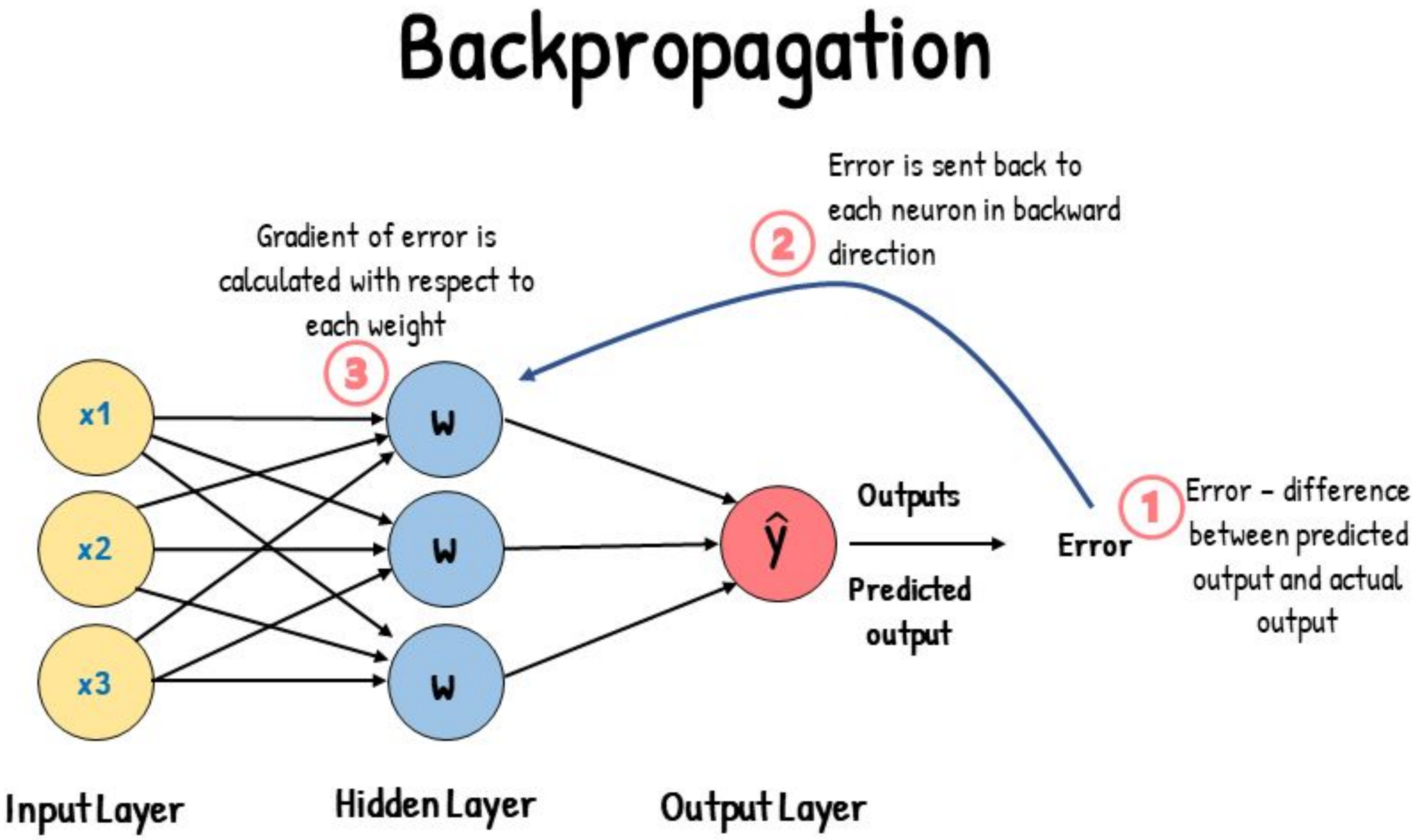
Needed For the Exam

Neural networks… more layers, more computation



Needed For the Exam

Neural networks… more layers, more computation

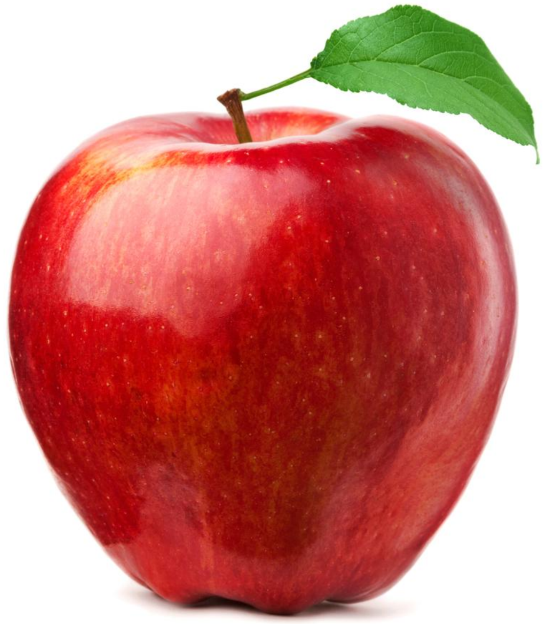


Needed For the Exam

Neural networks… Backpropagation overview

- Backpropagation is an iterative algorithm to optimize all parameters (weights and biases) of a neural networks
- Backpropagation uses Gradient Descend for optimization
- Backpropagation happens during training of neural network

Machine Learning Algorithm… Summary



If you need to eat an apple
Which one would you go for?

Section Agenda

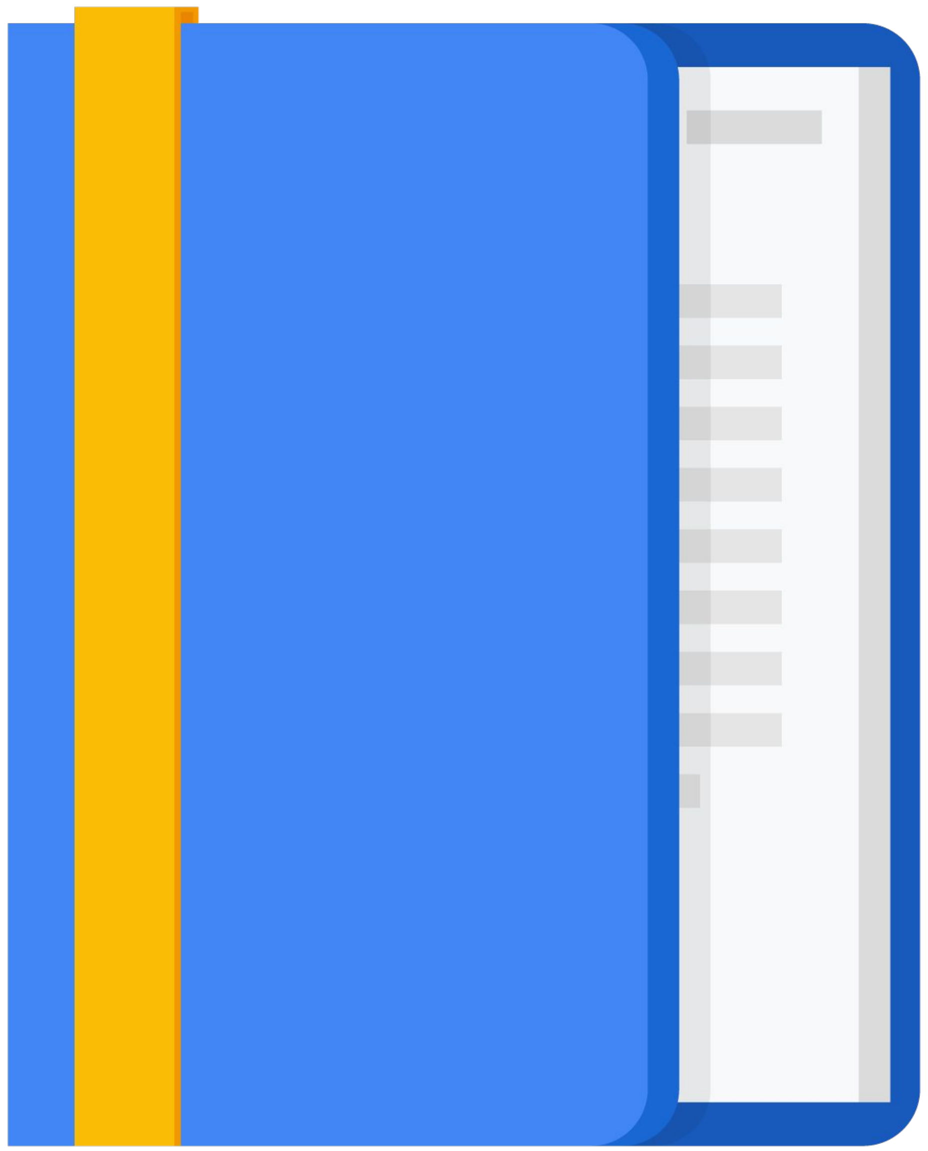
Introduction to ML Algorithms

Unsupervised Learning Algorithms

Supervised Learning Algorithms

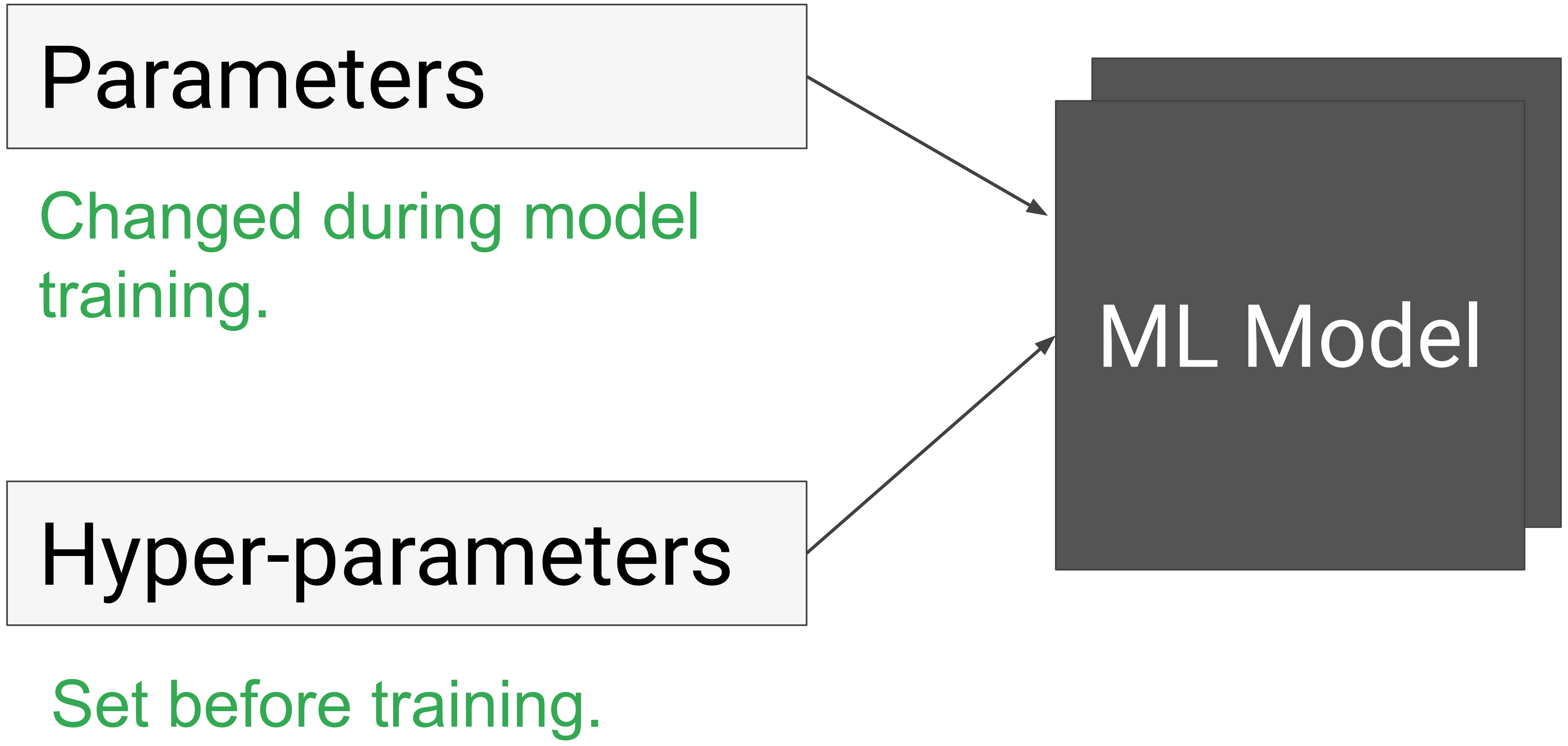
Loss functions

Gradient Descent



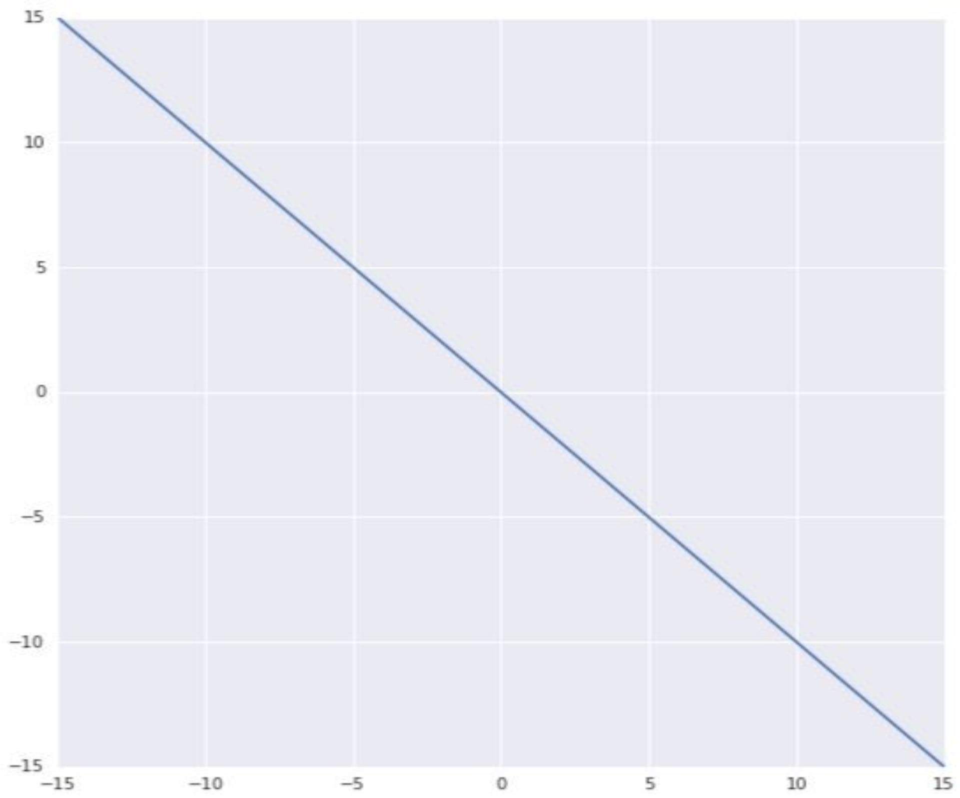
Needed For the Exam

ML models are mathematical functions with parameters and hyper-parameters



FYI: For your knowledge

Linear models have two types of parameters: Bias and weight



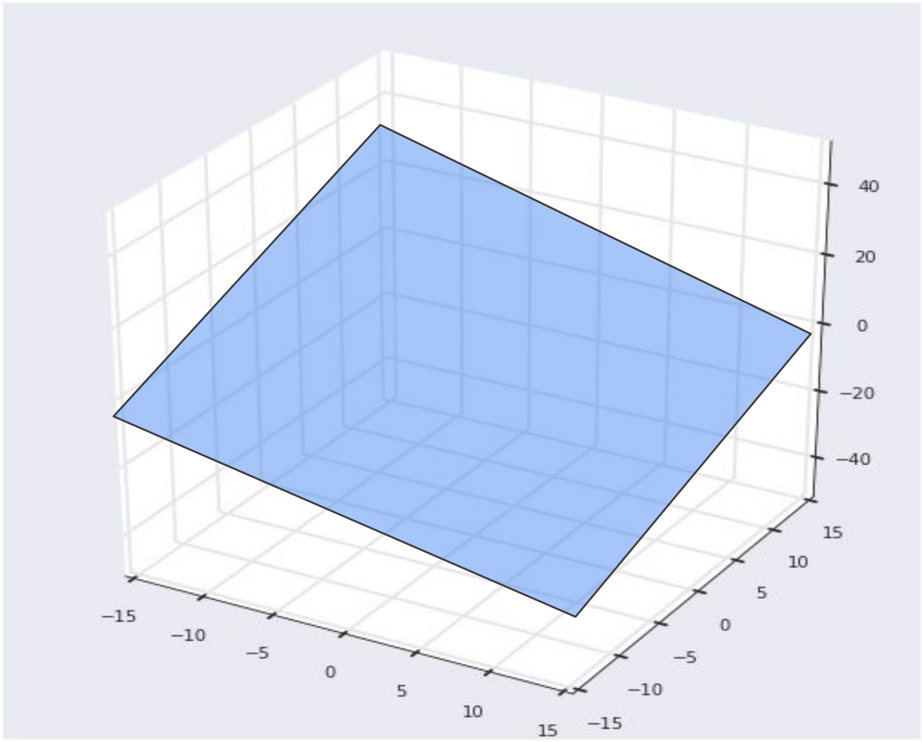
Linear

Output | Bias Term | Input | Weight

$$y = \boxed{b} + x \times \boxed{m}$$

$$y = \boxed{b} + X \times \boxed{w}$$

Model Parameters



Hyperplane

FYI: For your knowledge

Equation for a linear model tying mother's age and baby weight

The slope of the line is given by w1.

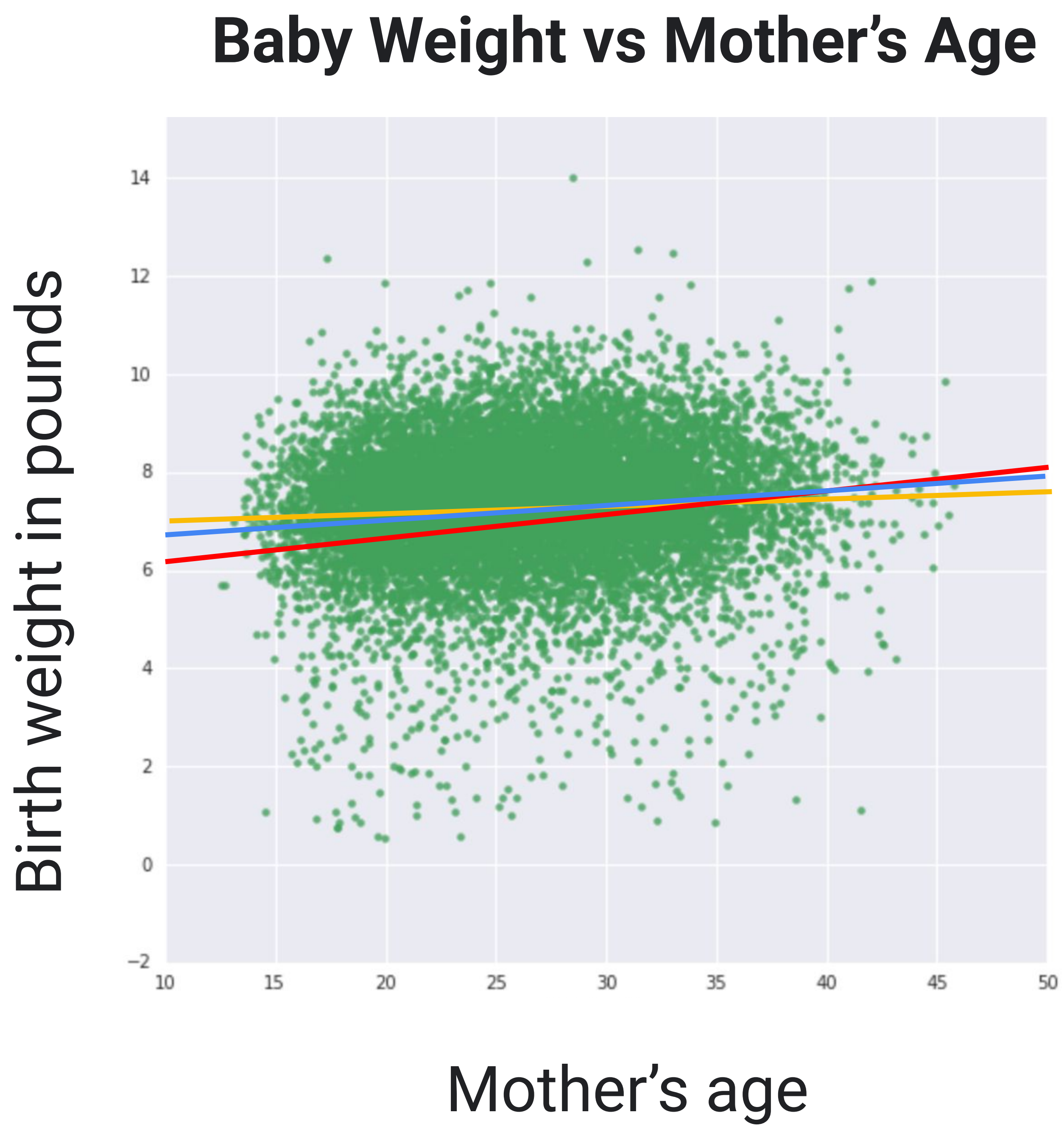
$$y = w_1 x_1 + b$$

- x_1 is the **feature** (e.g. mother's age)
- w_1 is the **weight** for x_1

Line: y= .02x + 6.83

Line: y= .03x + 6.49

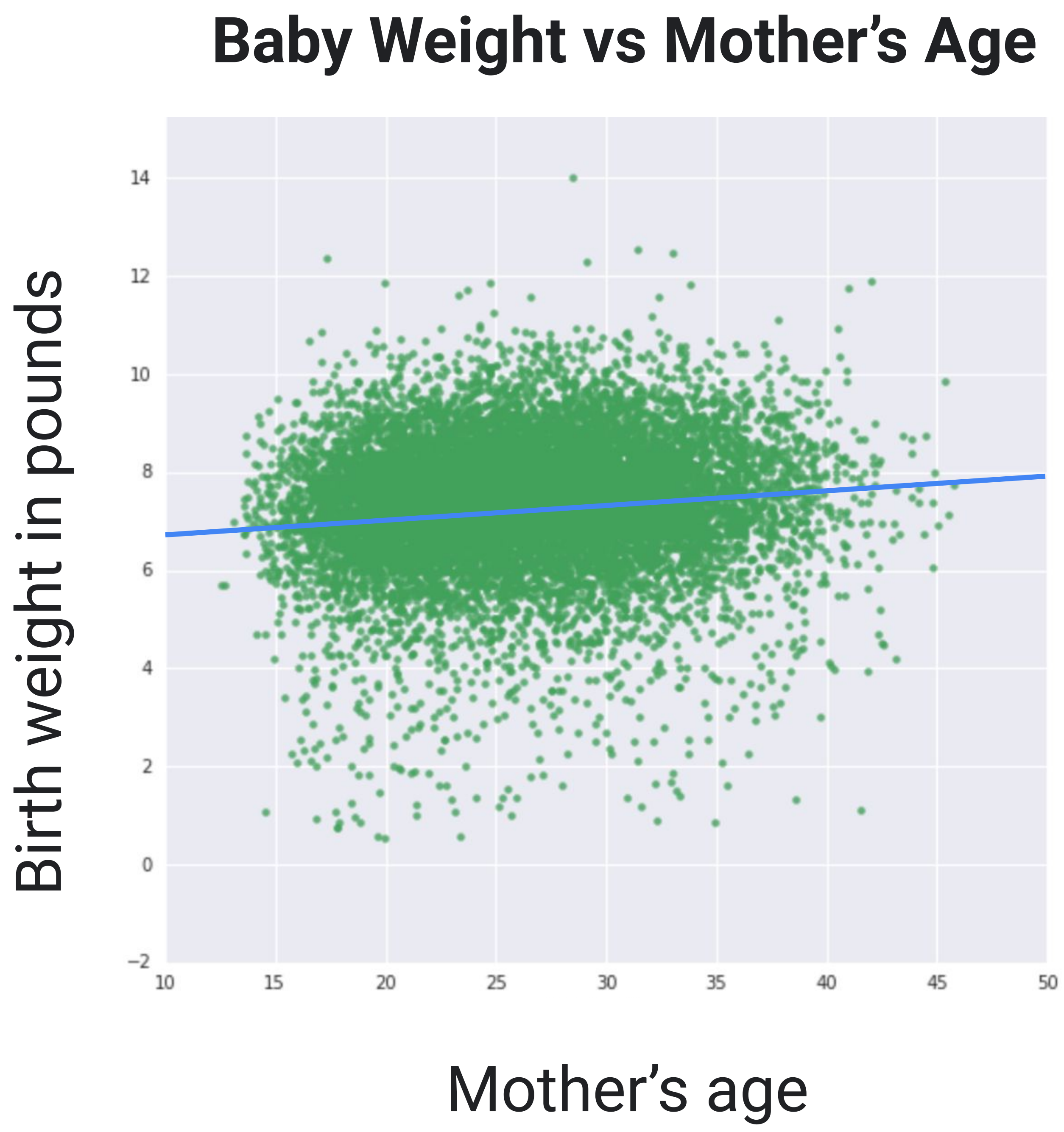
Line: y= .01x + 7.14



FYI: For your knowledge

Can’t we just solve the equation using all the data?

When an analytical solution is no longer an option, you use gradient descent.



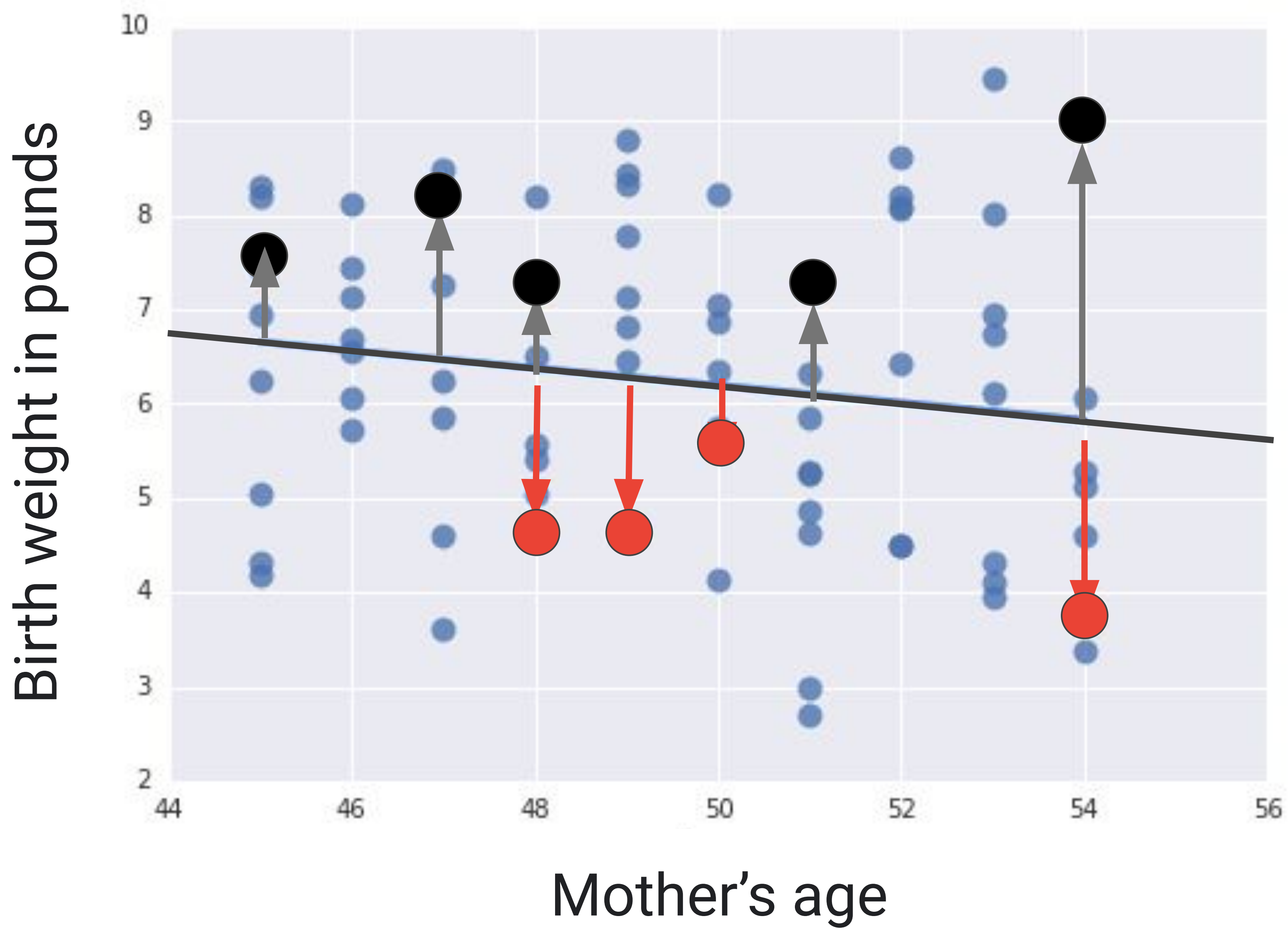
FYI: For your knowledge

Compose a loss function by calculating errors

Error = actual (true) - predicted value
Compute the errors:

```
+0.70
+1.10
+0.65
-1.20
-1.15
+1.10
+3.09
-2.10
```

Each error makes sense. How about all the errors added together?



Needed For the Exam

One loss function metric is Root Mean Squared Error (RMSE)

1 Get the errors for the training examples.

```
+0.70
+1.10
+0.65
-1.20
-1.15
+1.10
+3.09
-2.10
```

2 Compute the squares of the error values.

```
0.49
1.21
0.42
1.44
1.32
1.21
9.55
4.41
```

3 Compute the mean of the squared error values.

```
2.51
```

$$\sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2}$$

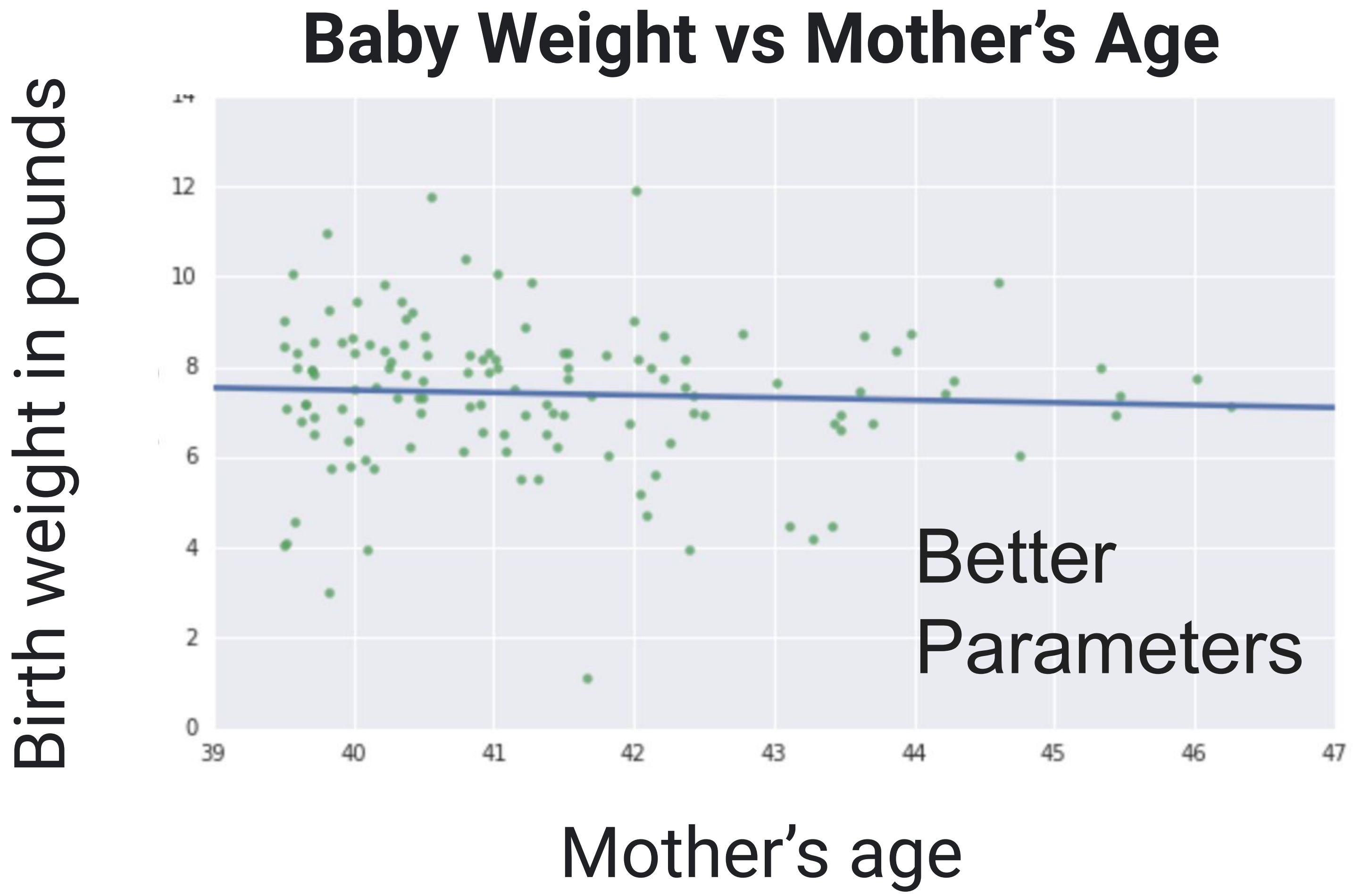
$\hat{Y}_i$ predicted value

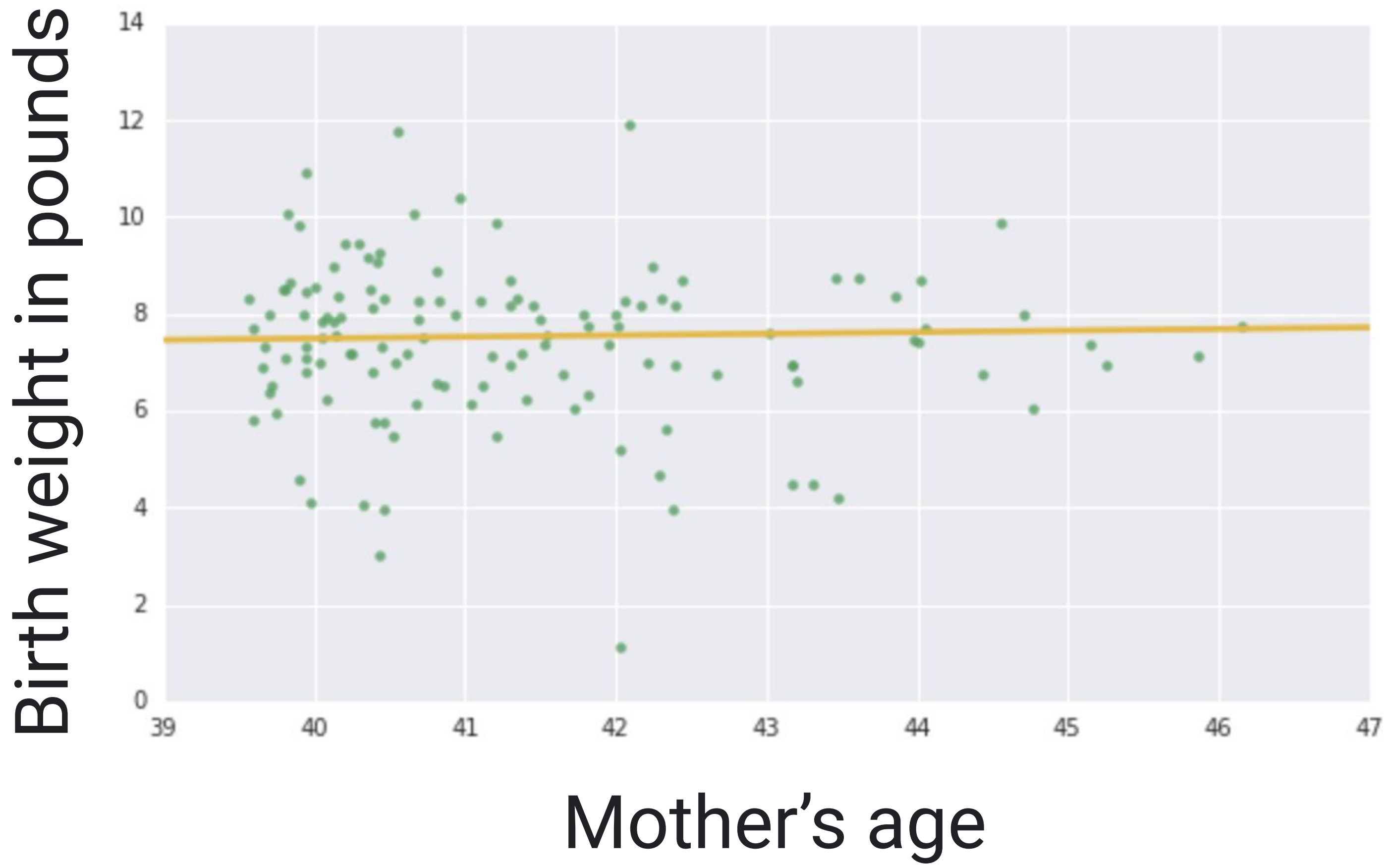
Y_i labeled value

4 Take a square root of the mean. **1.58**

Needed For the Exam

Lower RMSE indicates a better performing model





RMSE=.145

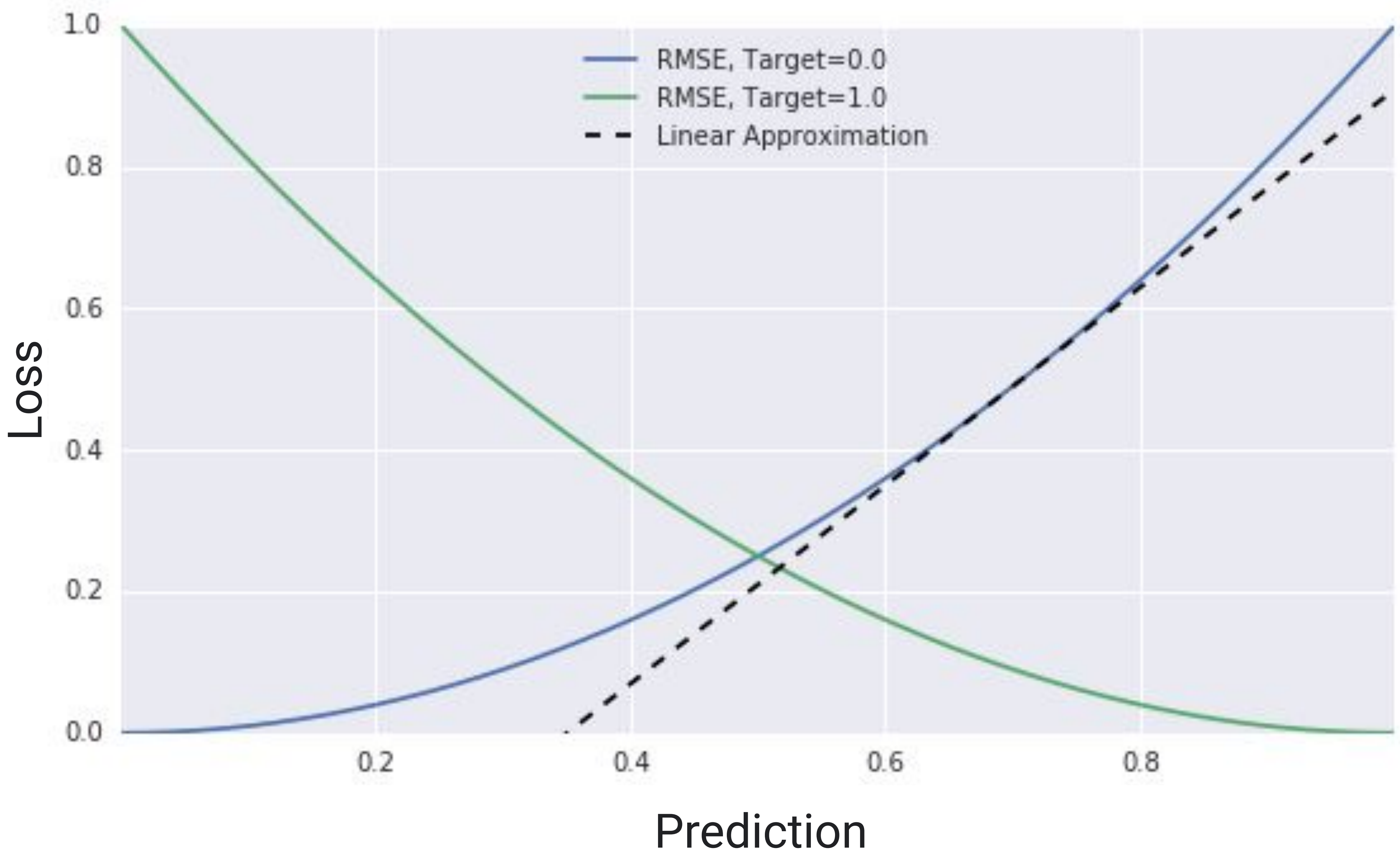
RMSE=.149

Need a way to find the best values for weight and bias.

FYI: For your knowledge

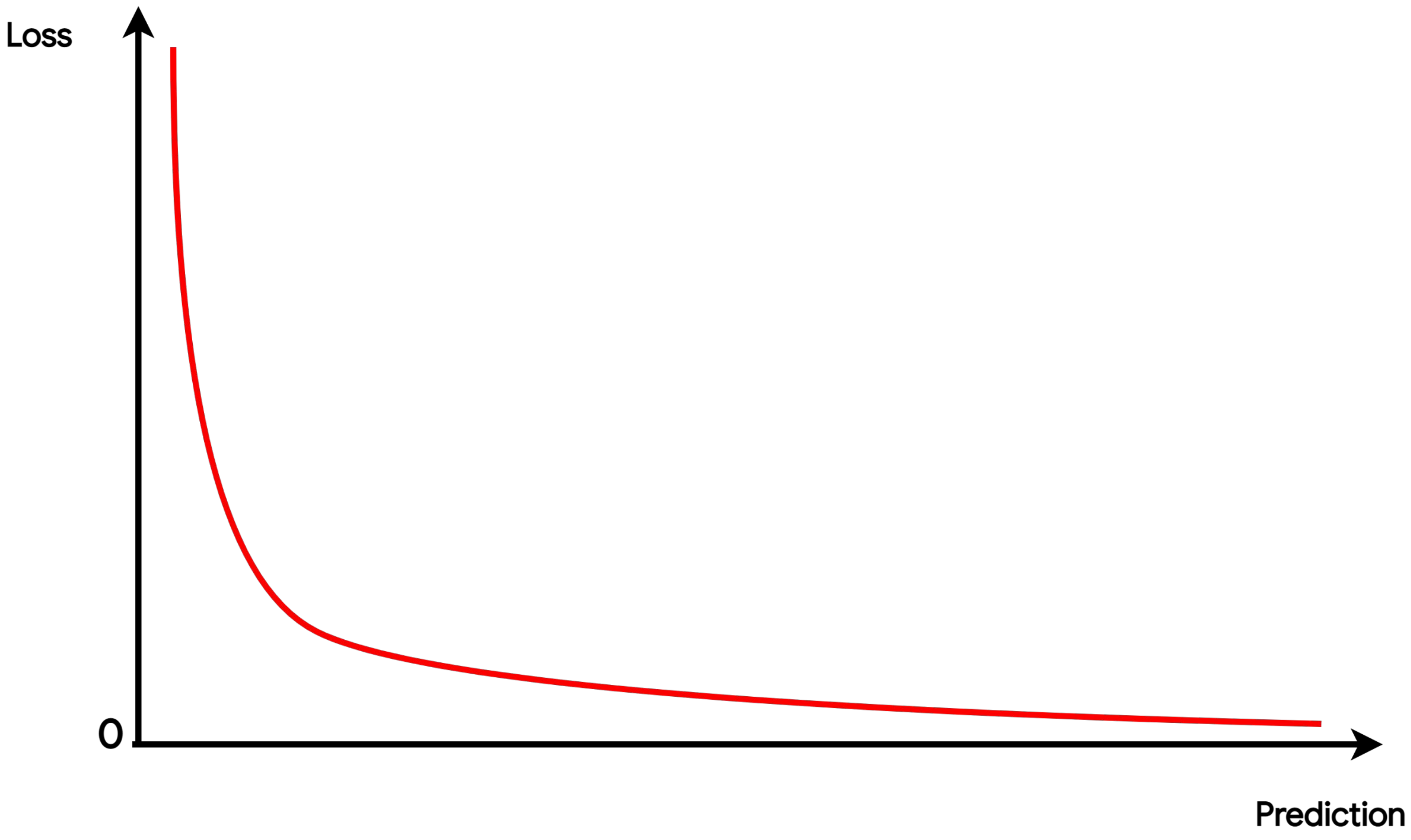
Problem: RMSE doesn't work as well for classification

RMSE doesn't penalize bad classifications appropriately.



Needed For the Exam

Solution: Cross Entropy does



FYI: For your knowledge

Section Agenda

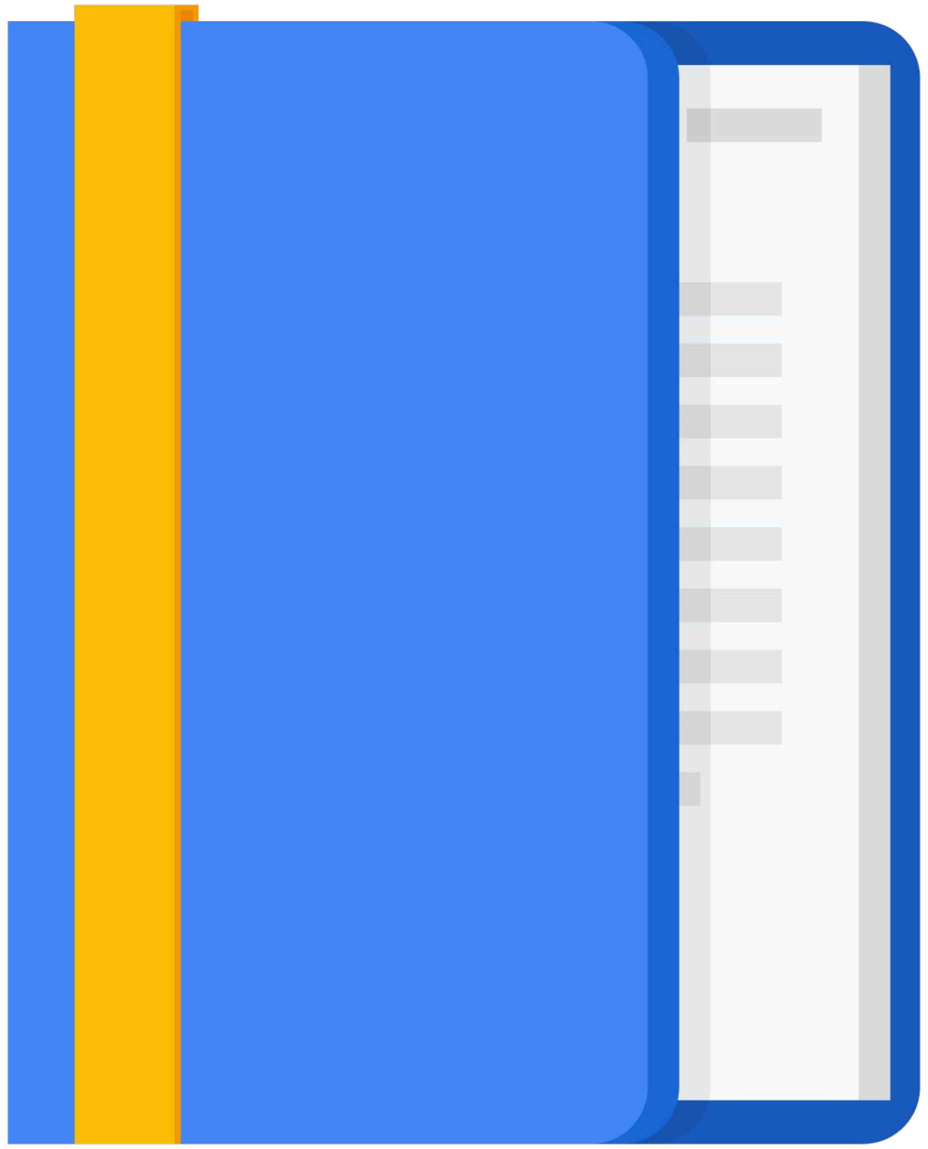
Introduction to ML Algorithms

Unsupervised Learning Algorithms

Supervised Learning Algorithms

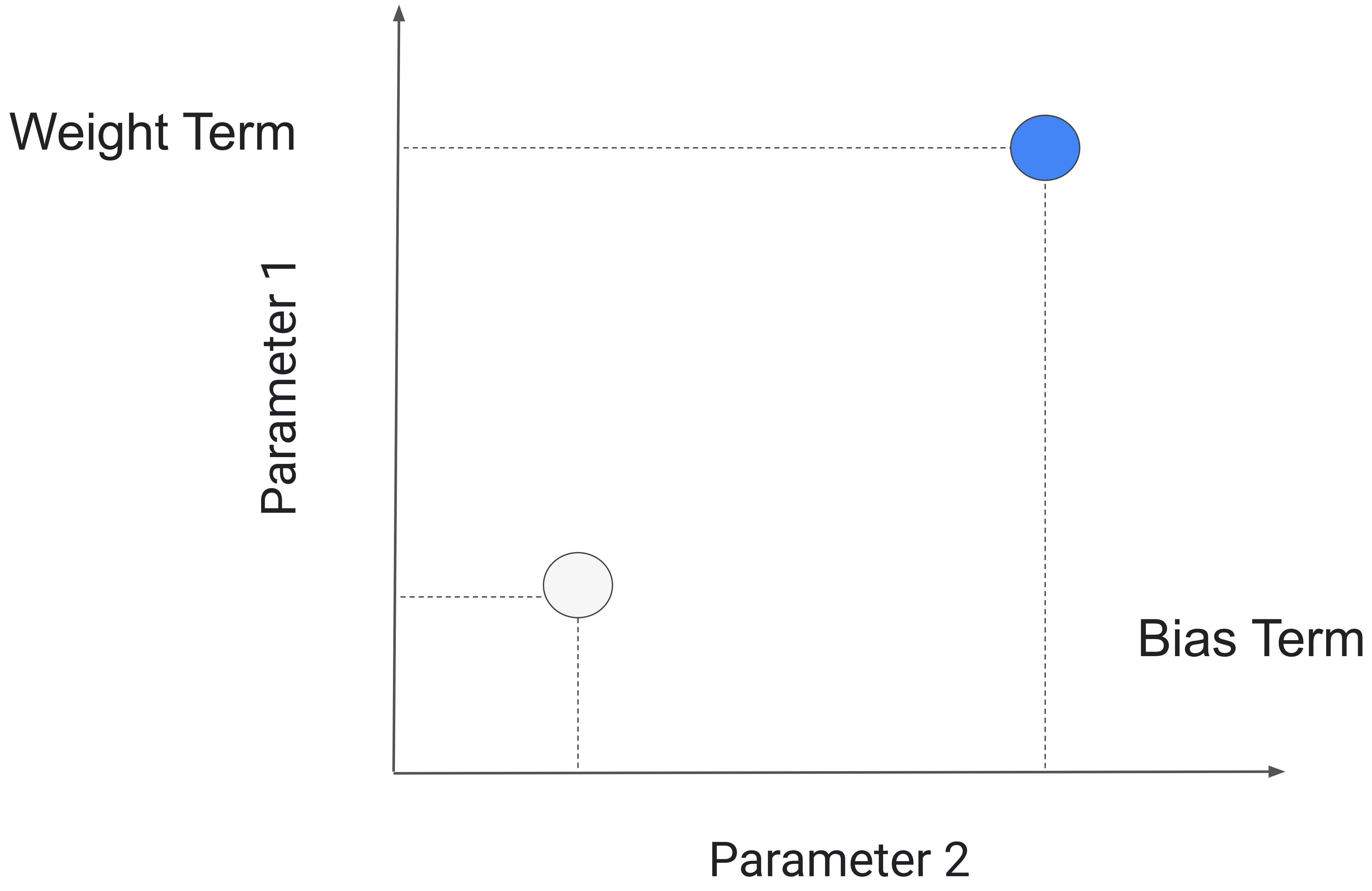
Loss functions

Gradient Descent

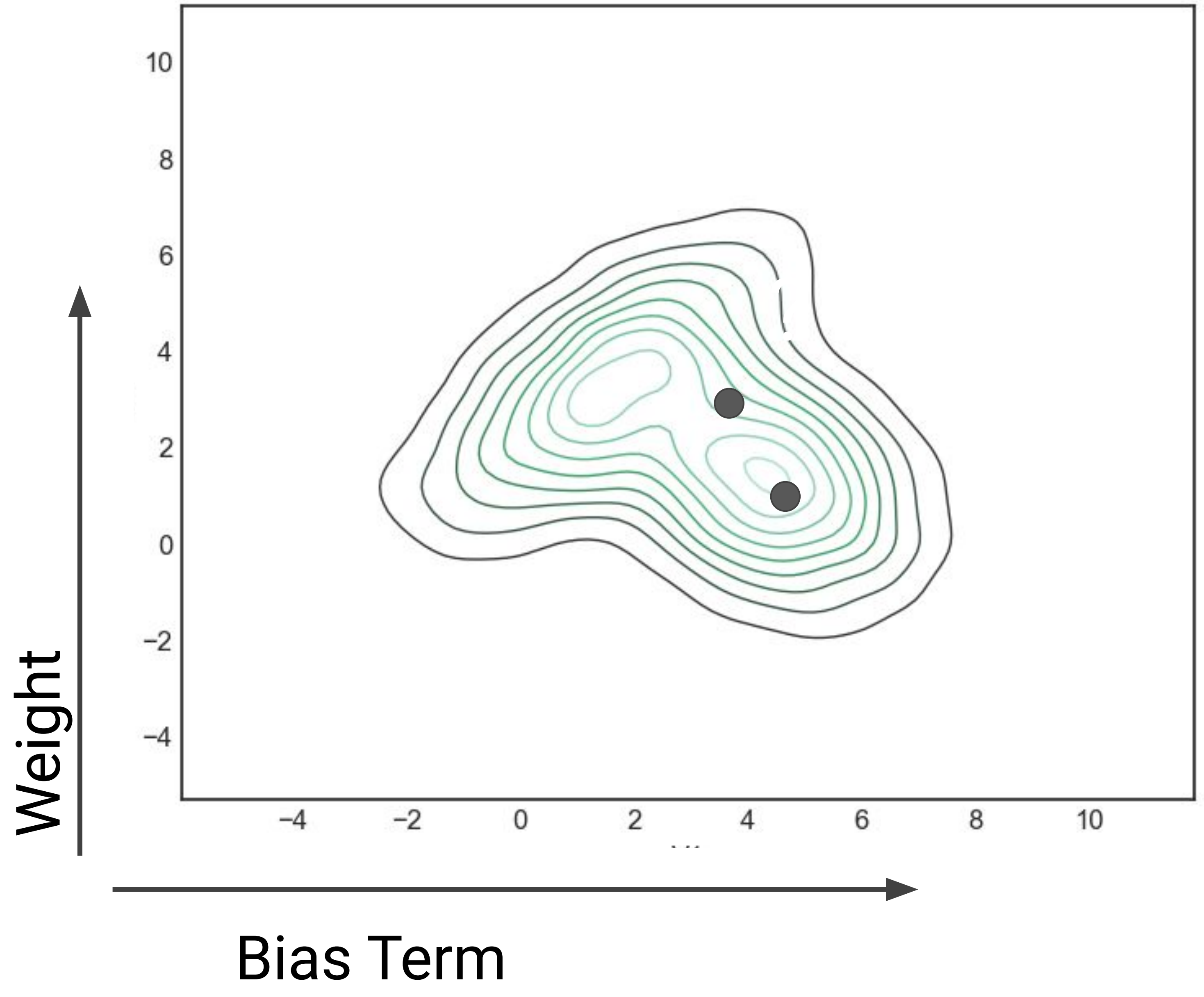


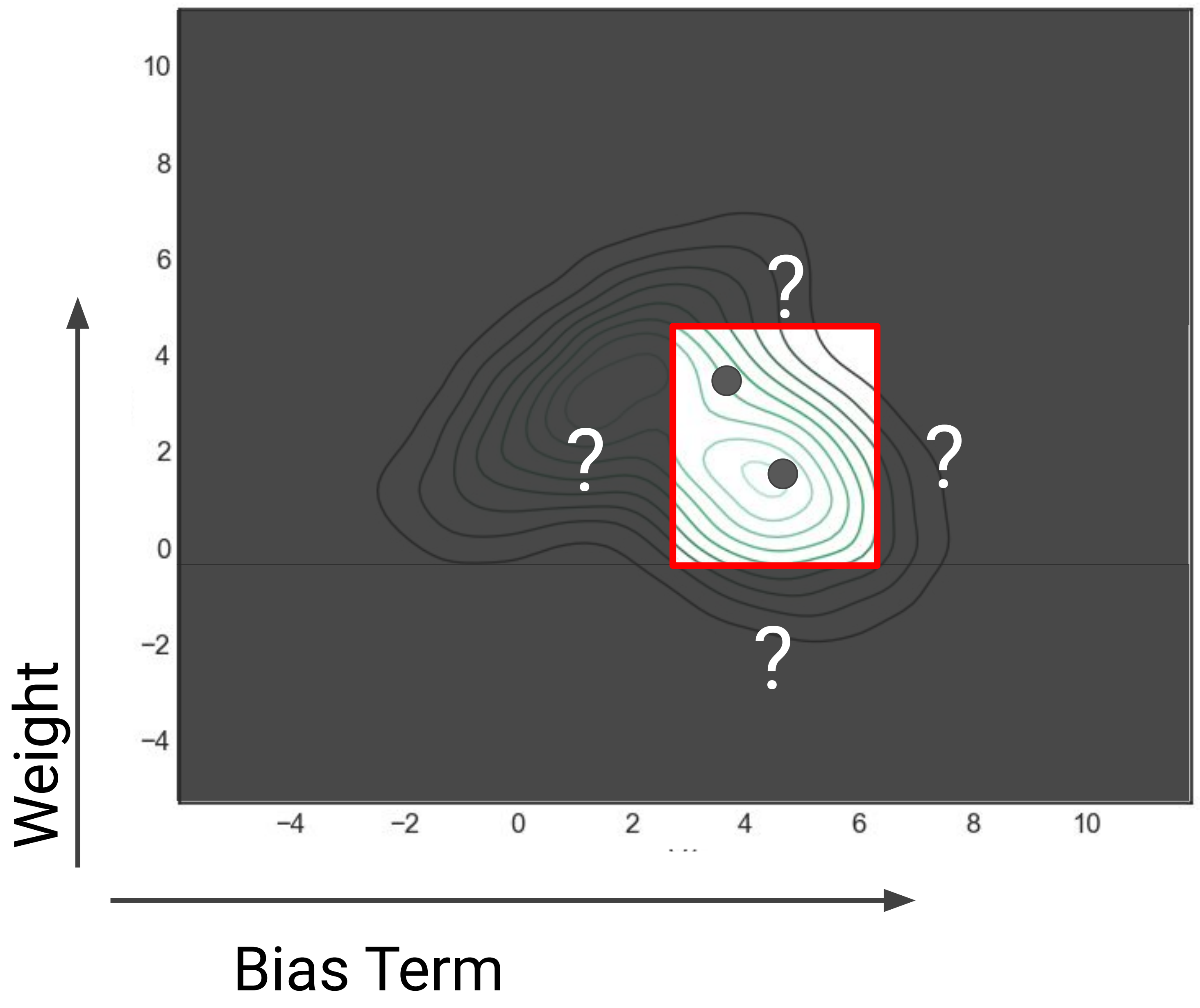
FYI: For your knowledge

Searching in parameter-space



Loss functions lead to loss surfaces





FYI: For your knowledge

Finding the bottom

Which direction should I head?

How large or small a step?

FYI: For your knowledge

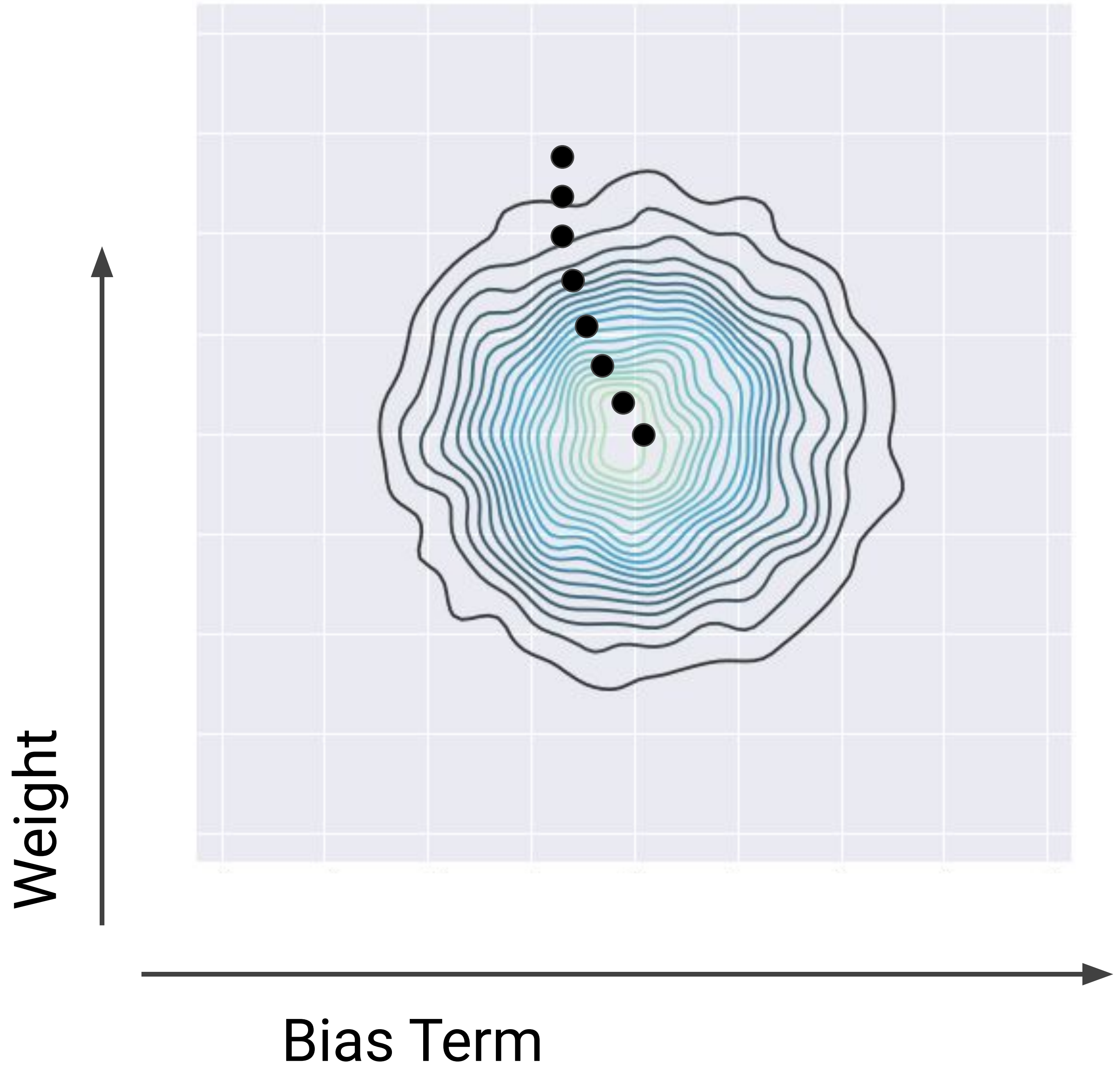
A simple algorithm to find the minimum

```
while loss is > Epsilon:
  direction = computeDirection()
  for i in range(self.params):
    self.params[i] = //
            self.params[i] //
                + stepSize * direction[i]
  loss = computeLoss()
```

Epsilon = A tiny Constant

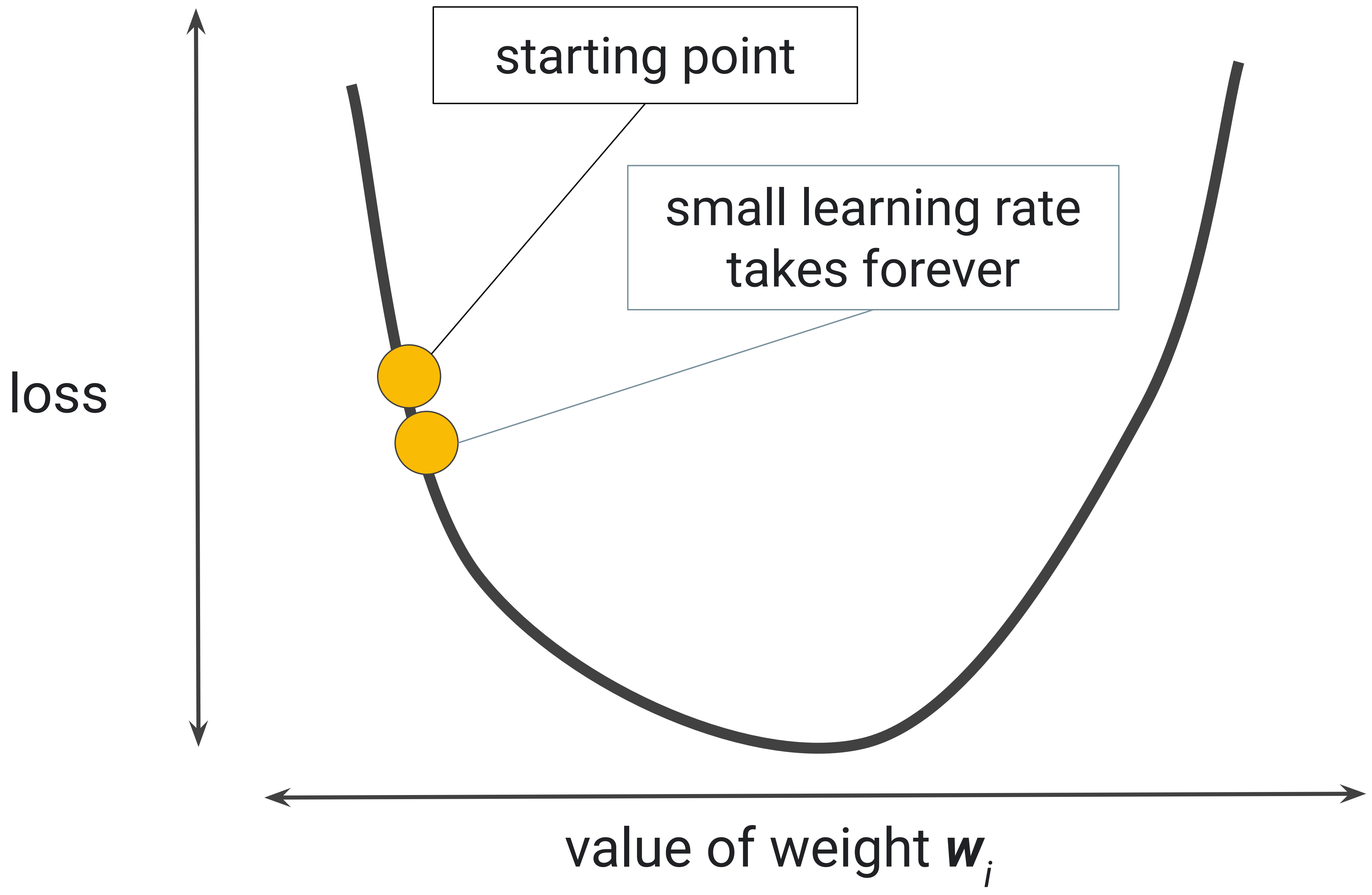
FYI: For your knowledge

Search for a minima by descending the gradient



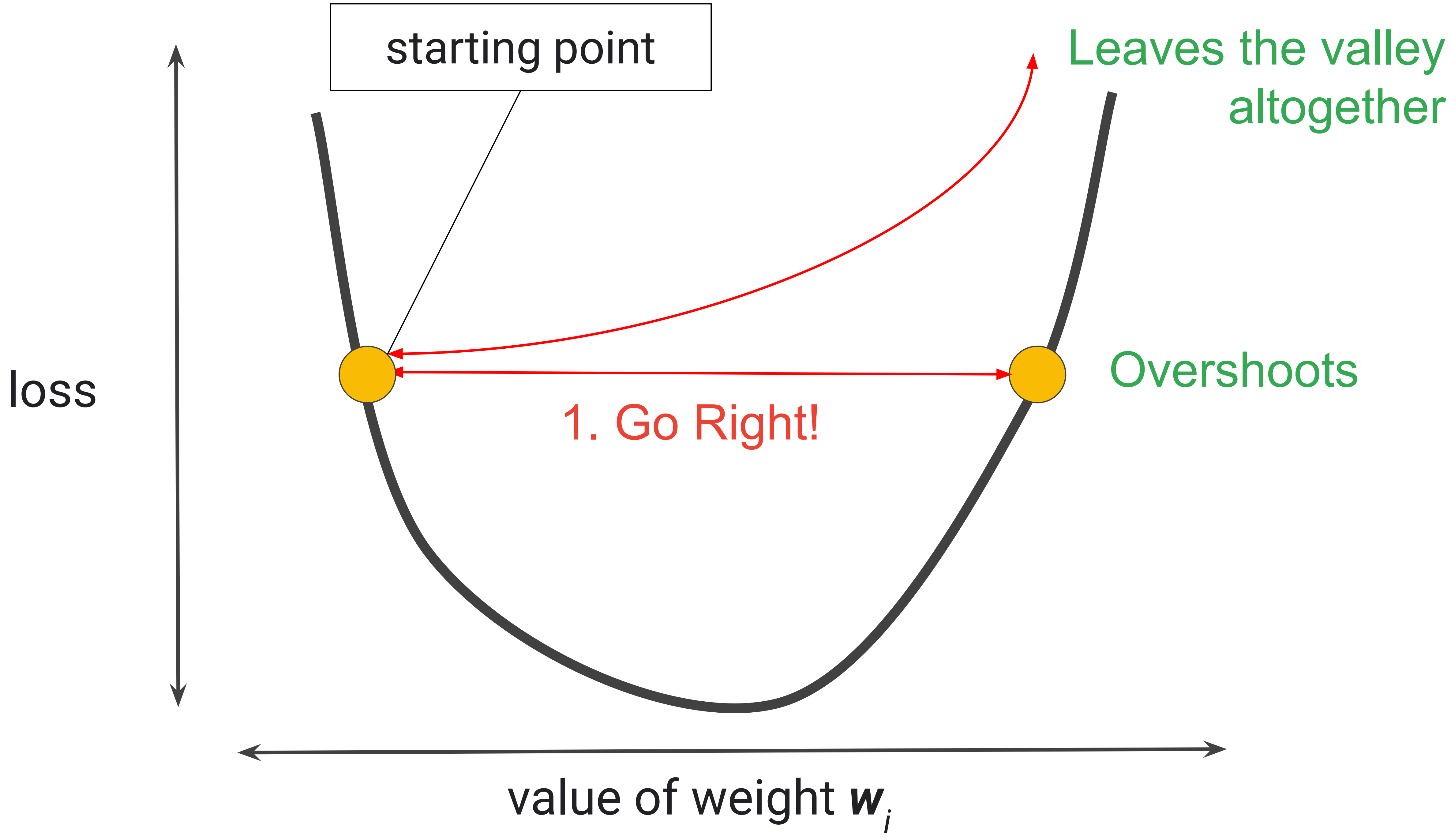
FYI: For your knowledge

Small step sizes can take a very long time to converge



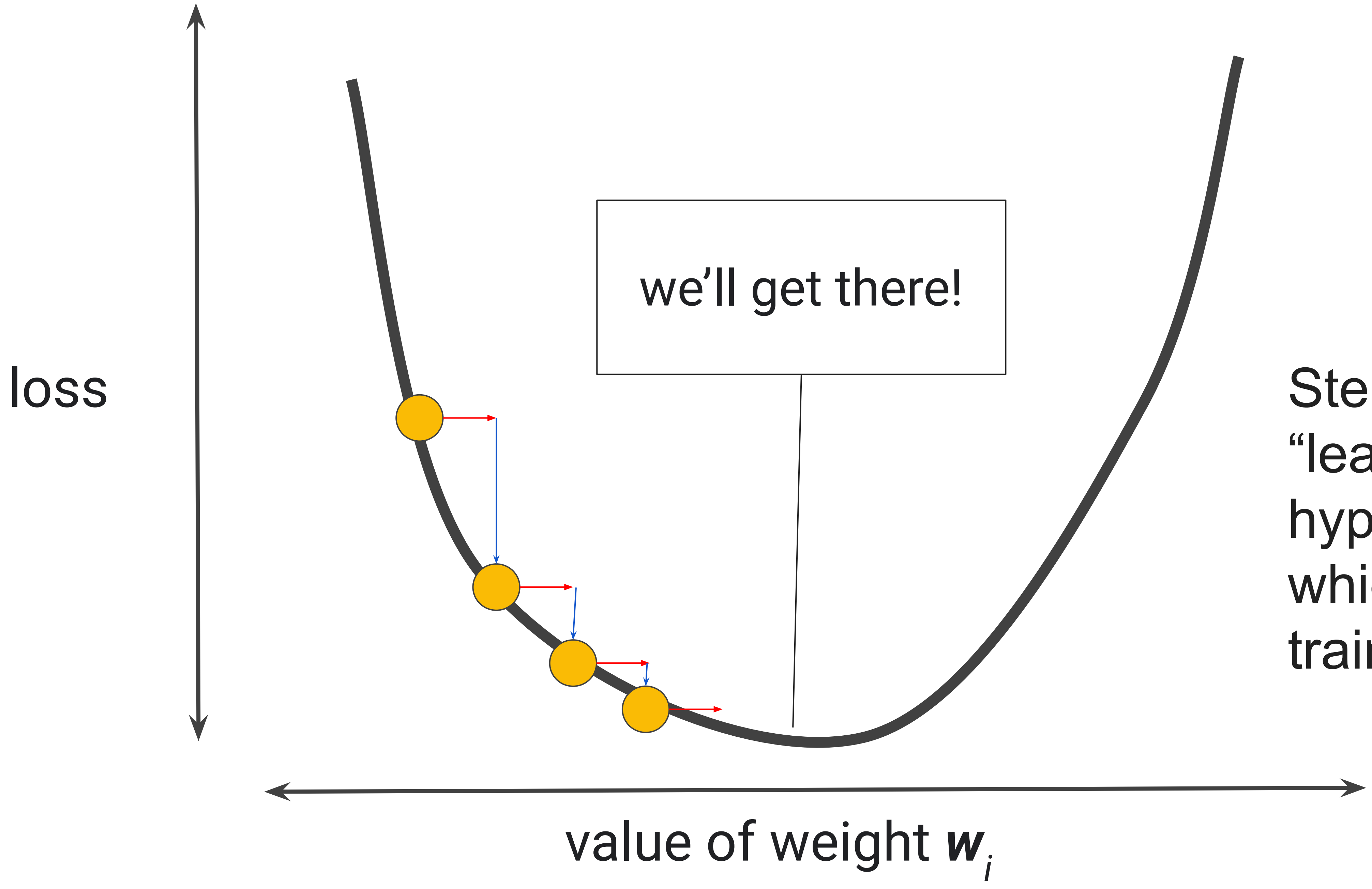
FYI: For your knowledge

Large step sizes may never converge to the true minimum



FYI: For your knowledge

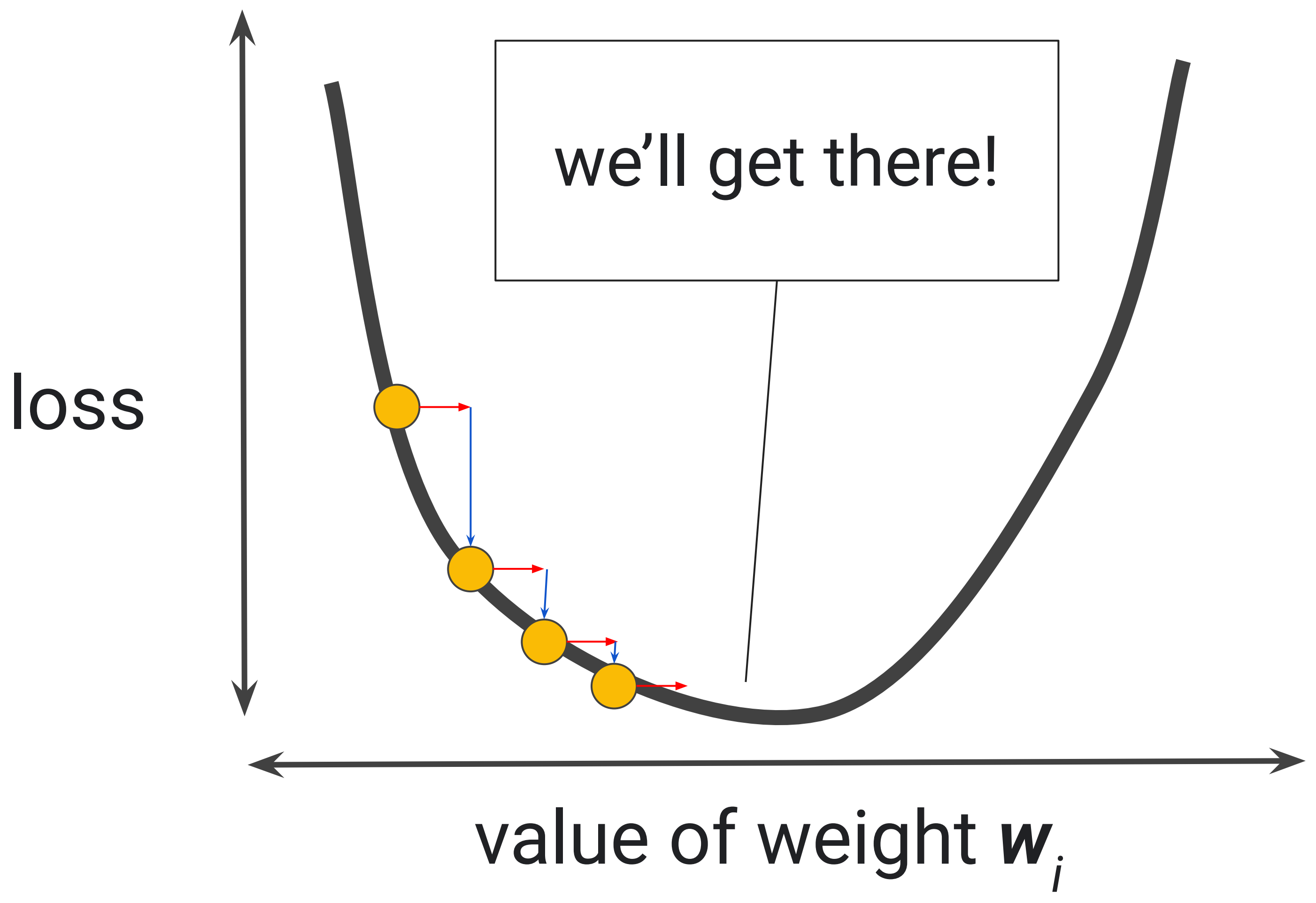
A correct and constant step size can be difficult to find



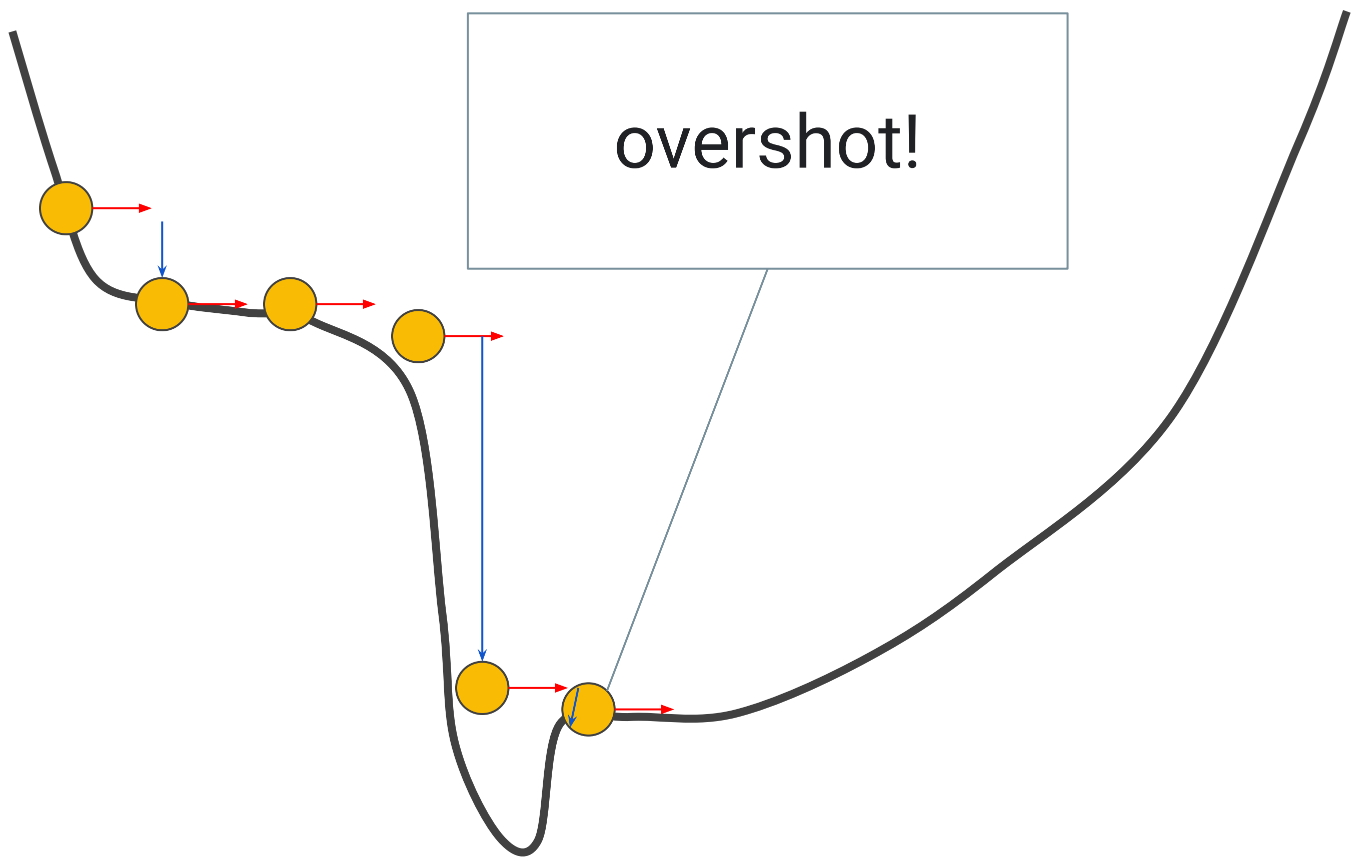
Step size or "learning rate" is a hyper-parameter which is set before training.

FYI: For your knowledge

A correct and constant step size can be difficult to find



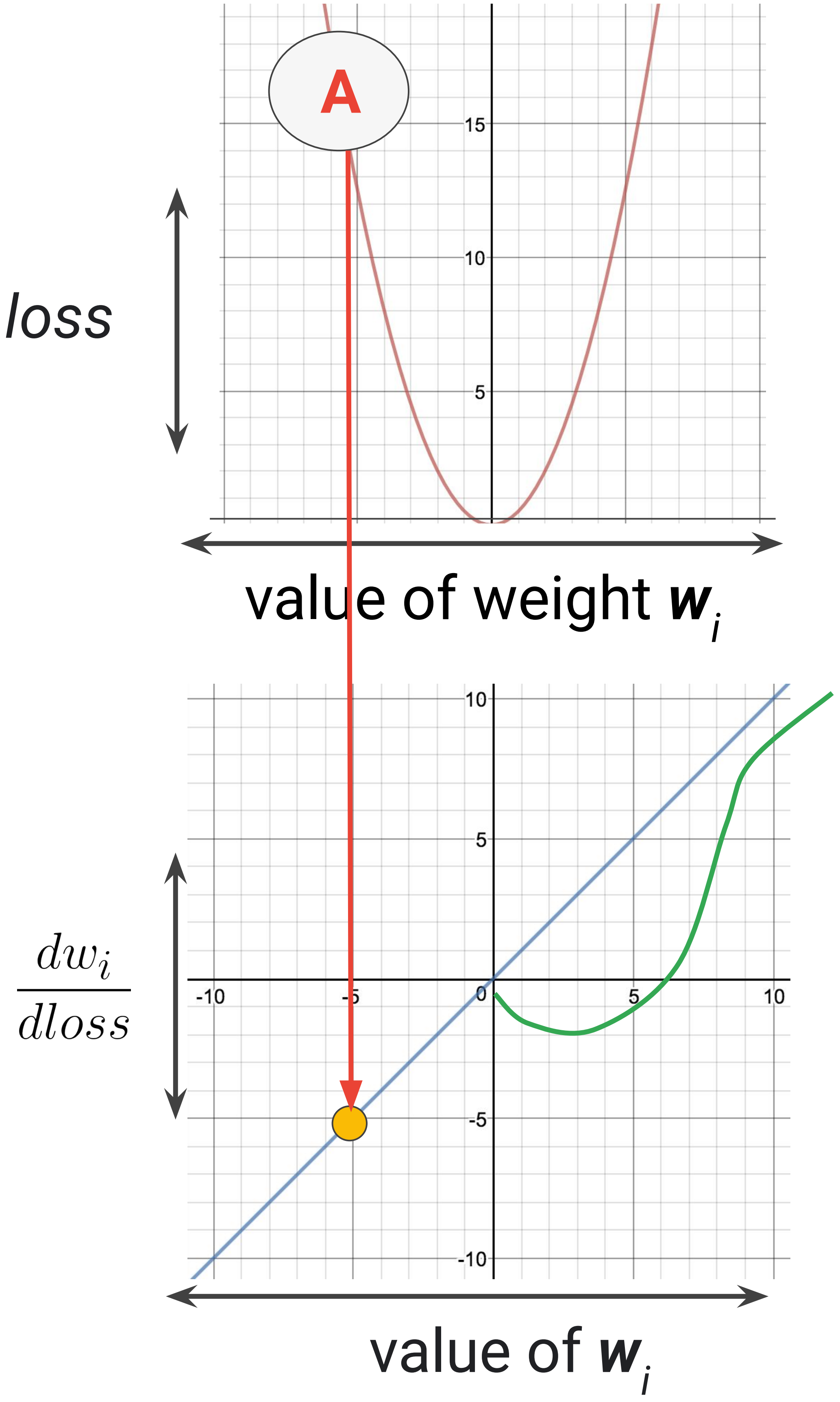
Step size or "learning rate" is a hyper-parameter which is set before training.



One size does not fit all models.

The Loss Function slope provides direction and step size in your search

FYI: For your knowledge



Loss Function
(e.g. RMSE)

Remember, your goal is to find the minimum loss which is where the Loss Function slope is 0.

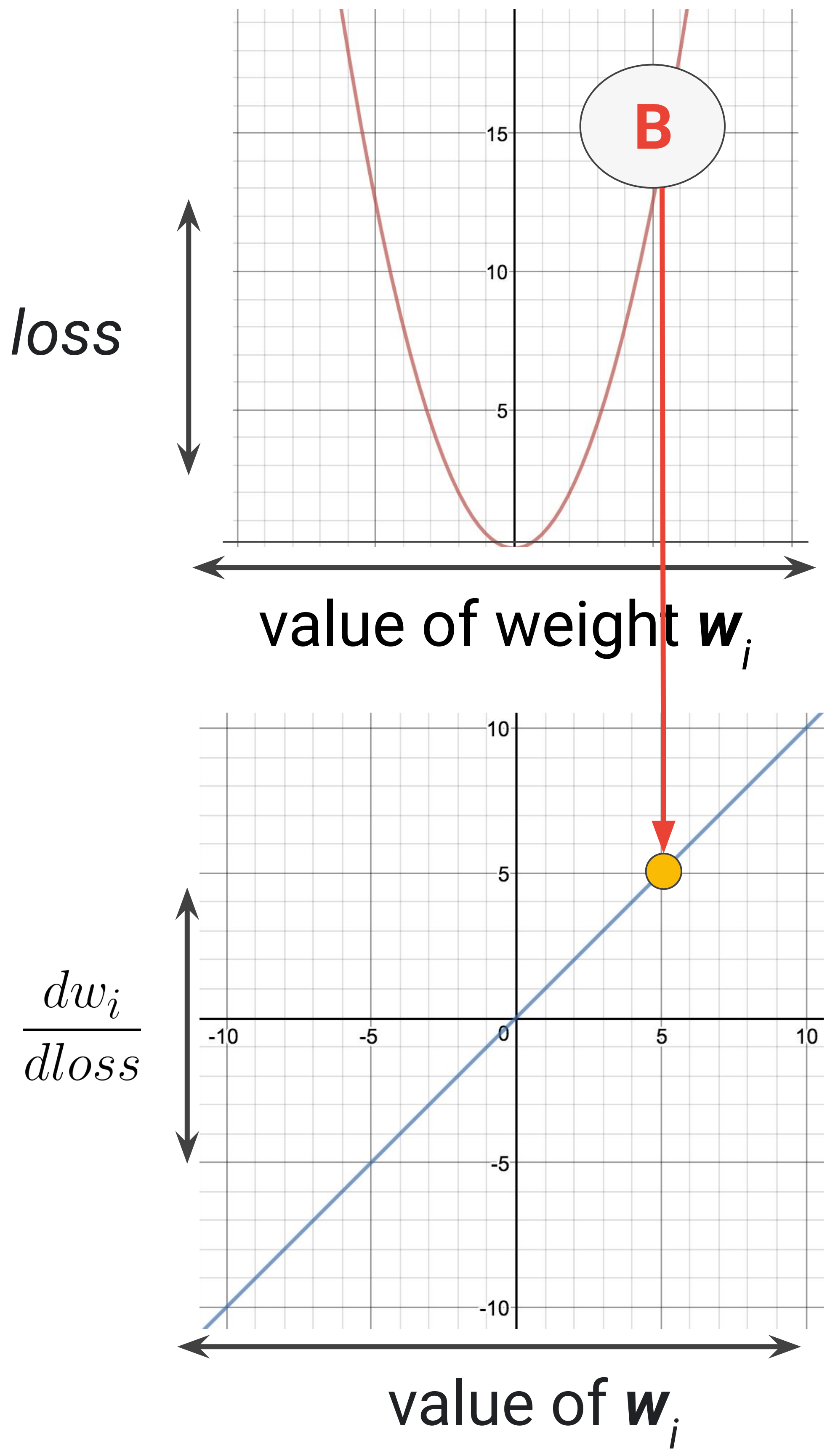
Loss Function Slope (Derivative)

Slope is Negative
Direction: Go Right!

Magnitude is (-5)
Step Size: Big

FYI: For your knowledge

The Loss Function slope provides direction and step size in your search



Loss Function
(e.g. RMSE)

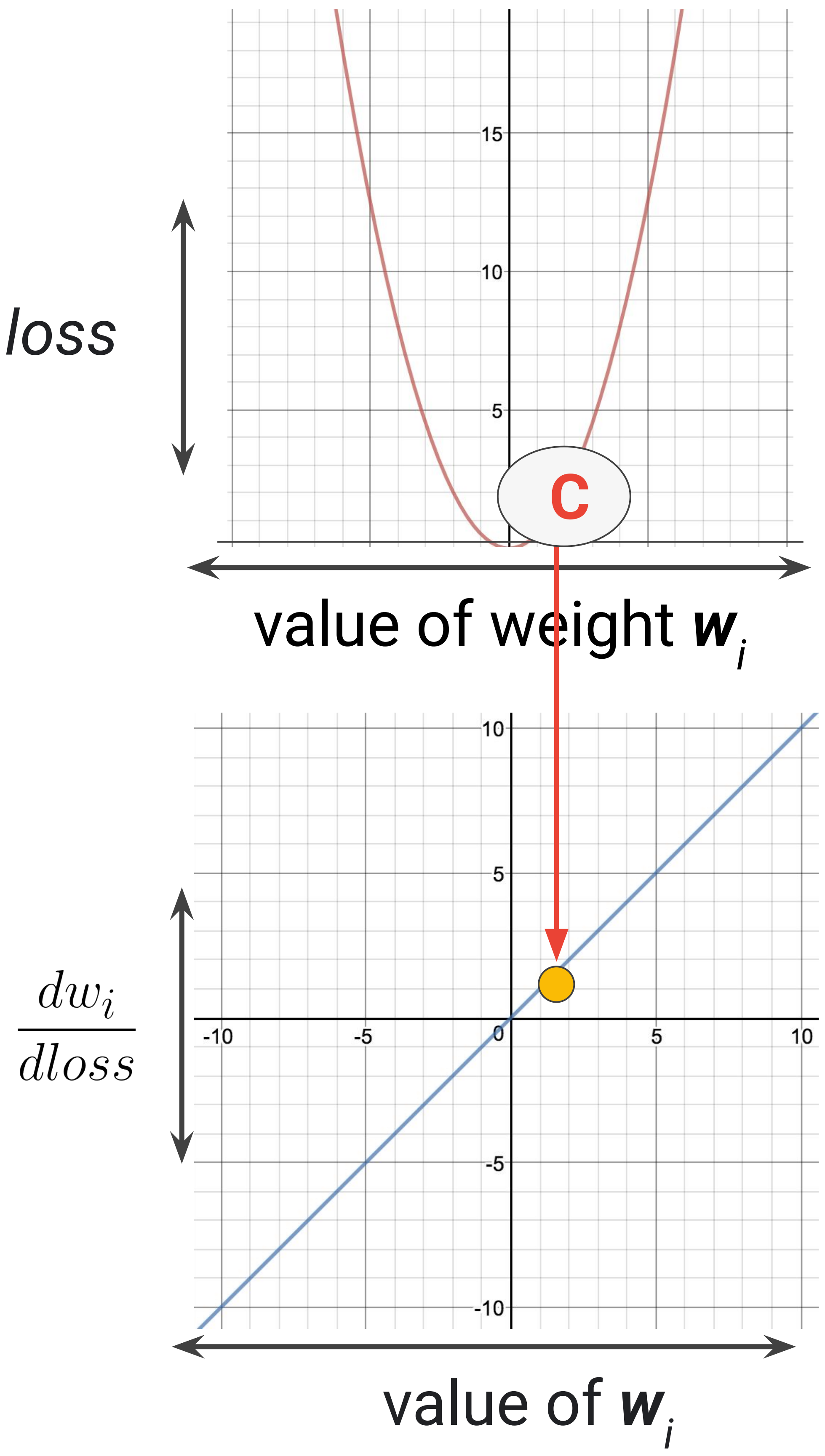
Slope is Positive
Direction: Go Left!

Magnitude is 5
Step Size: Big

Loss Function Slope (Derivative)

FYI: For your knowledge

The Loss Function slope provides direction and step size in your search



Loss Function
(e.g. RMSE)

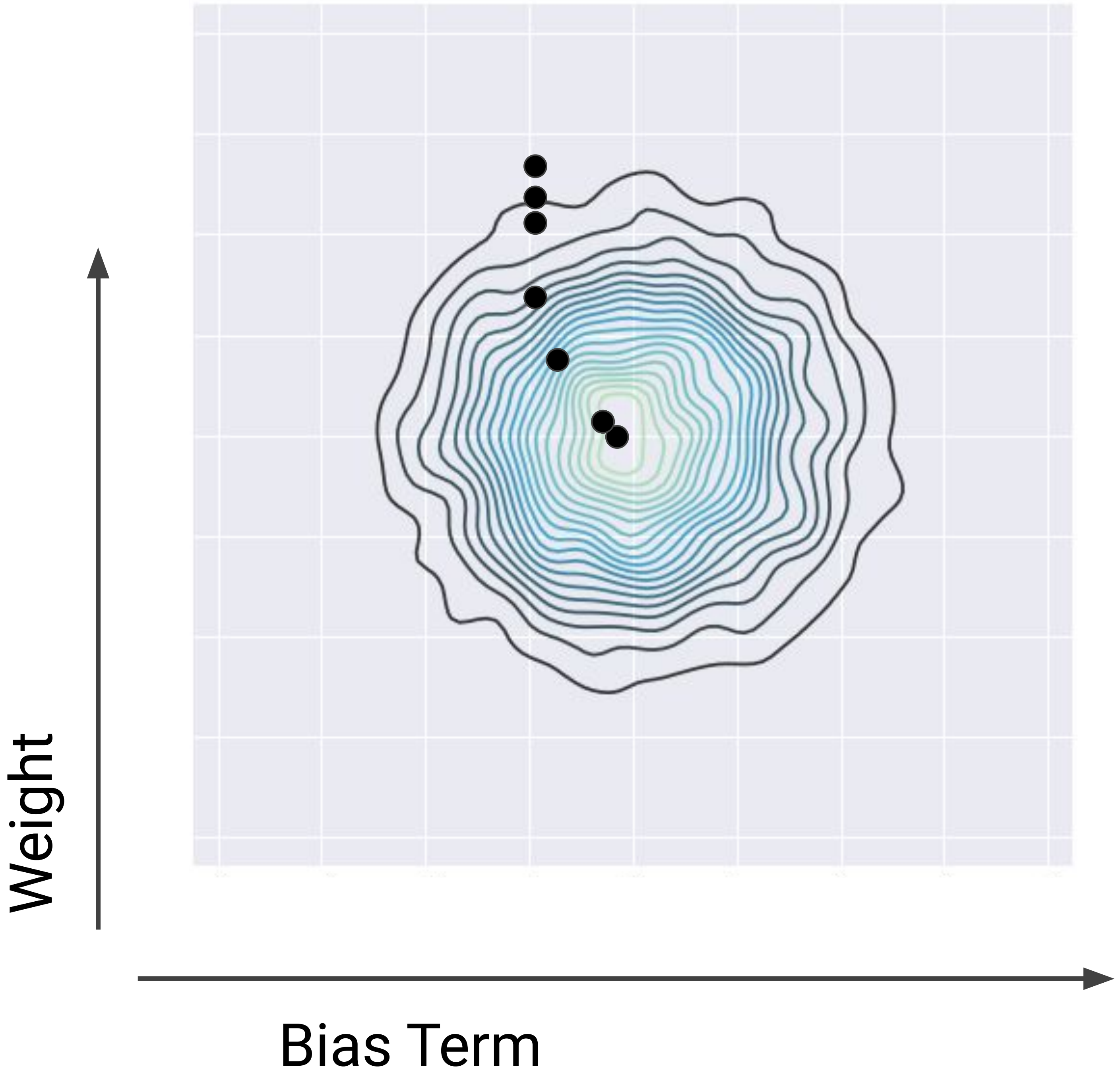
Slope is Positive
Direction: Go Left!

Magnitude is 2
Step Size

Loss Function Slope (Derivative)

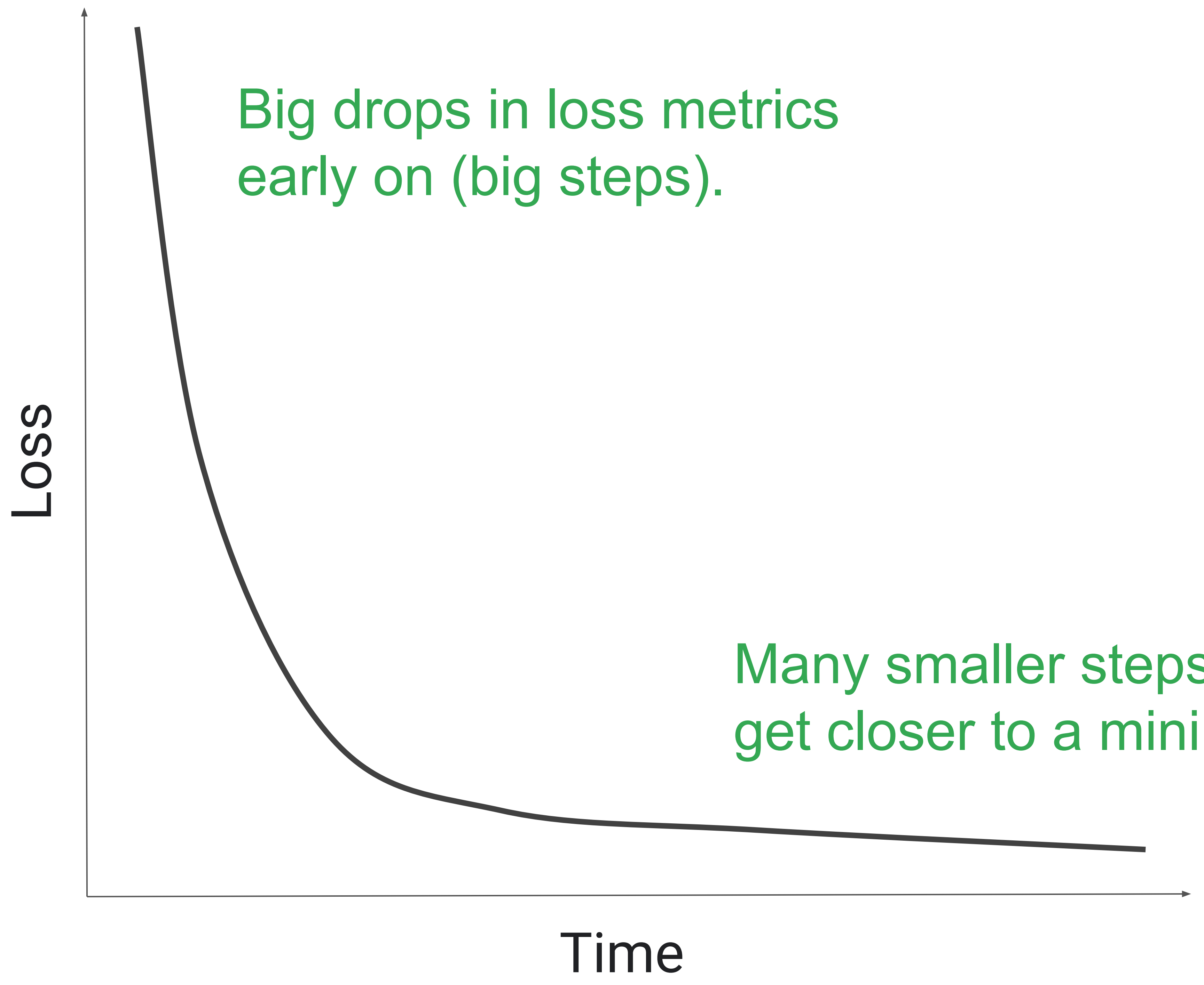
FYI: For your knowledge

Are you done yet?



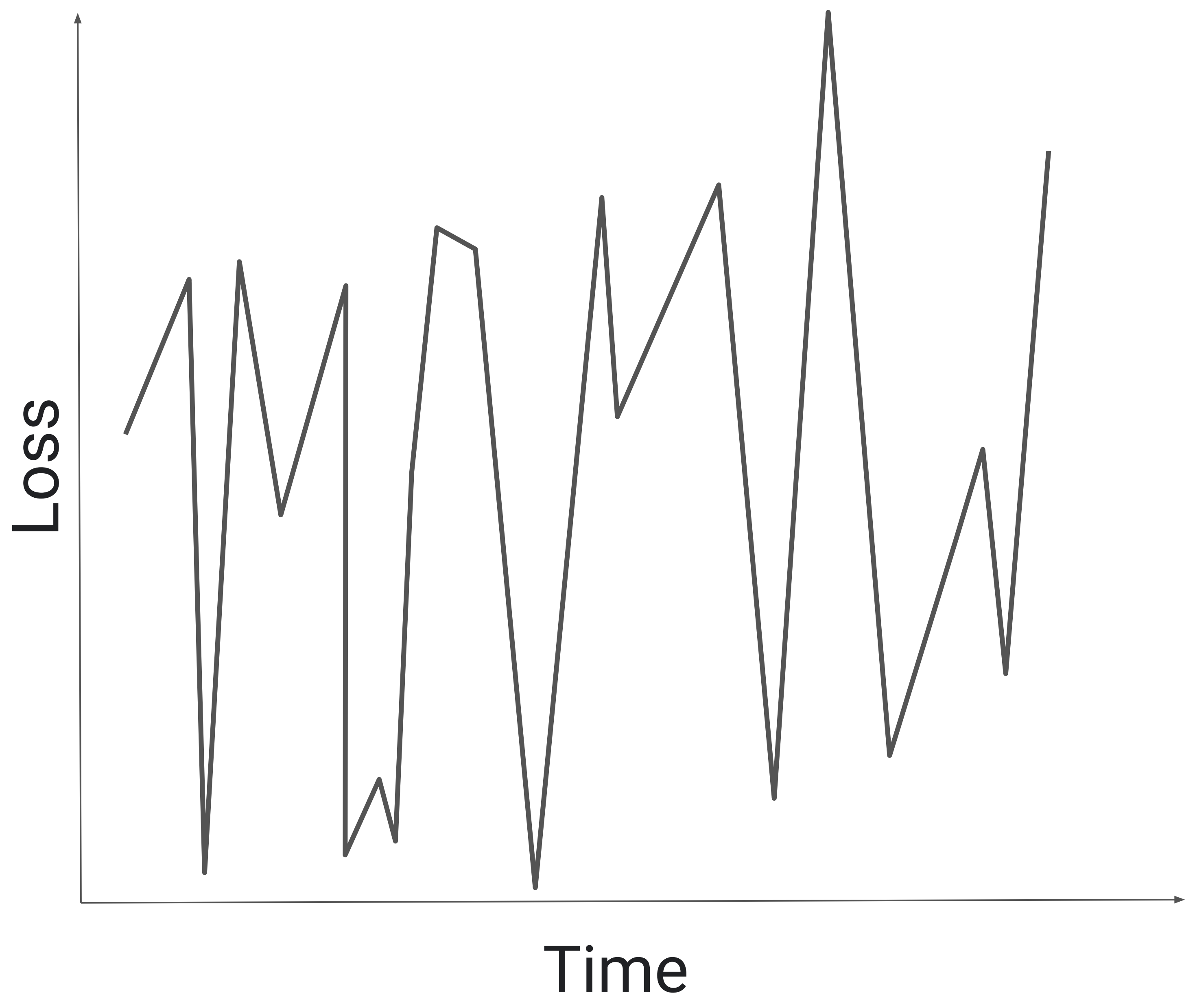

```
while loss is > Epsilon:
   derivative = computeDerivative()
   for i in range(self.params):
      self.params[i] = //
                  self.params[i] //
                        - derivative[i]
   loss = computeLoss()
```

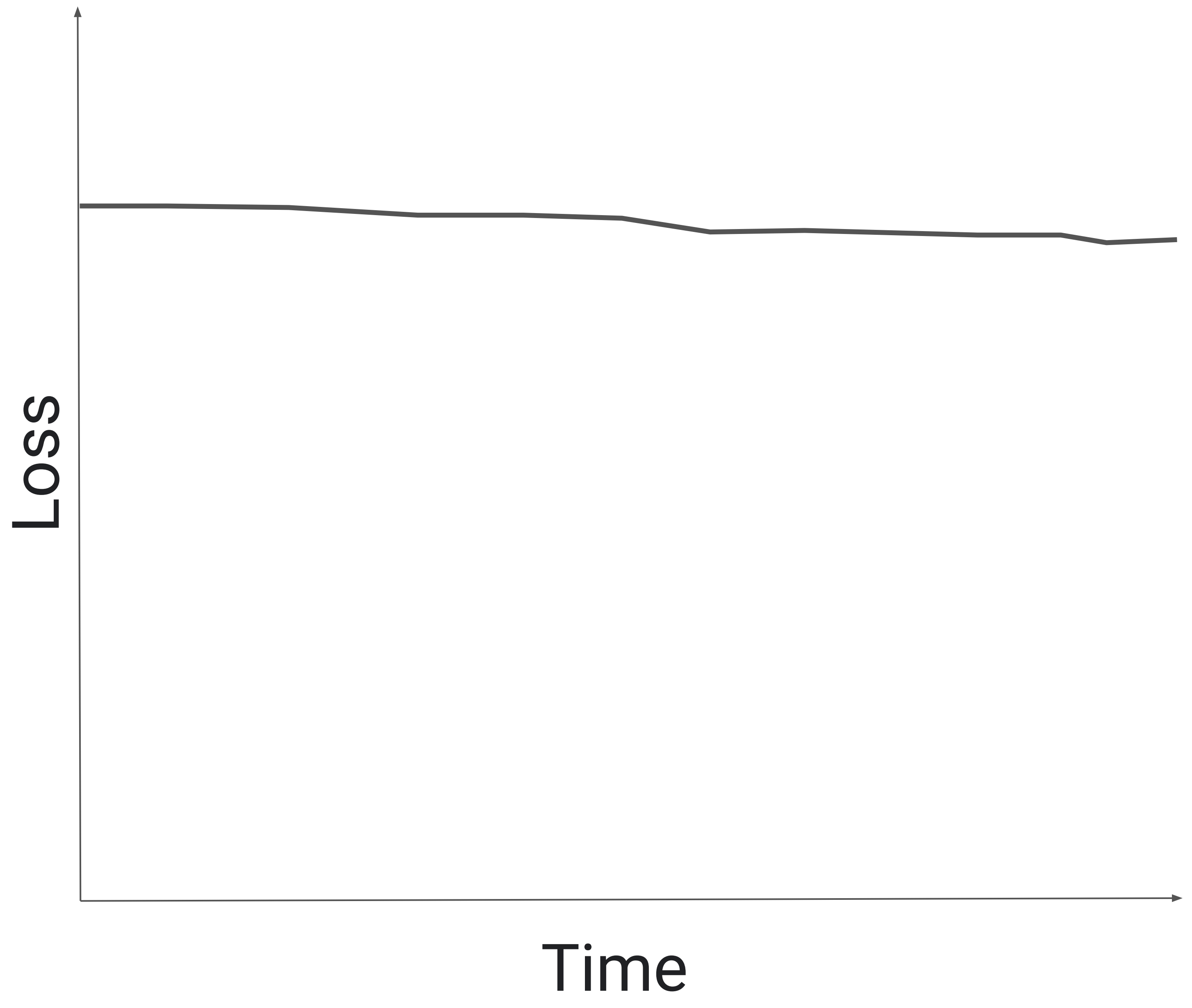
A typical loss curve



Needed For the Exam

Troubleshooting a Loss Curve





FYI: For your knowledge

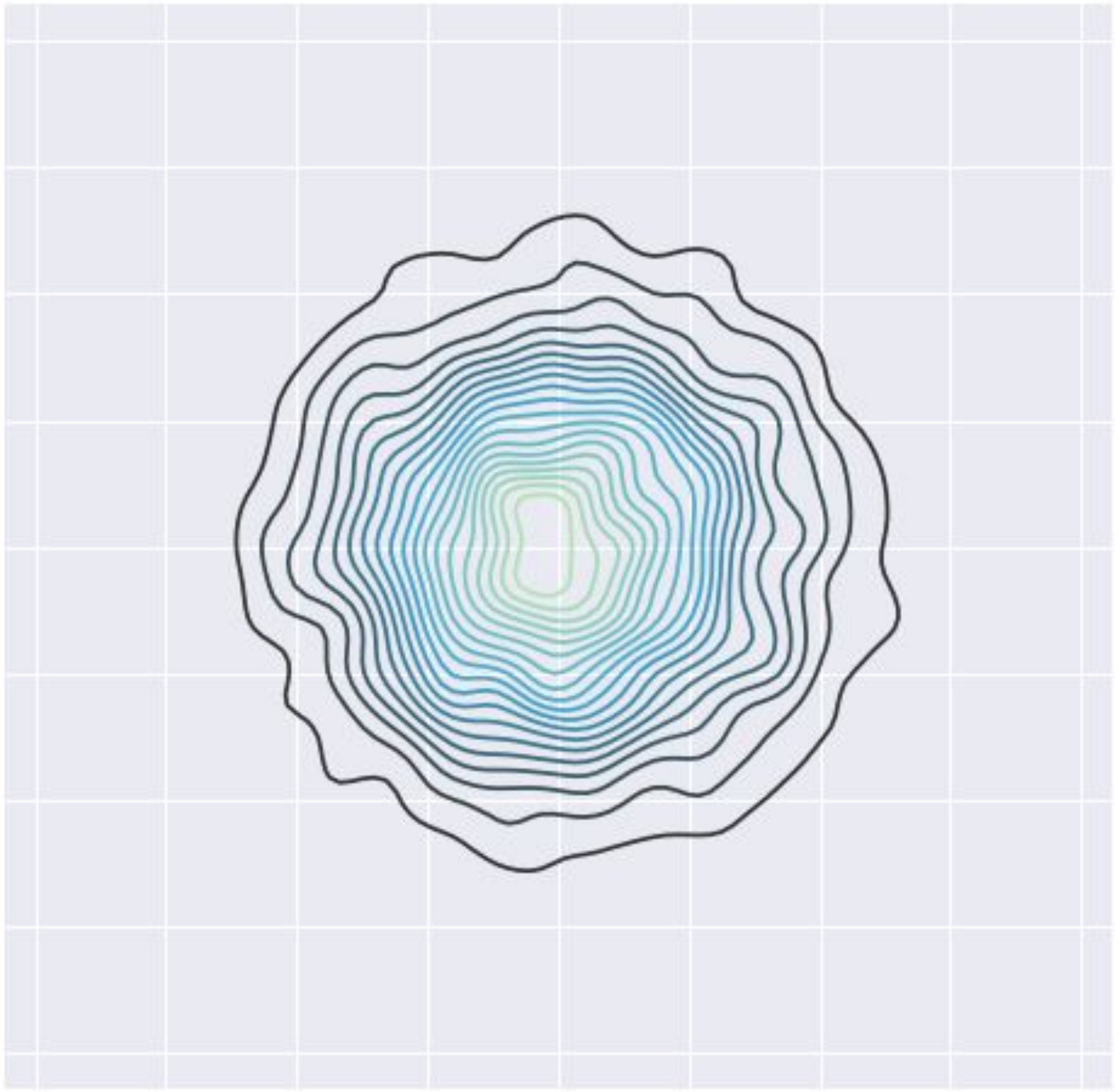
Adding a scaling hyperparameter

```
while loss is > Epsilon:
    derivative = computeDerivative()
  for i in range(self.params):
    self.params[i] = //
            self.params[i] //
                  - learning_rate //
                  * derivative[i]
  loss = computeLoss()
```

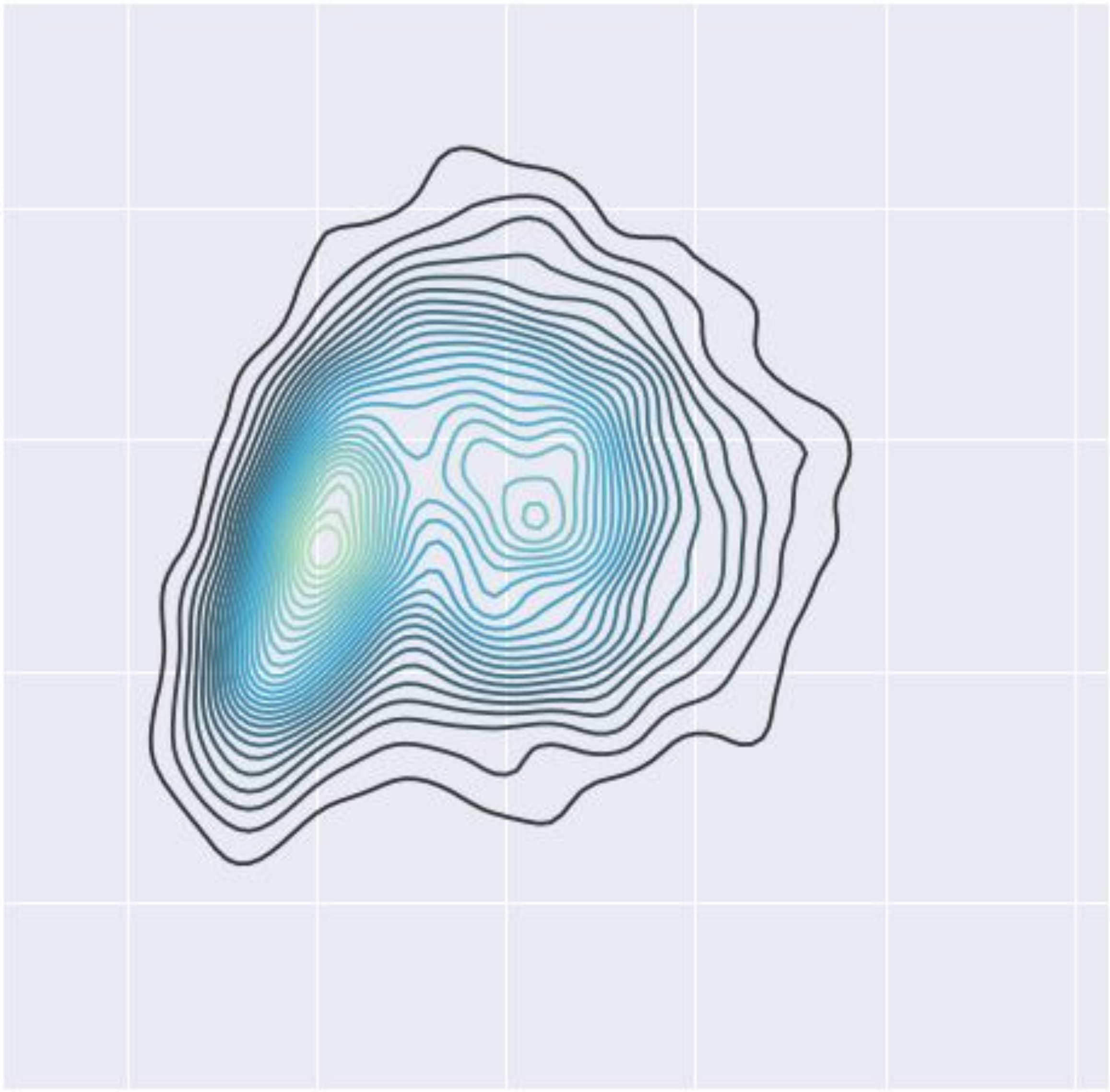
FYI: For your knowledge

Problem: Multiple minima may exist

Loss Surface with a global minimum



Loss Surface with more than one minima



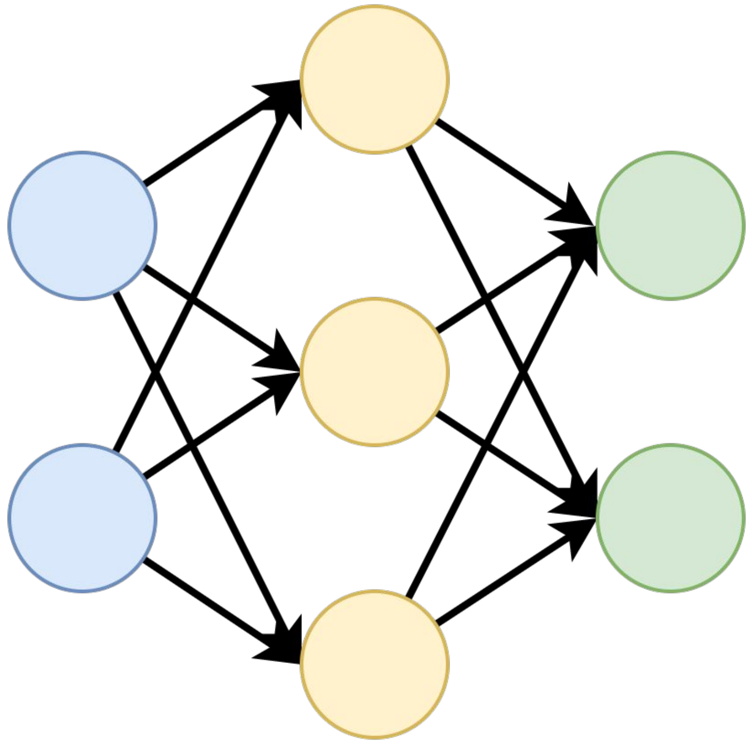
How do we get down… Optimizers

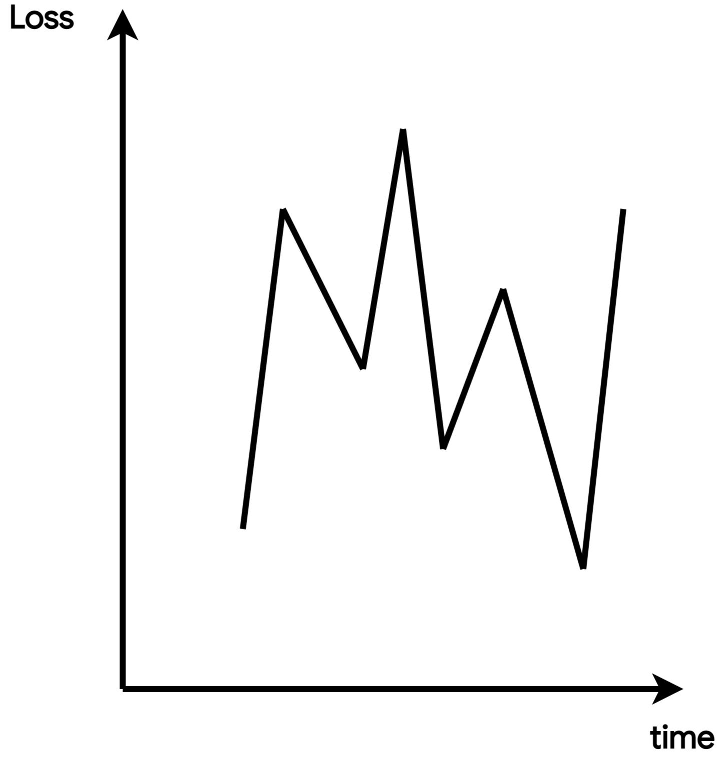
FYI: For your knowledge

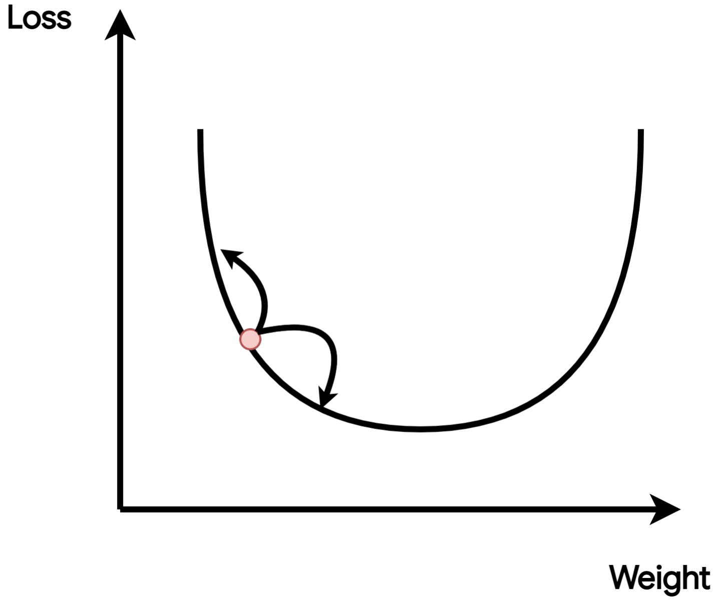
FYI: For your knowledge

Solution 1: Stochastic Gradient Descend

- Instead of computing the Gradient on the whole dataset, do it for every single sample



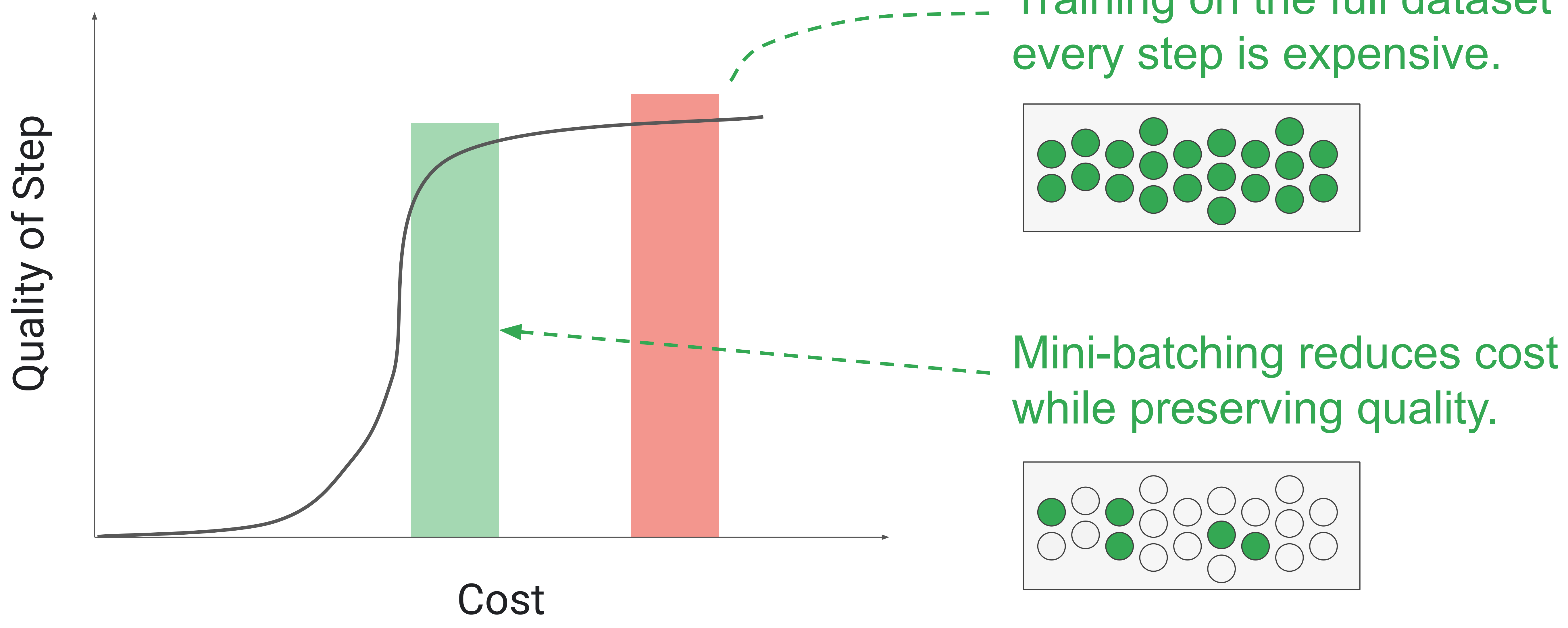




Stochastic just does not work. Too noisy jumps

FYI: For your knowledge

Solution 2: Mini-batch Gradient Descend… Good compromise



Typical values for batch size: 10 - 1000 examples.

FYI: For your knowledge

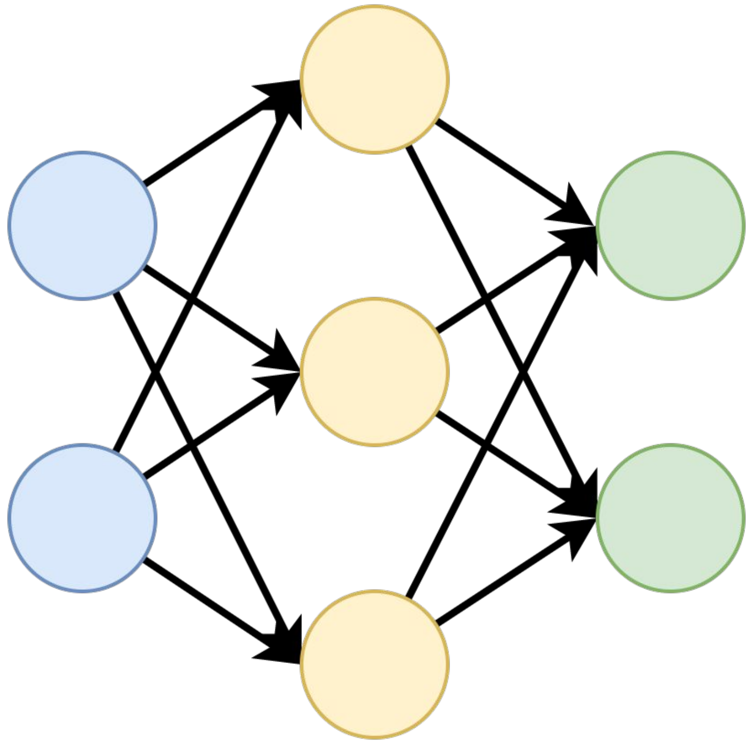
Solution 3: We can add Momentum to SGD

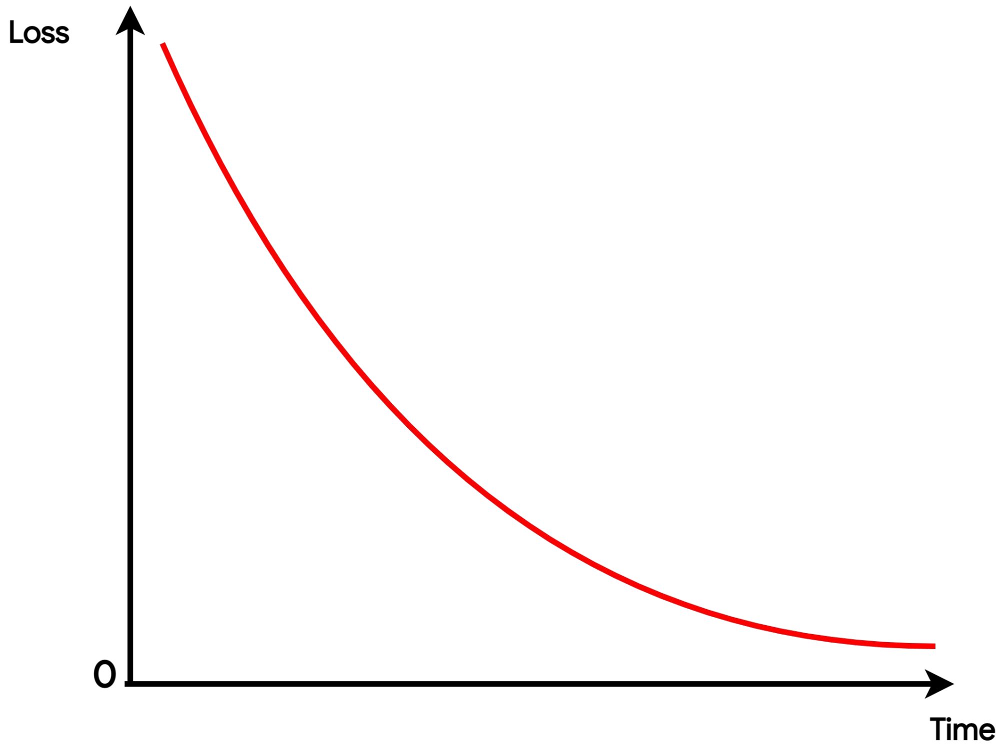
FYI: For your knowledge

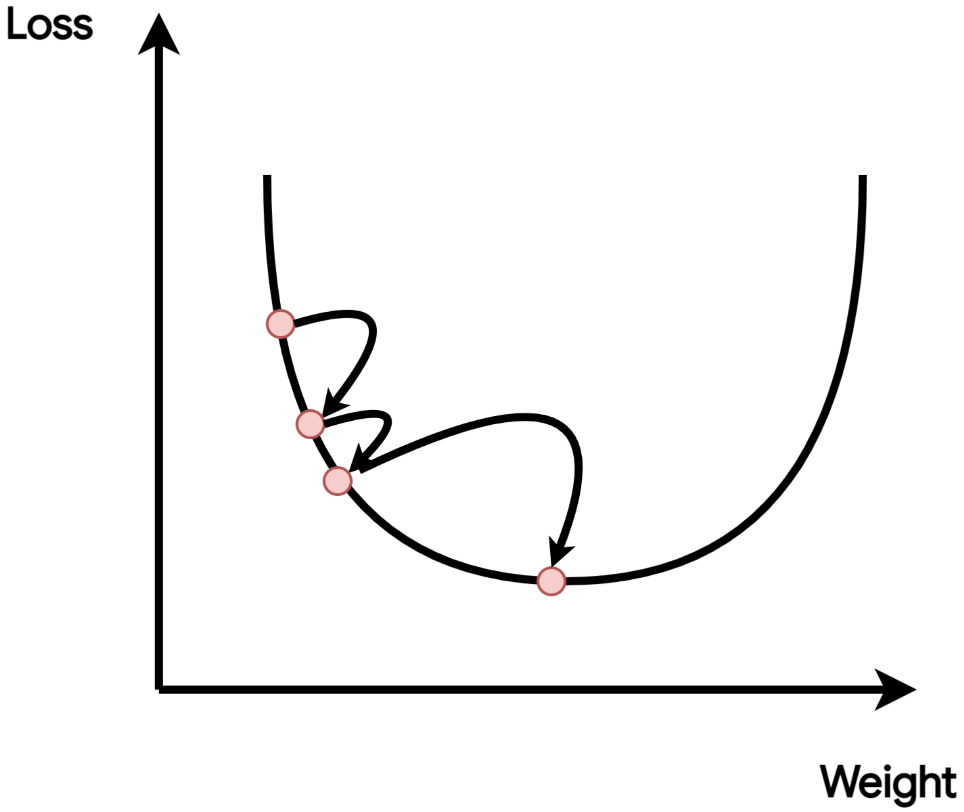
ADAM: Adaptive Moment Estimation

FYI: For your knowledge

Adam lets you learn faster and precisely







Needed For the Exam

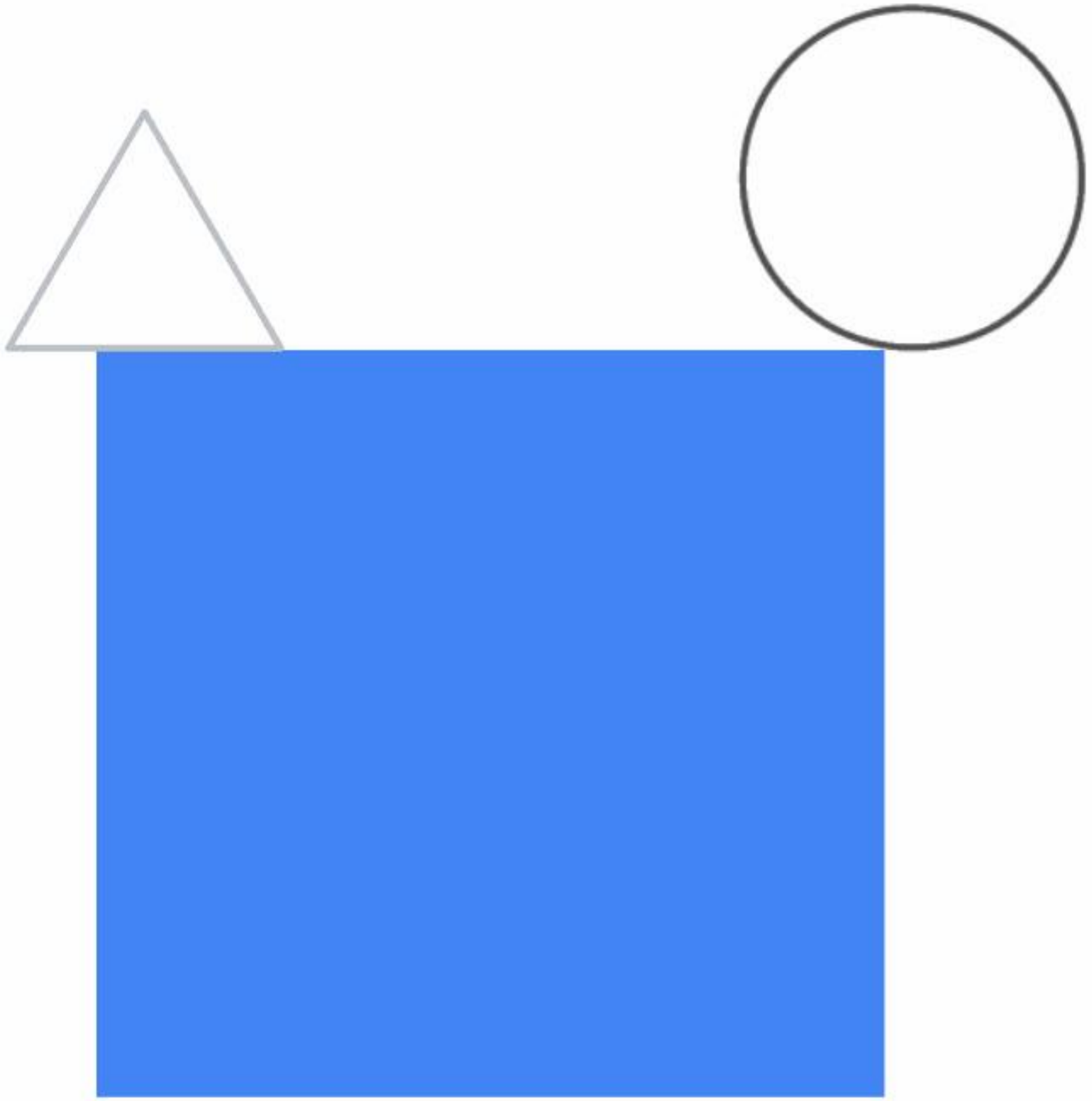
In summary…

- Gradient Descend cannot work with a huge dataset
- Mini-Batch can be a good compromise to find the global minima with large dataset
- ADAM is one of the most used Optimizer when it comes for training large number of parameters such as NN.

Break time

We will resume in:

<<8:00->>



Session 1: agenda

01 Streaming Data in Google Cloud with Pub/Sub and DataFlow

02 How Google Does Machine Learning

03 Machine Learning Basics: Part1- Algorithms

04 Machine Learning Basics: Part2 - Model Metrics

Machine Learning Basics: Model Metrics

Part 2

Instructor: Ben Ahmed

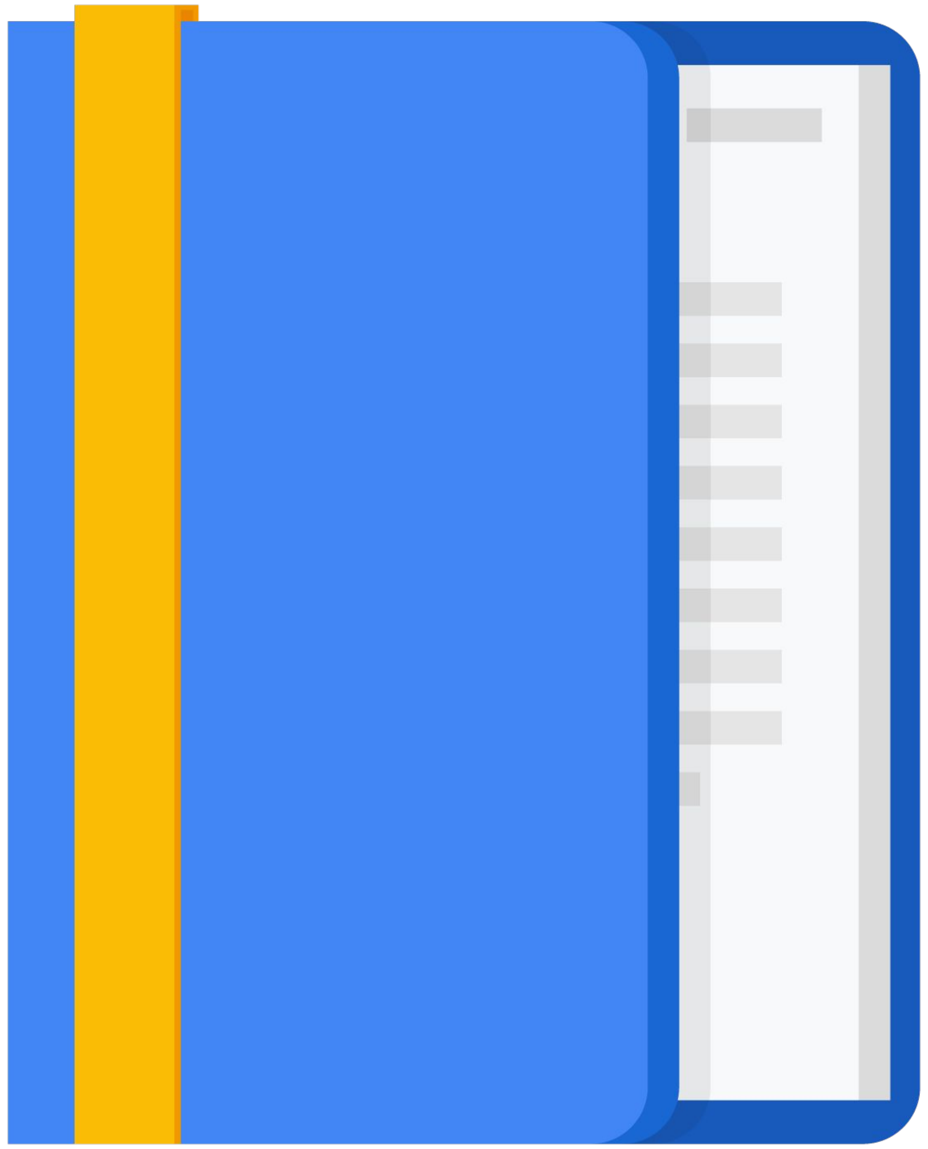
Section Agenda

Generalization in ML

Sampling

Regularizations

ML Model Performance Evaluation



Needed For the Exam

Suppose we want to predict duration of pregnancy based on mother's weight gain in pounds

What is the error measure to optimize?

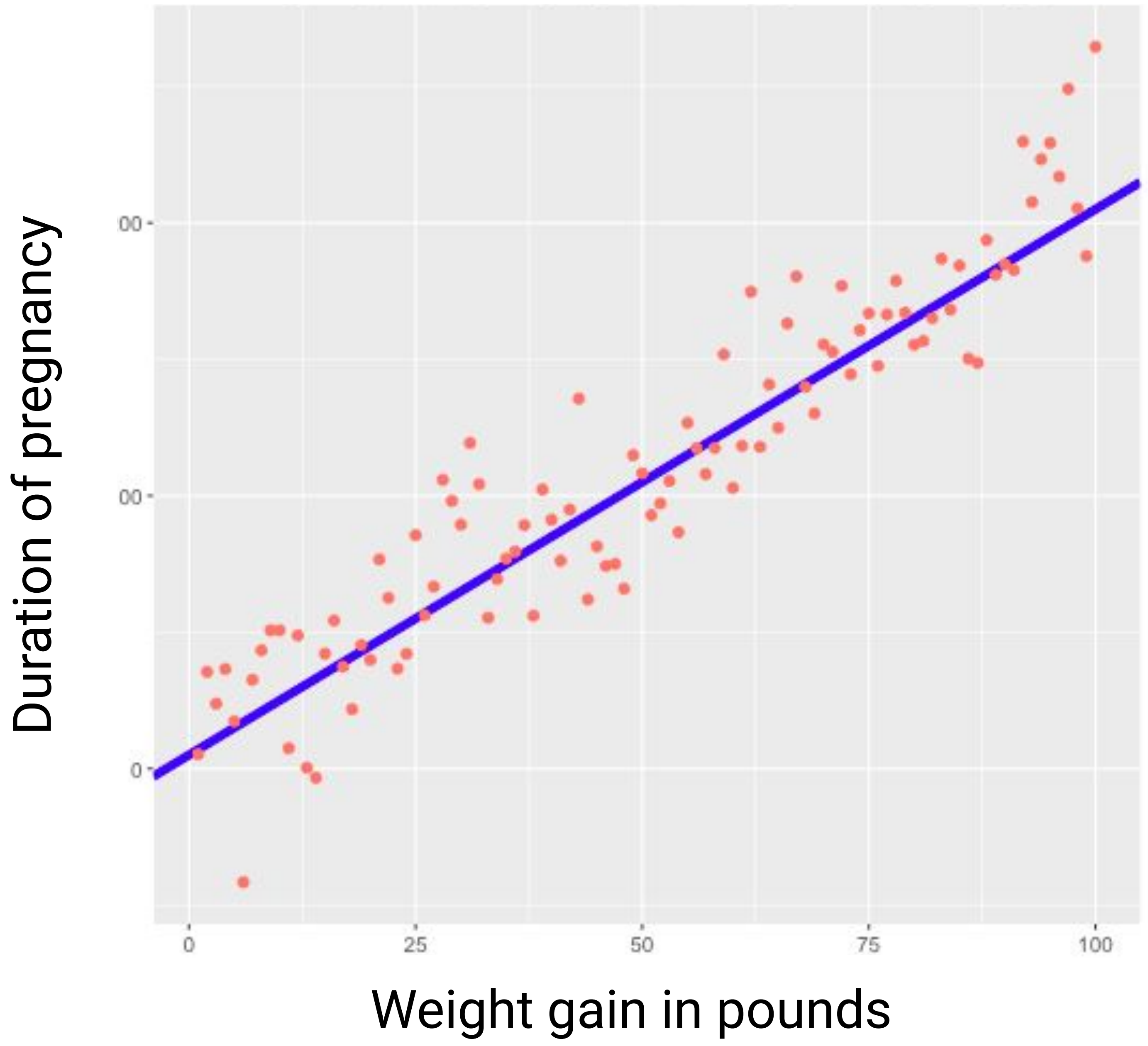
Needed For the Exam

Model 1 is a linear model using linear regression

Red = training examples

Blue = model prediction for each baby

RMSE = 2.224



Needed For the Exam

Model 2 has more free parameters

RMSE = 0

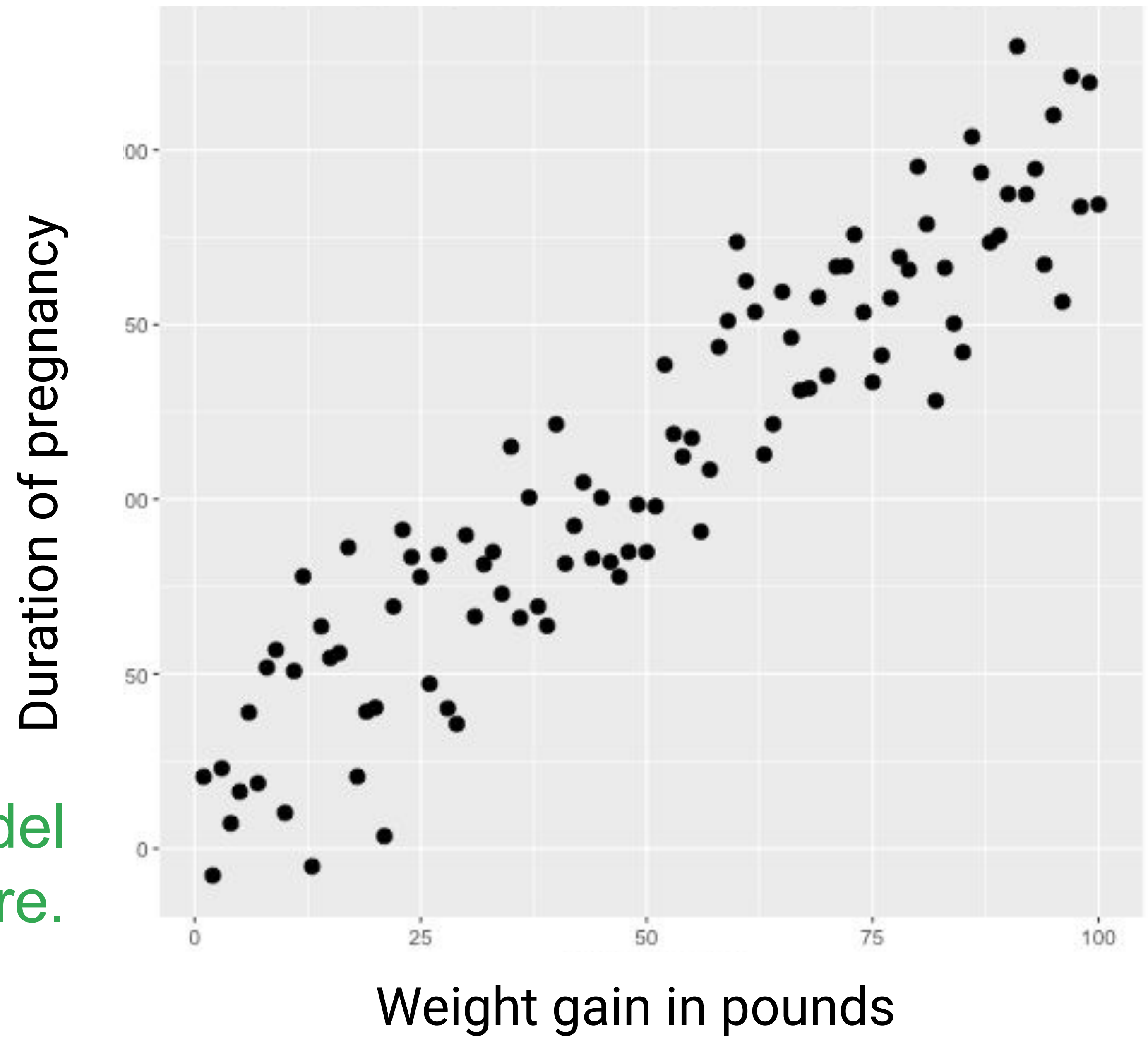
Which model is better?

How can you tell?

Does the model generalize to new data?

Need data that were not used in training.

New data the model hasn't seen before.



Needed For the Exam

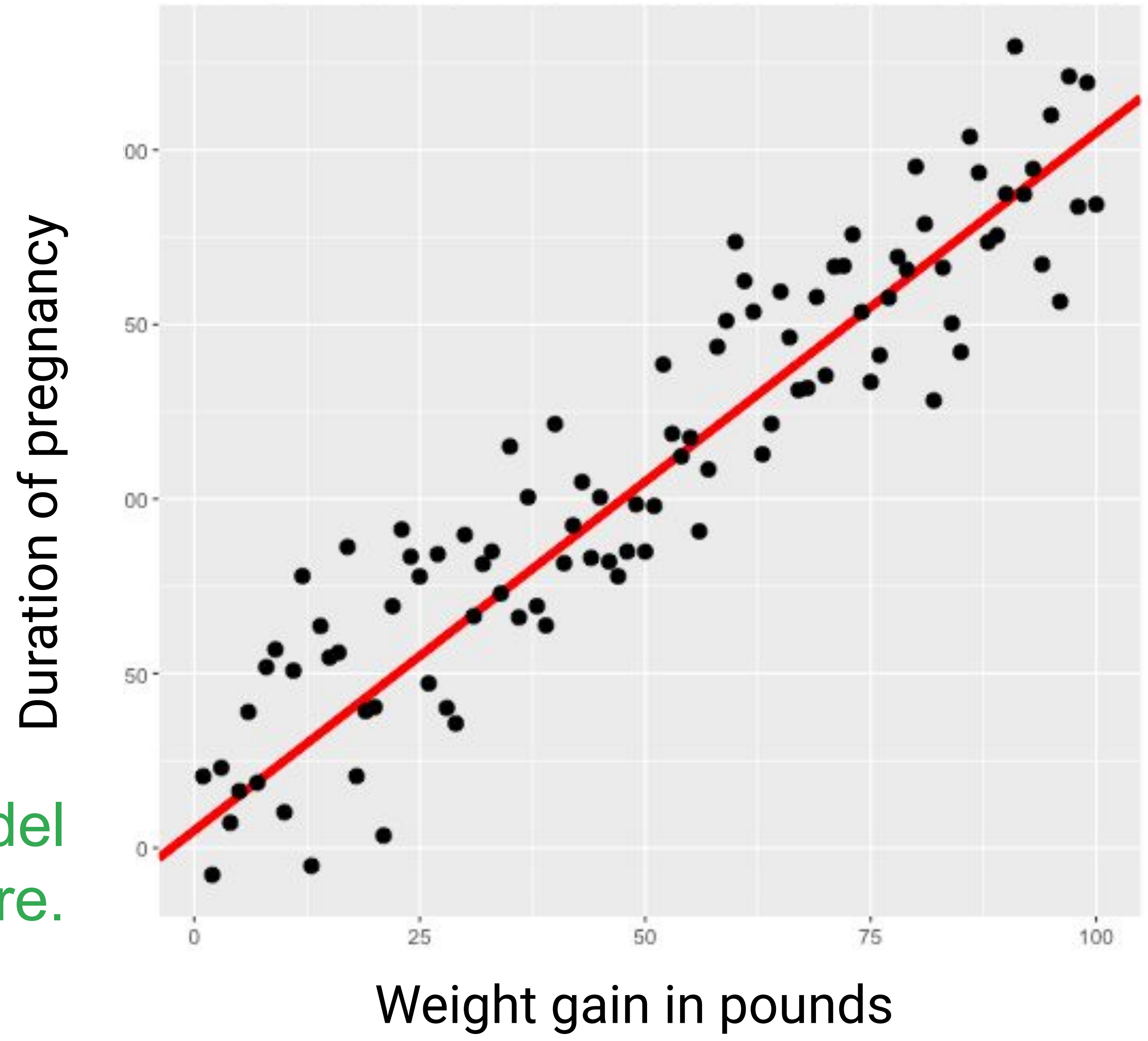
Model 1 generalizes well

Old RMSE = 2.224

New RMSE = 2.198

Pretty similar = good

New data the model hasn't seen before.



Needed For the Exam

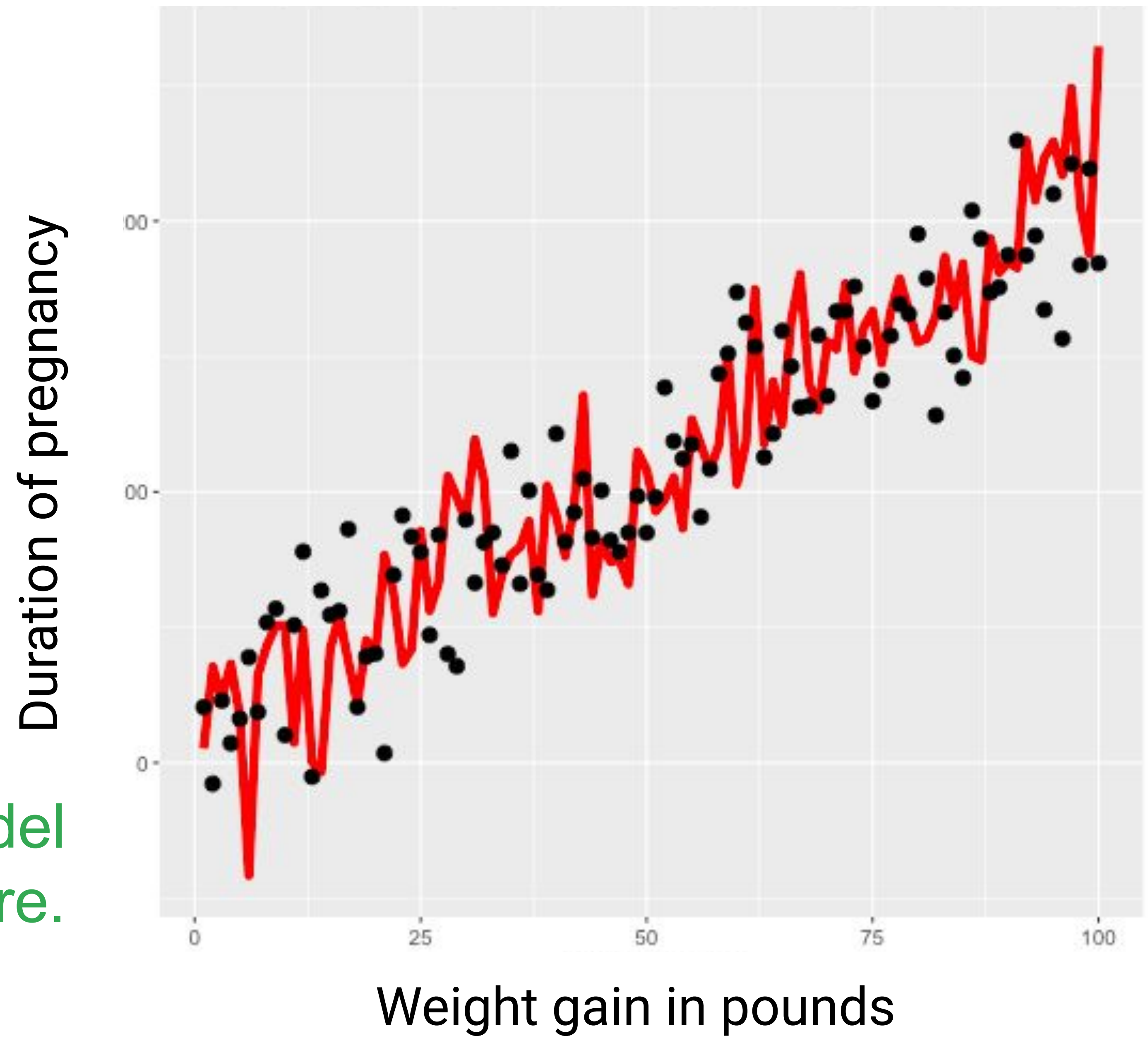
Model 2 does not generalize well

Old RMSE = 0

New RMSE = 3.2

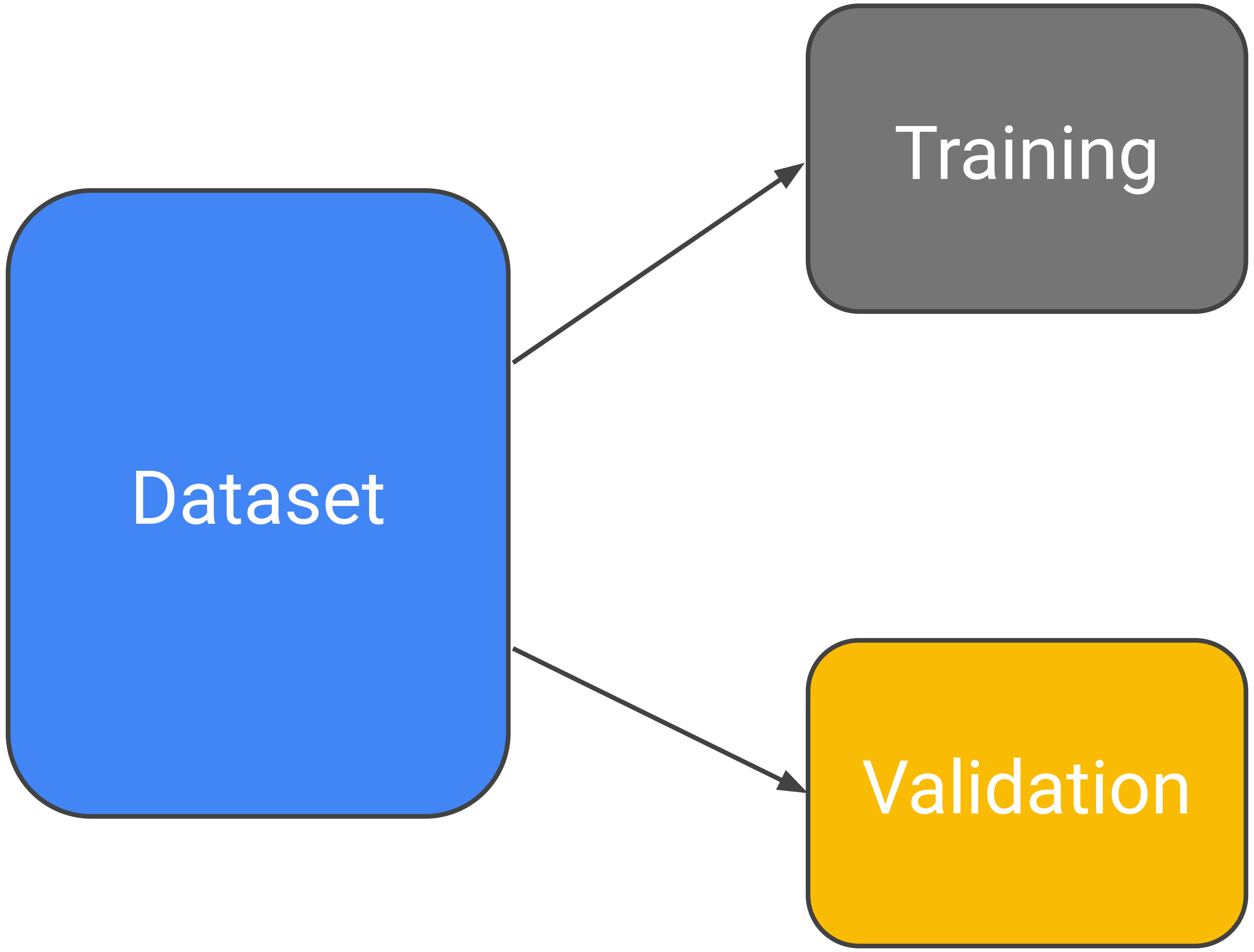
This is a red flag

New data the model hasn't seen before.



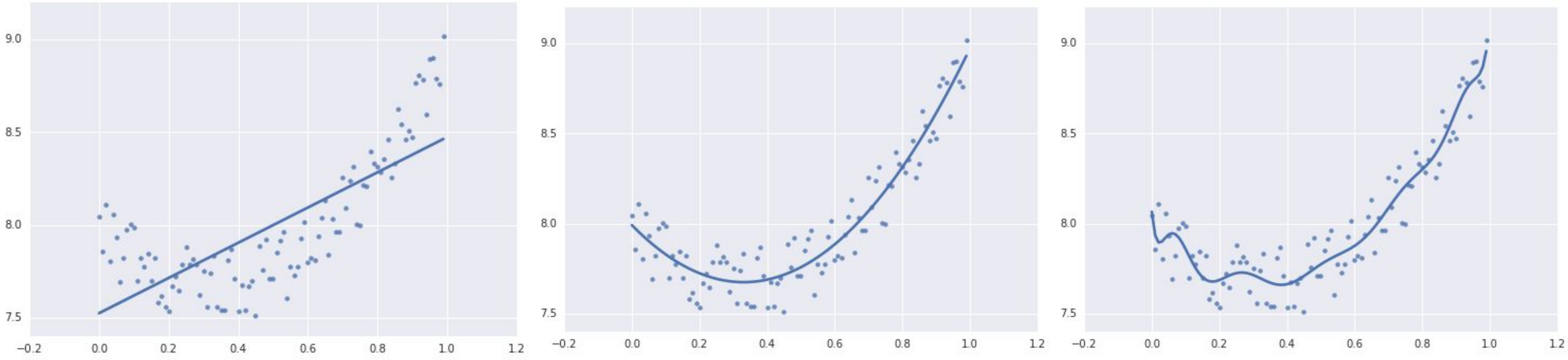
Needed For the Exam

Split the dataset and experiment with models



Needed For the Exam

Beware of overfitting as you increase model complexity

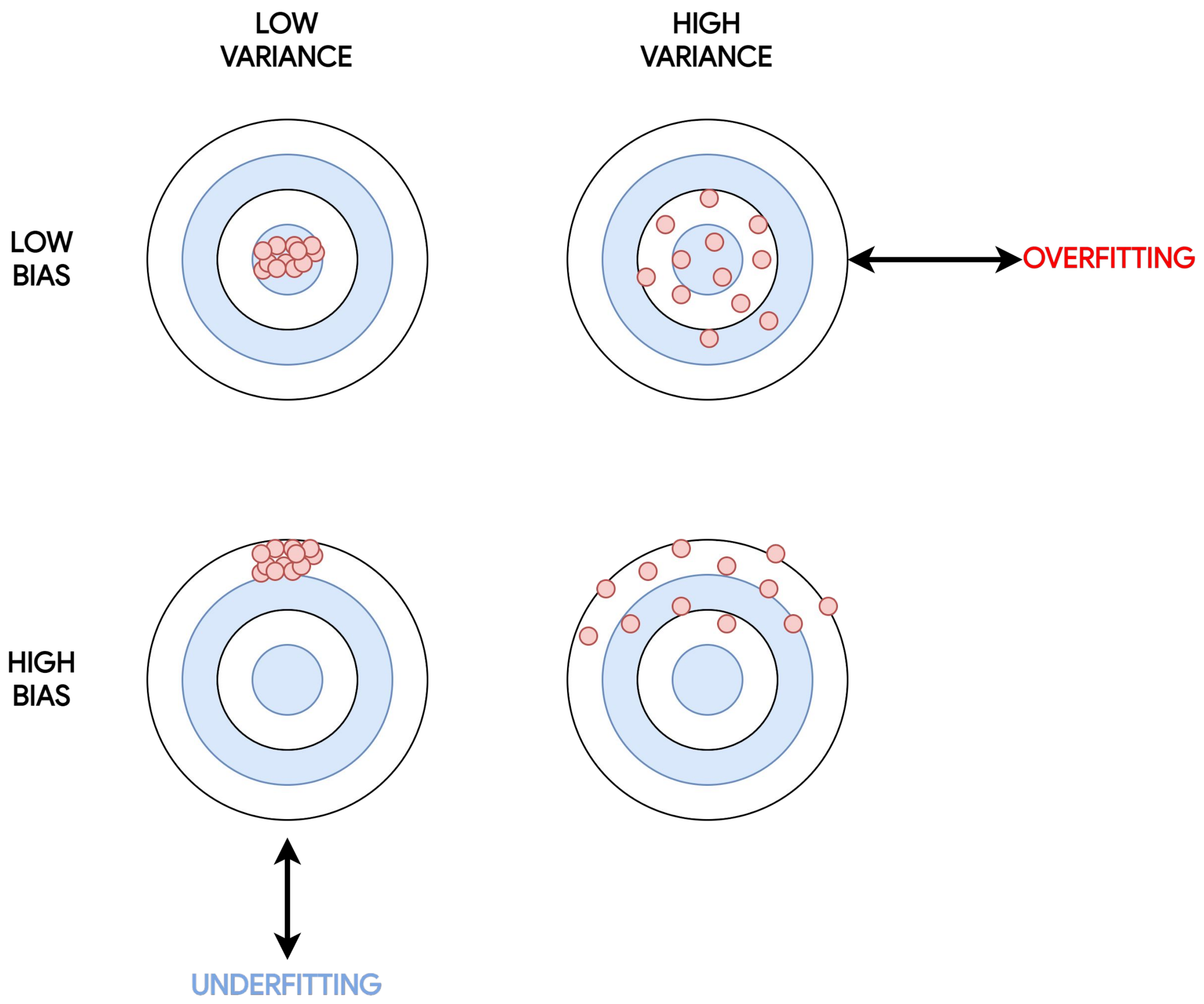


Underfit — High Bias

Fit

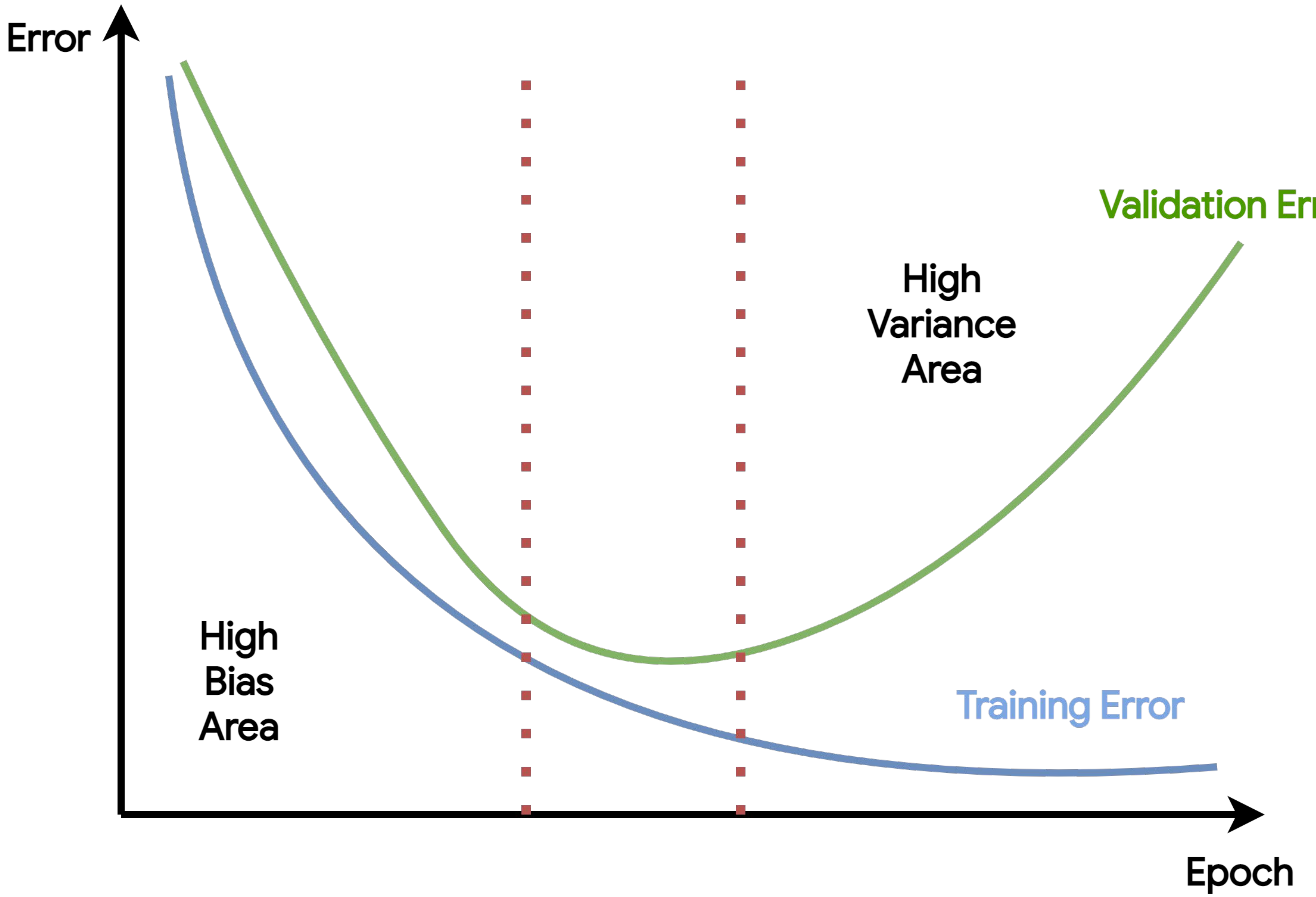
Overfit — High Variance

Bias VS Variance…



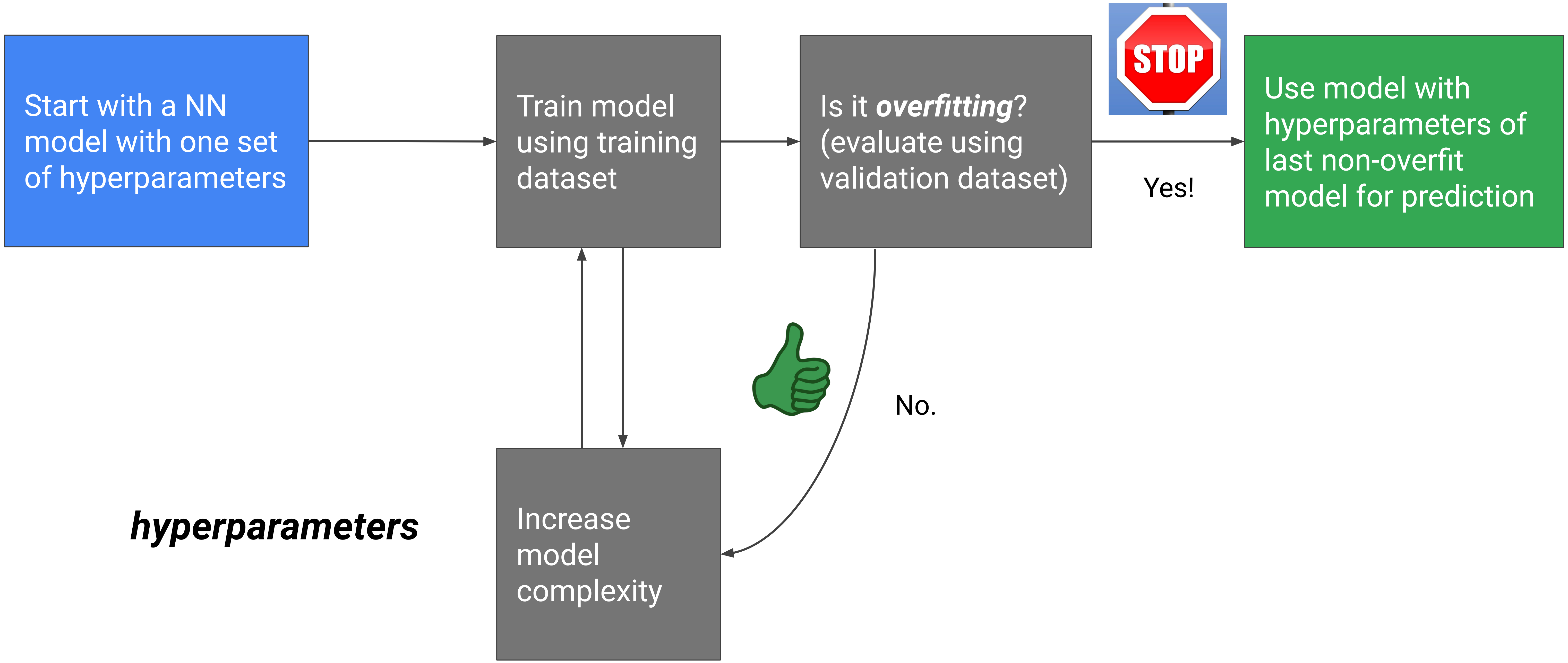
FYI: For your knowledge

Bias VS Variance… Trade off



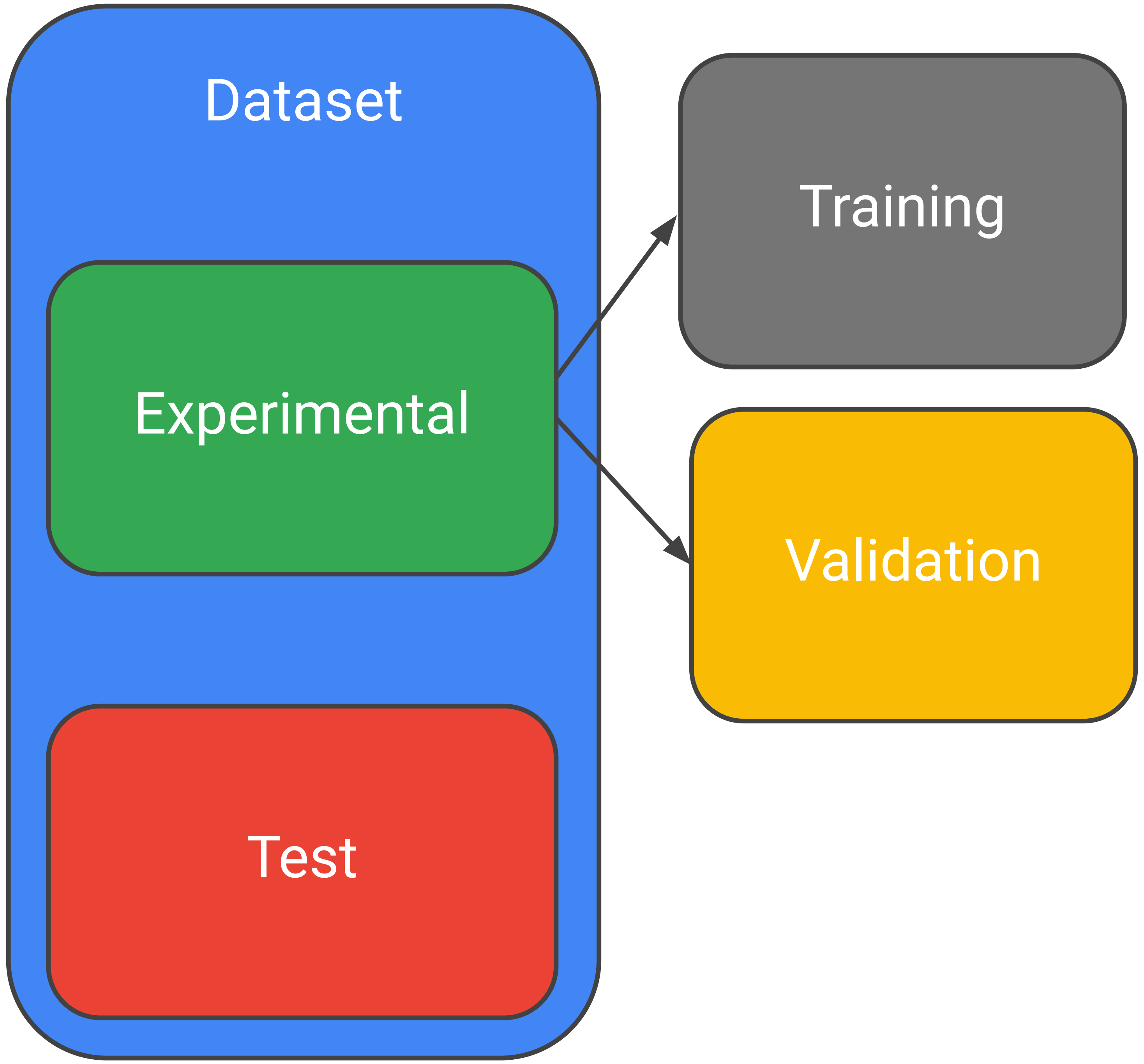
Needed For the Exam

You can use the validation dataset to experiment with model complexity



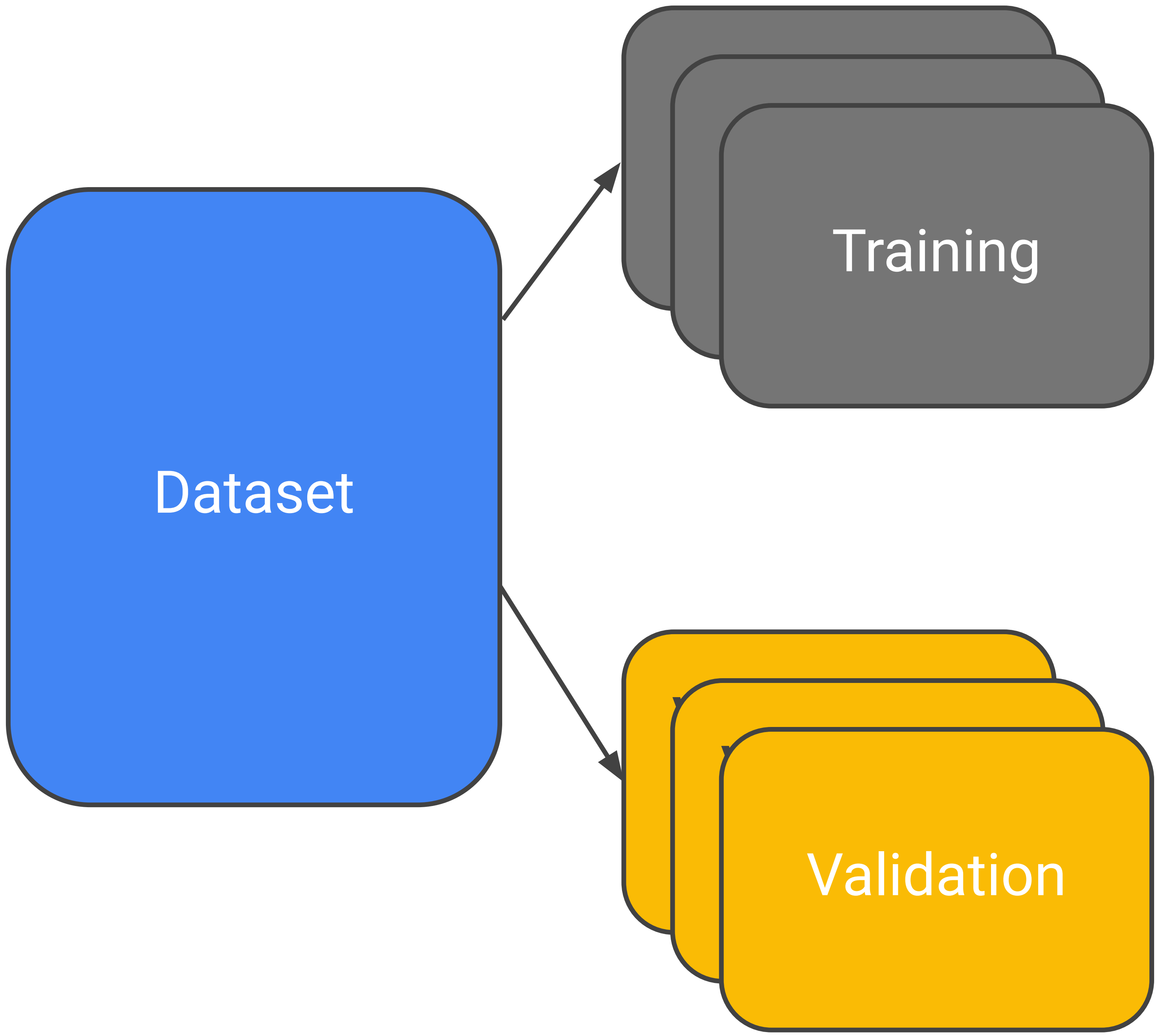
Needed For the Exam

Evaluate the final model with independent test data



Needed For the Exam

Evaluate the final model with cross-validation



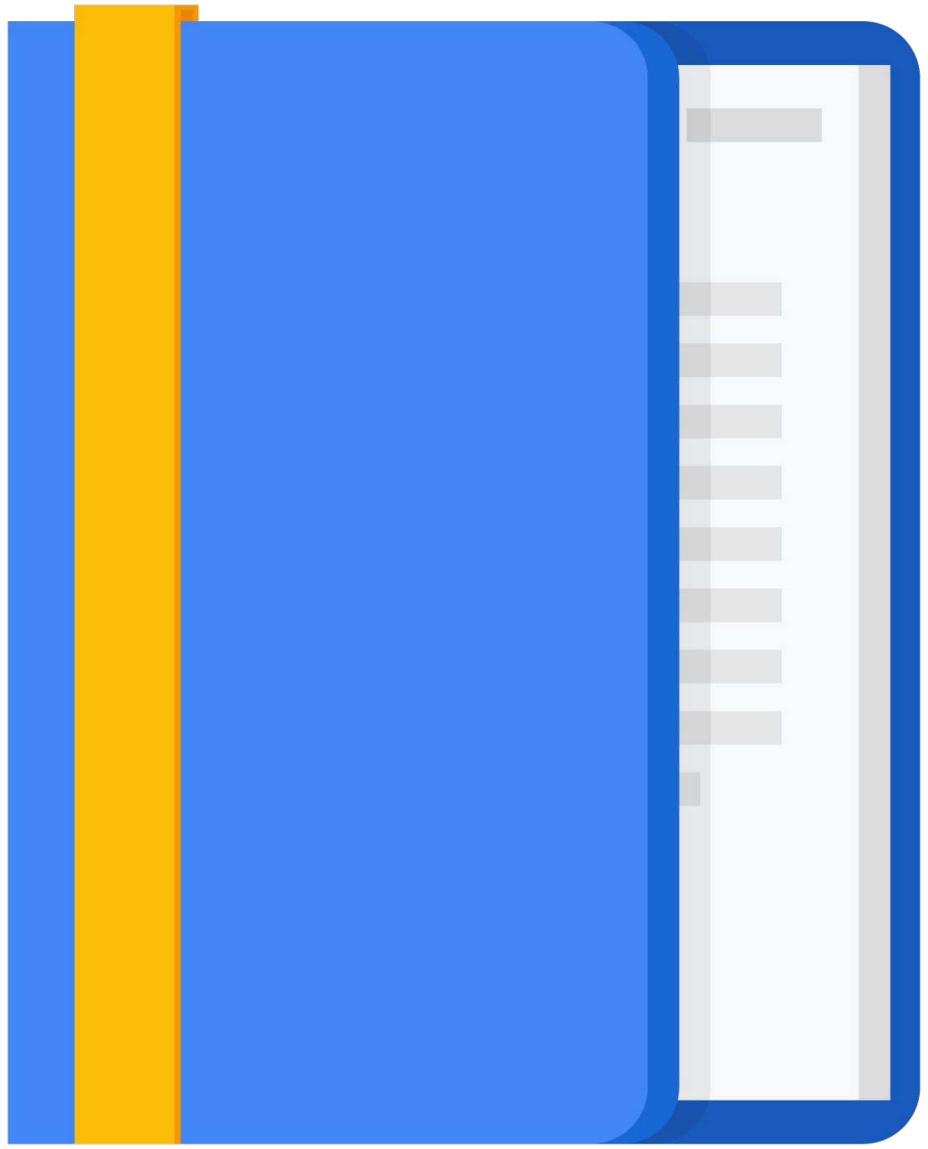
Section Agenda

Generalization in ML

Sampling

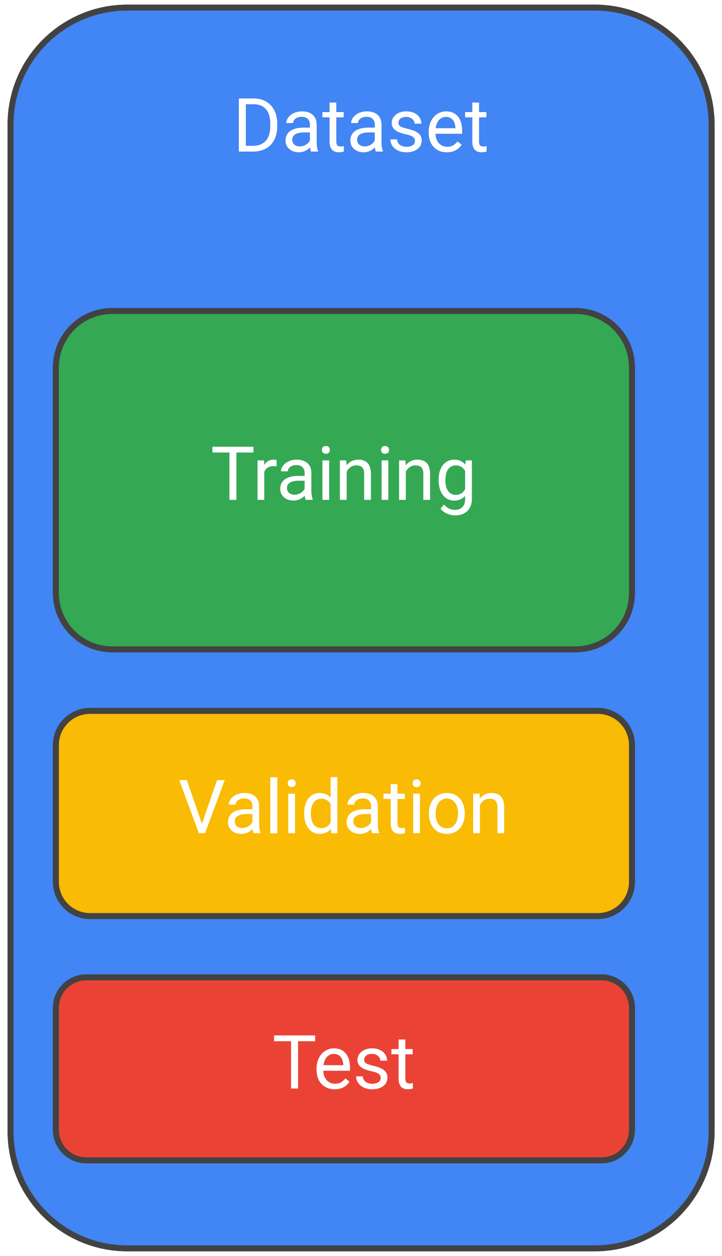
Regularizations

ML Model Performance Evaluation



FYI: For your knowledge

We often have large datasets in BigQuery that we want to use for machine learning



Row	date	airline	departure_airport	departure_schedule	arrival_airport	arrival_delay
1	2004-08-07	TZ	SRQ	1255	IND	-14.0
2	2004-03-05	TZ	SRQ	2117	IND	-9.0
3	2004-04-12	TZ	SRQ	2000	IND	-17.0
4	2003-04-16	TZ	SRQ	1215	IND	-5.0
5	2005-03-20	TZ	SRQ	645	IND	14.0
6	2003-04-06	TZ	SRQ	1235	IND	-8.0

FYI: For your knowledge

It's easy to get a random 80% of your dataset for training

```
#standardSQL
SELECT
  date,
  airline,
  departure_airport,
  departure_schedule,
  arrival_airport,
  arrival_delay
FROM
  `bigquery-samples.airline_ontime_data.flights`
WHERE
  RAND() < 0.8
```

RAND will return a number between 0 and 1.

However, experimentation requires repeatability

You need to know which specific data was involved in training, validation, and testing.

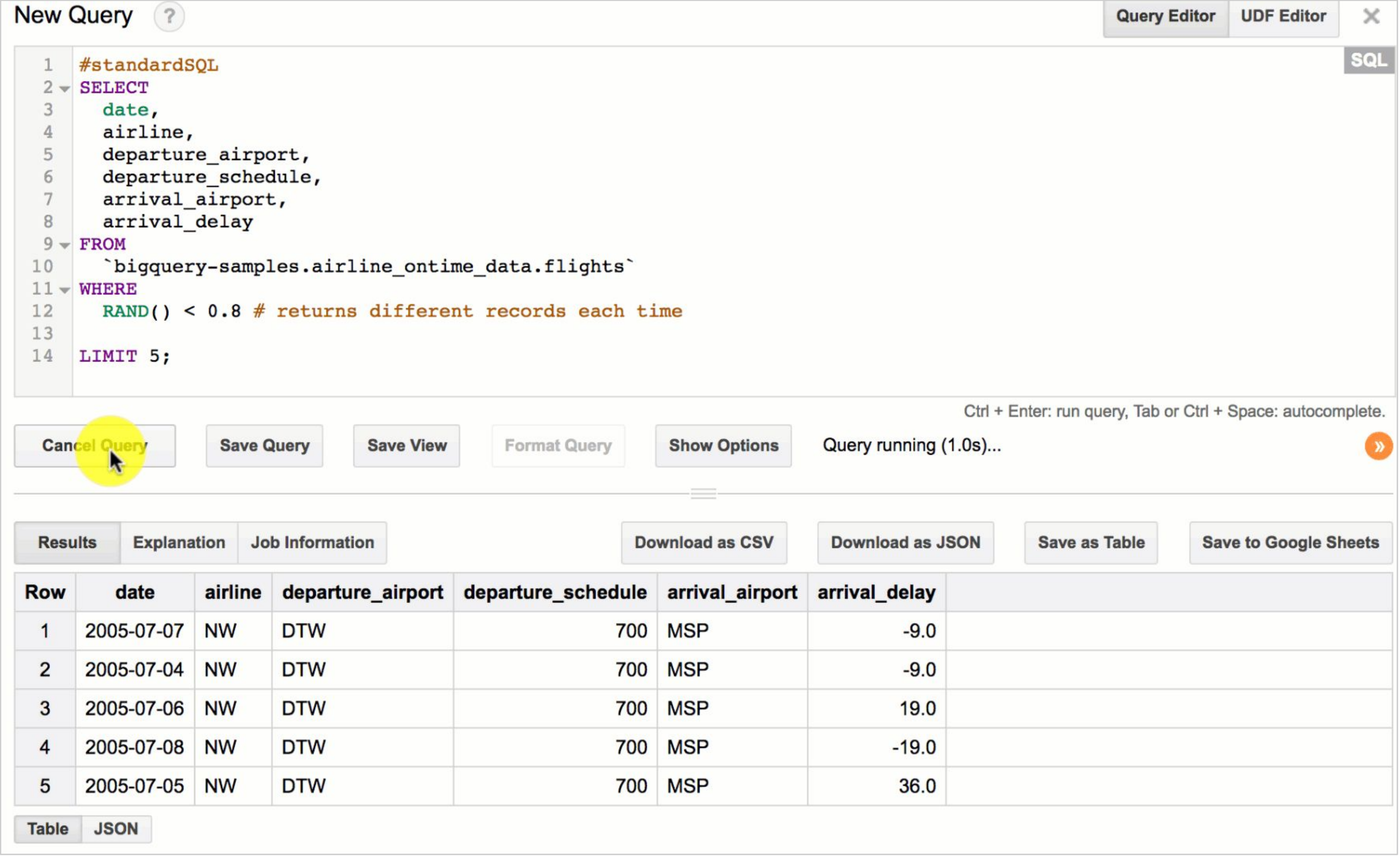
FYI: For your knowledge

Naive random splitting is not repeatable

Order of rows in BigQuery is not certain without ORDER BY.

Hard to identify and split the remaining 20% of data for validation and testing.

RAND() will return different results each time →



FYI: For your knowledge

Solution: Split a dataset into training/validation/test using the hashing and modulo operators

```
#standardSQL
SELECT
  date,
  airline,
  departure_airport,
  departure_schedule,
  arrival_airport,
  arrival_delay
FROM
 `bigquery-samples.airline_ontime_data.flights`

WHERE
  MOD(ABS(FARM_FINGERPRINT(date)),10) < 8
```

Note: Even though we select date, our model wouldn't actually use it during training.

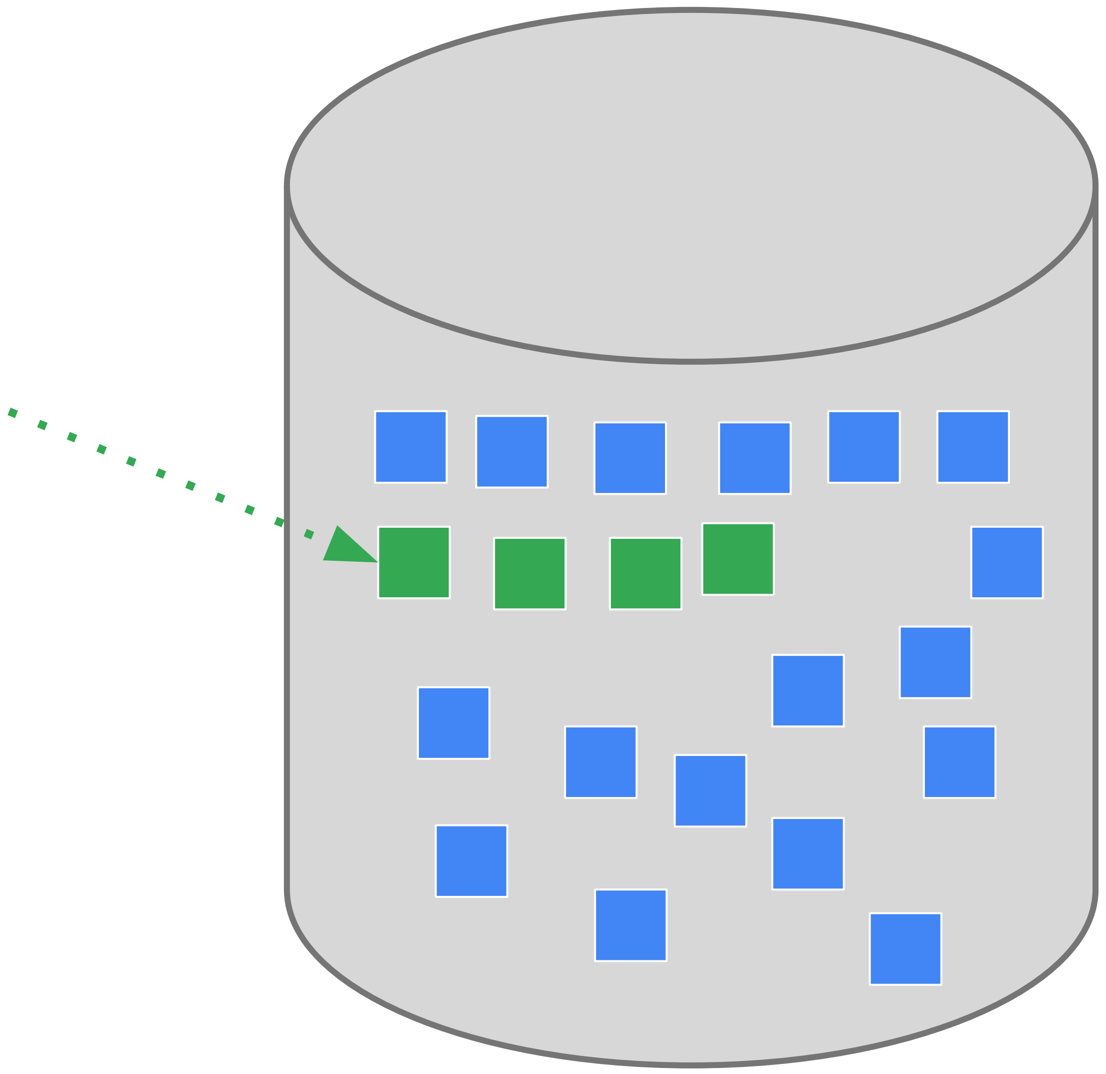
Hash value on the Date will always return the same value.

Then we can use a modulo operator to only pull 80% of that data based on the last few hash digits.

FYI: For your knowledge

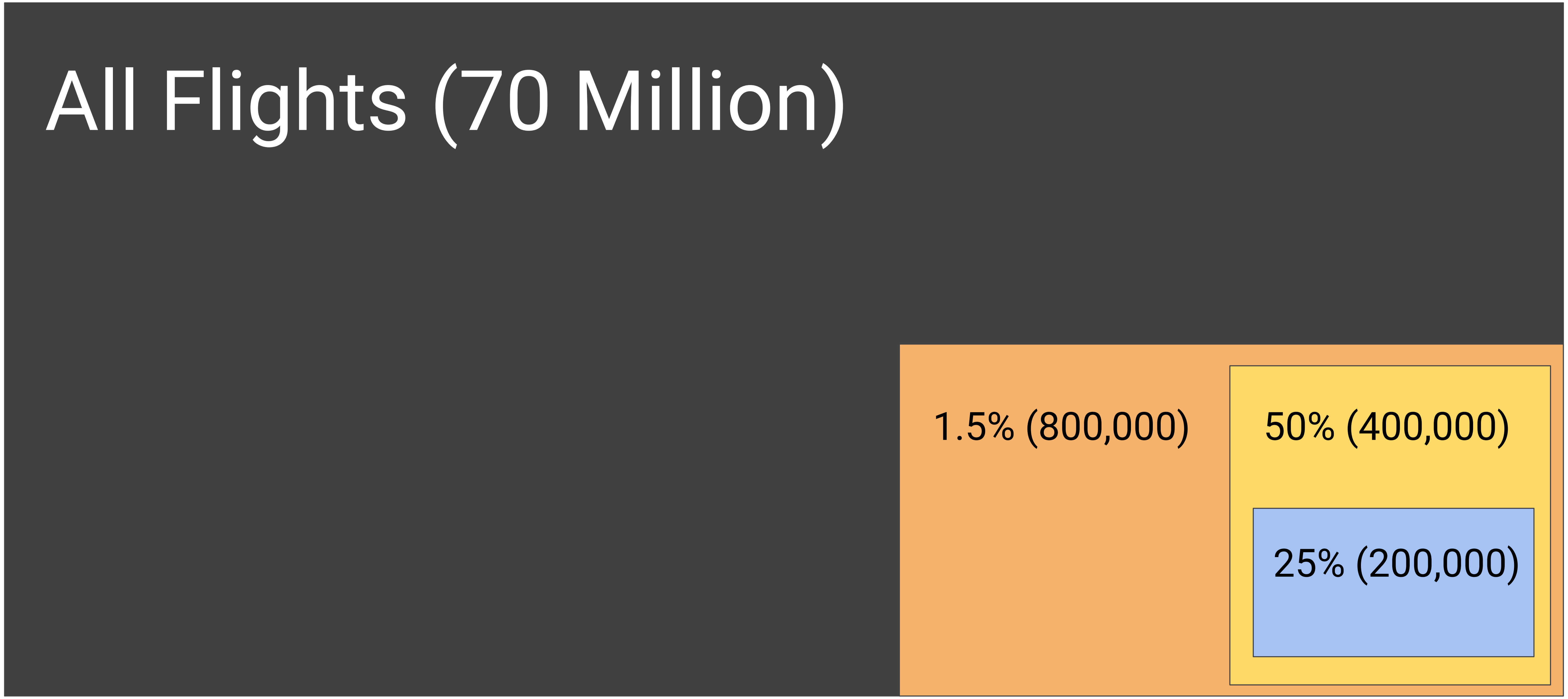
Developing the ML model software on the entire dataset can be expensive; you want to to develop on a smaller sample

Develop your TensorFlow code on a small subset of data, then scale it out to the cloud.



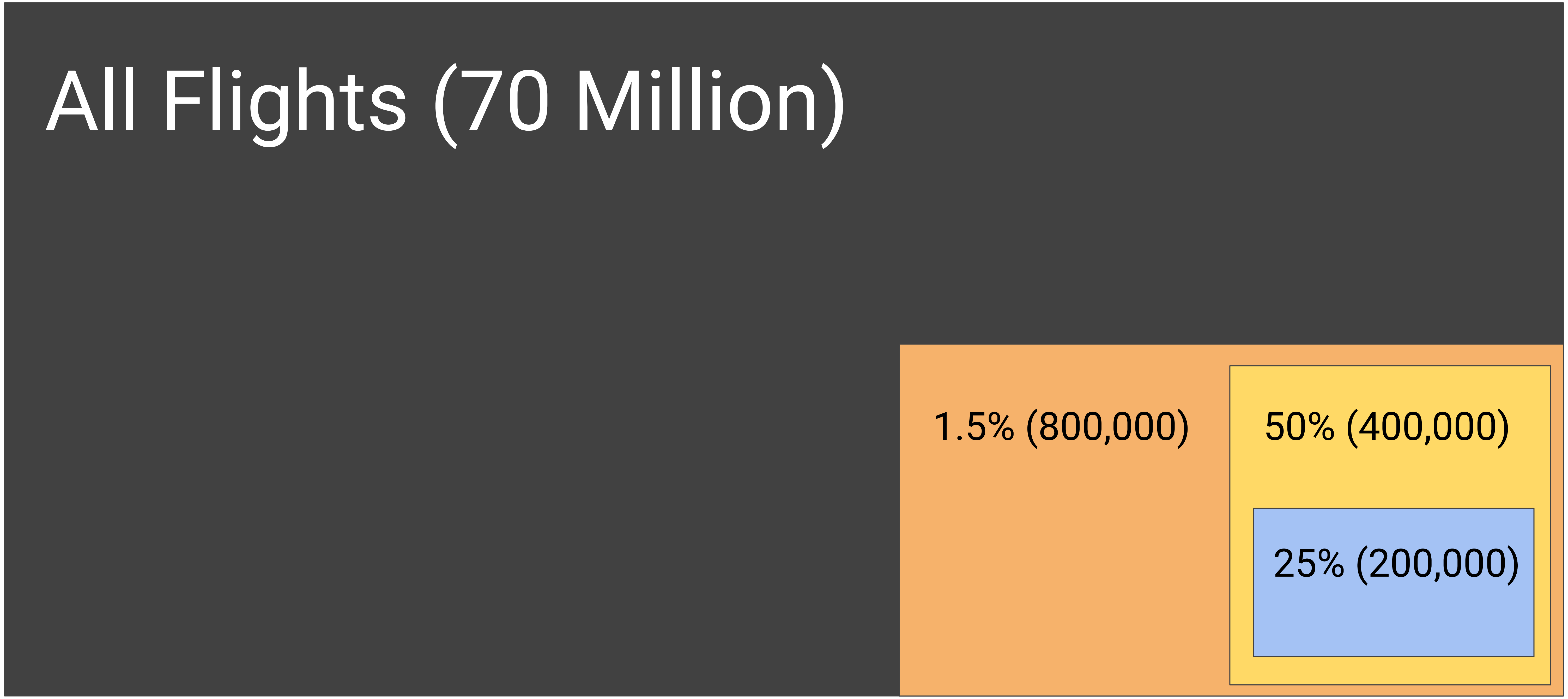
FYI: For your knowledge

How we want to split our data



FYI: For your knowledge

How we want to split our data



We can extend this to creating 3 splits

```
#standardSQL
SELECT
  date,
  airline,
  departure_airport,
  departure_schedule,
  arrival_airport,
  arrival_delay
FROM
 `bigquery-samples.airline_ontime_data.flights`
WHERE
  MOD(ABS(FARM_FINGERPRINT(date)),70) = 0
     AND
  MOD(ABS(FARM_FINGERPRINT(date)),700) >= 350
    AND
  MOD(ABS(FARM_FINGERPRINT(date)),700) < 525
```

Then take 1 in 70 flights.

Ignore the 50% of the dataset (training).

Choose data between 350 and 524 which is a new 25% sample for Validation.

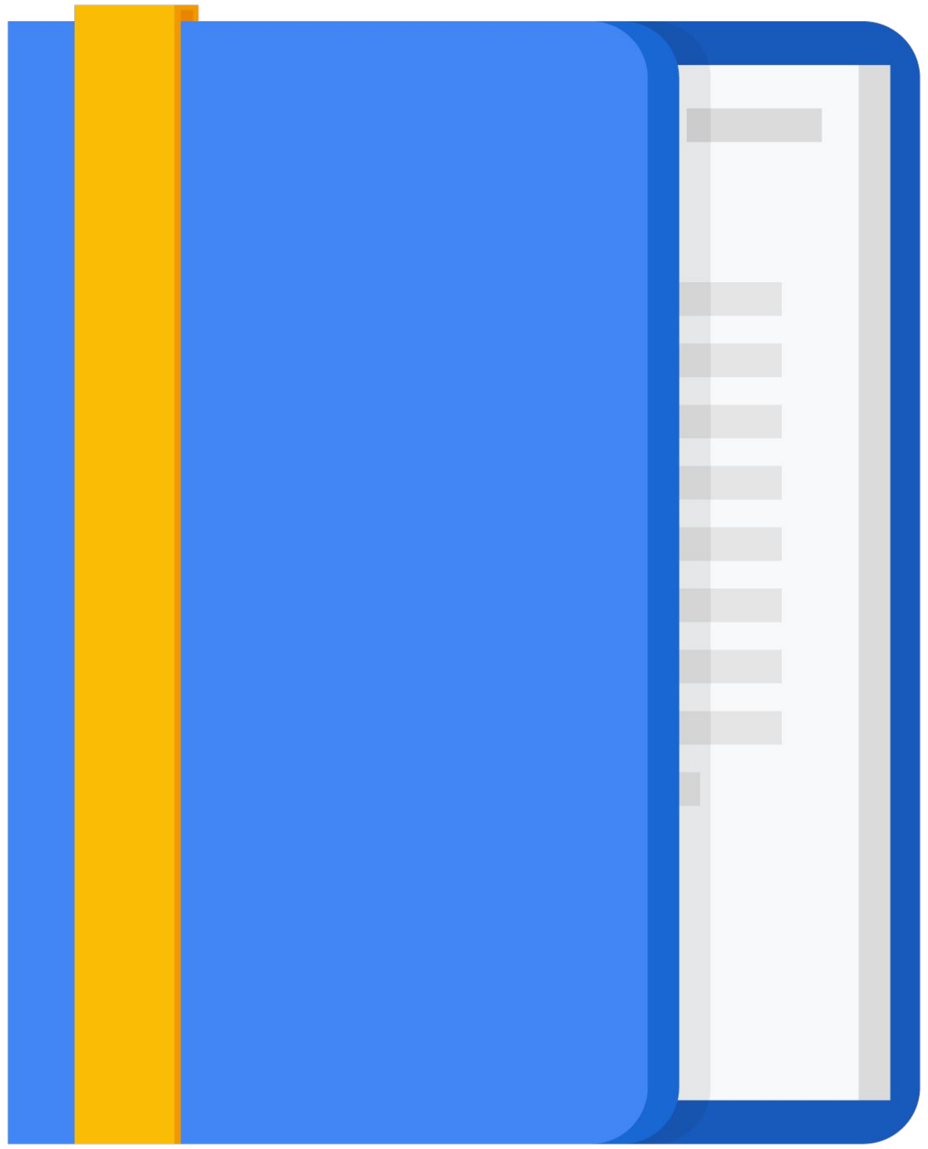
Course Agenda

Generalization in ML

Sampling

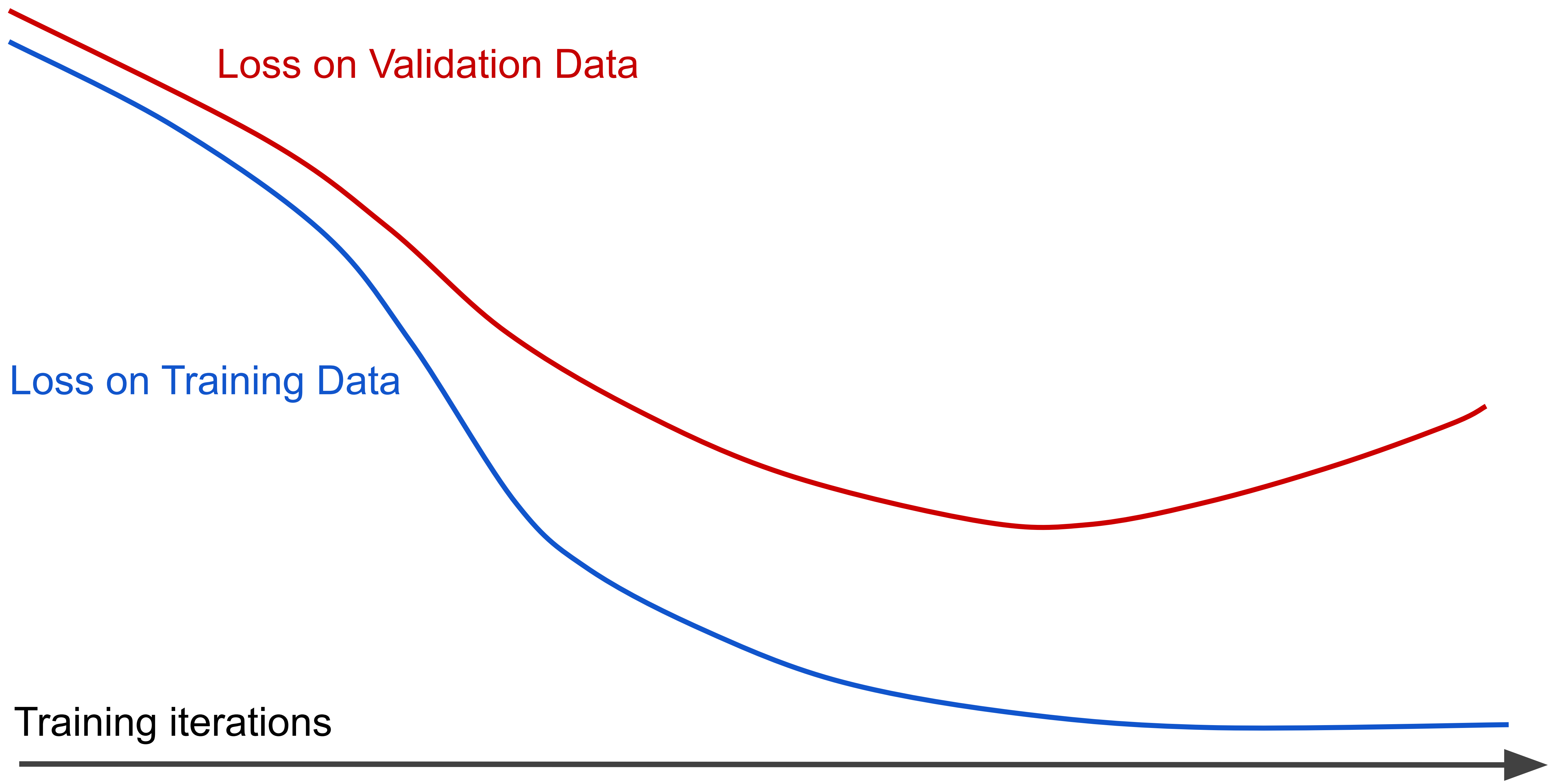
Regularizations

ML Model Performance Evaluation



Needed For the Exam

What is happening here? How can we address this?



REGULARIZATION!

Needed For the Exam

The simpler the better

Regularization is used when…

- You have too many features in your model
- You don't know which feature should be discarded
- You want to reduce overfitting

Don't cook with every spice in the spice rack!

Needed For the Exam

Occam's razor

When presented with competing hypothetical answers to a problem, one should select the one that makes the fewest assumptions. The idea is attributed to William of Ockham (c. 1287–1347).

source:https://en.wikipedia.org/wiki/Occam%27s_razor

FYI: For your knowledge

Factor in model complexity when calculating error

Minimize: *loss*(Data|Model) + complexity(Model)

aim for low training error

...but balance against complexity

Optimal model complexity is data-dependent, so requires hyperparameter tuning.

Needed For the Exam

Regularization is a major field of ML research

Early Stopping

Parameter Norm Penalties

- L1 regularization
- L2 regularization
- Max-norm regularization

Dataset Augmentation

Noise Robustness

Sparse Representations

. . .

We will look into these methods.

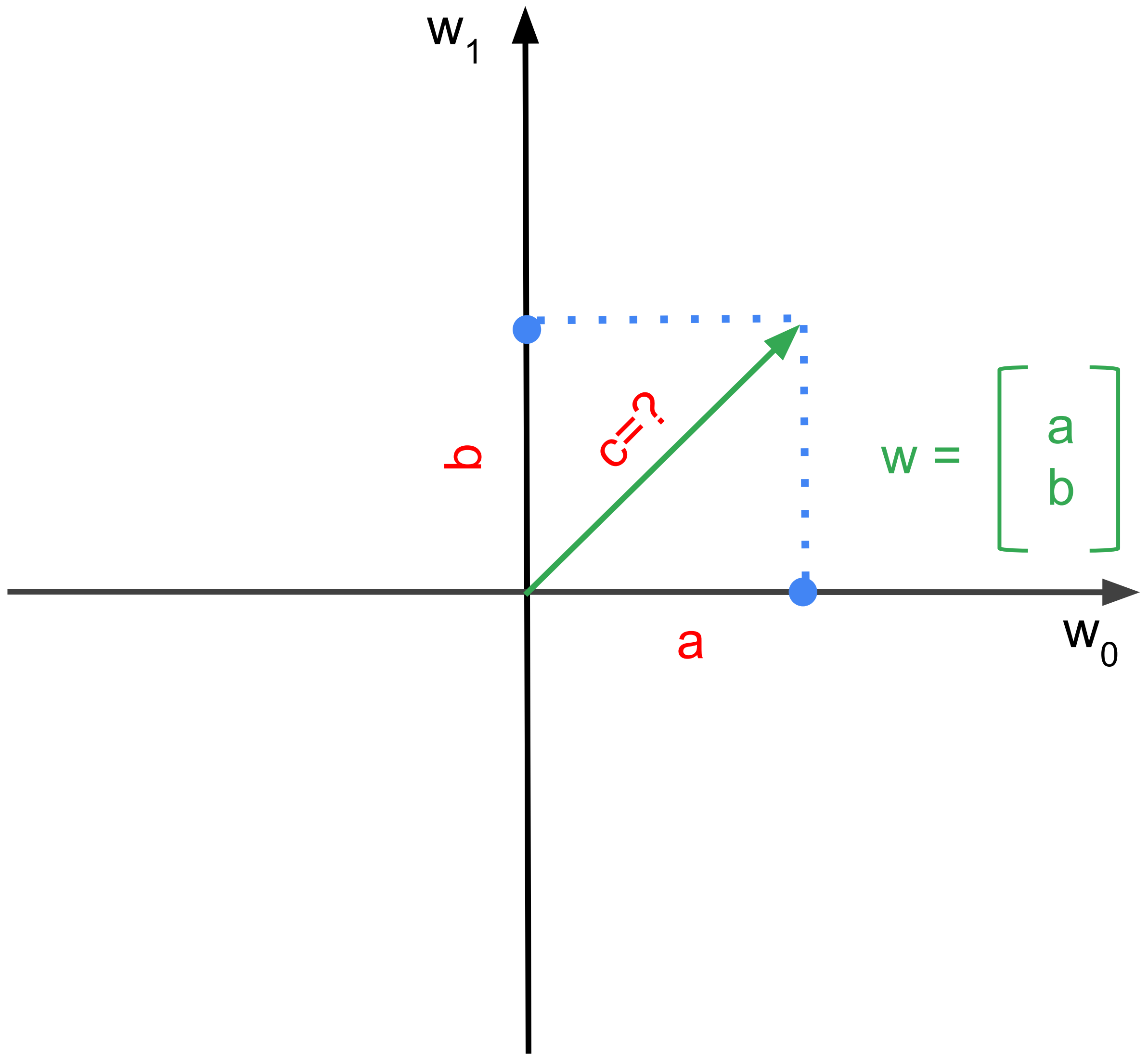
FYI: For your knowledge

How can we measure model complexity?

FYI: For your knowledge

L2 vs. L1 Norm

$$w = [w_0, w_1, \dots, w_n]^T$$



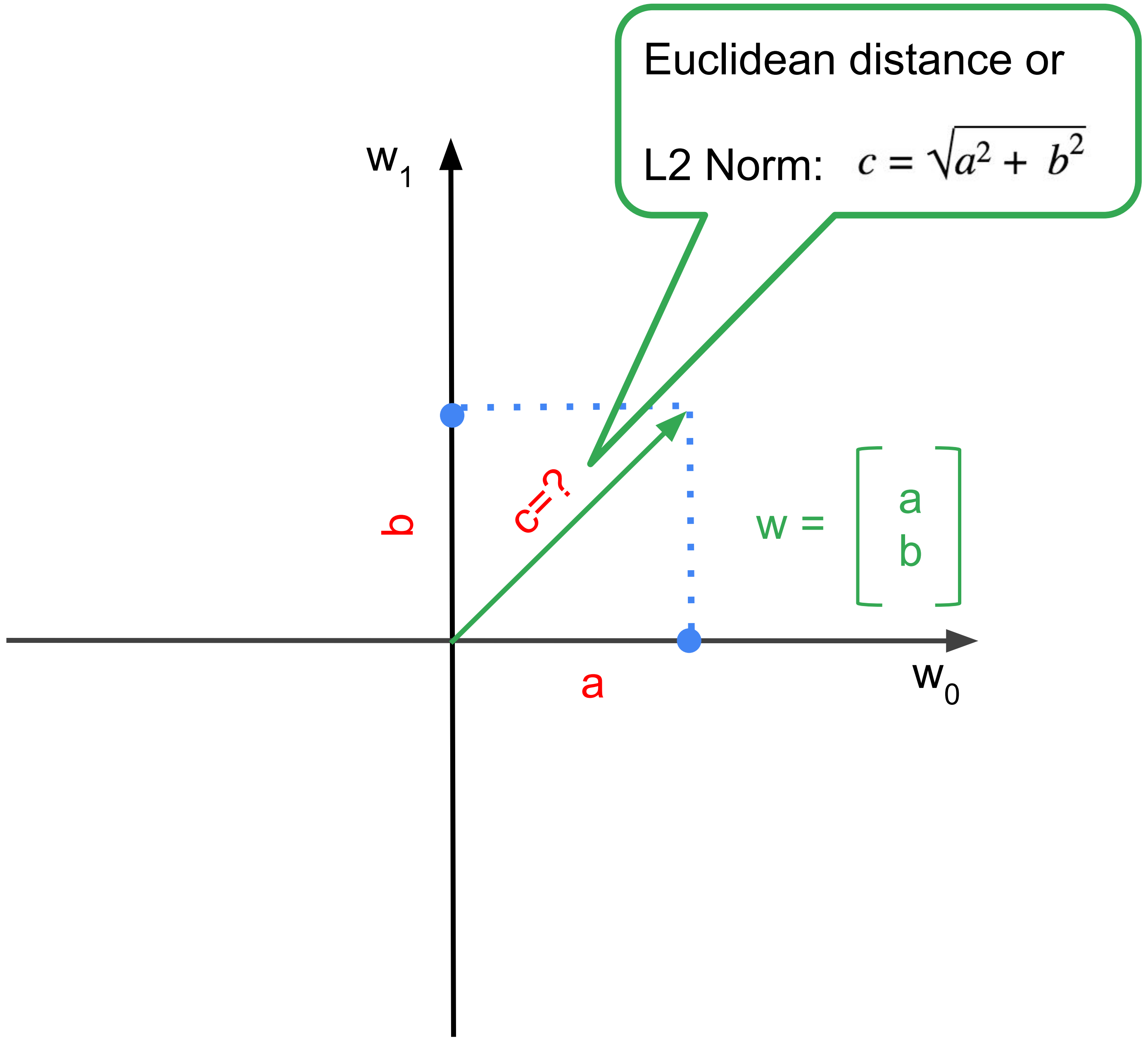
FYI: For your knowledge

L2 vs. L1 Norm

$$w = [w_0, w_1, \dots, w_n]^T$$

$$\|w\|_2 = (w_0^2 + w_1^2 + \dots + w_n^2)^{1/2}$$

$$\|w\|_1 = (|w_0| + |w_1| + \dots + |w_n|)$$



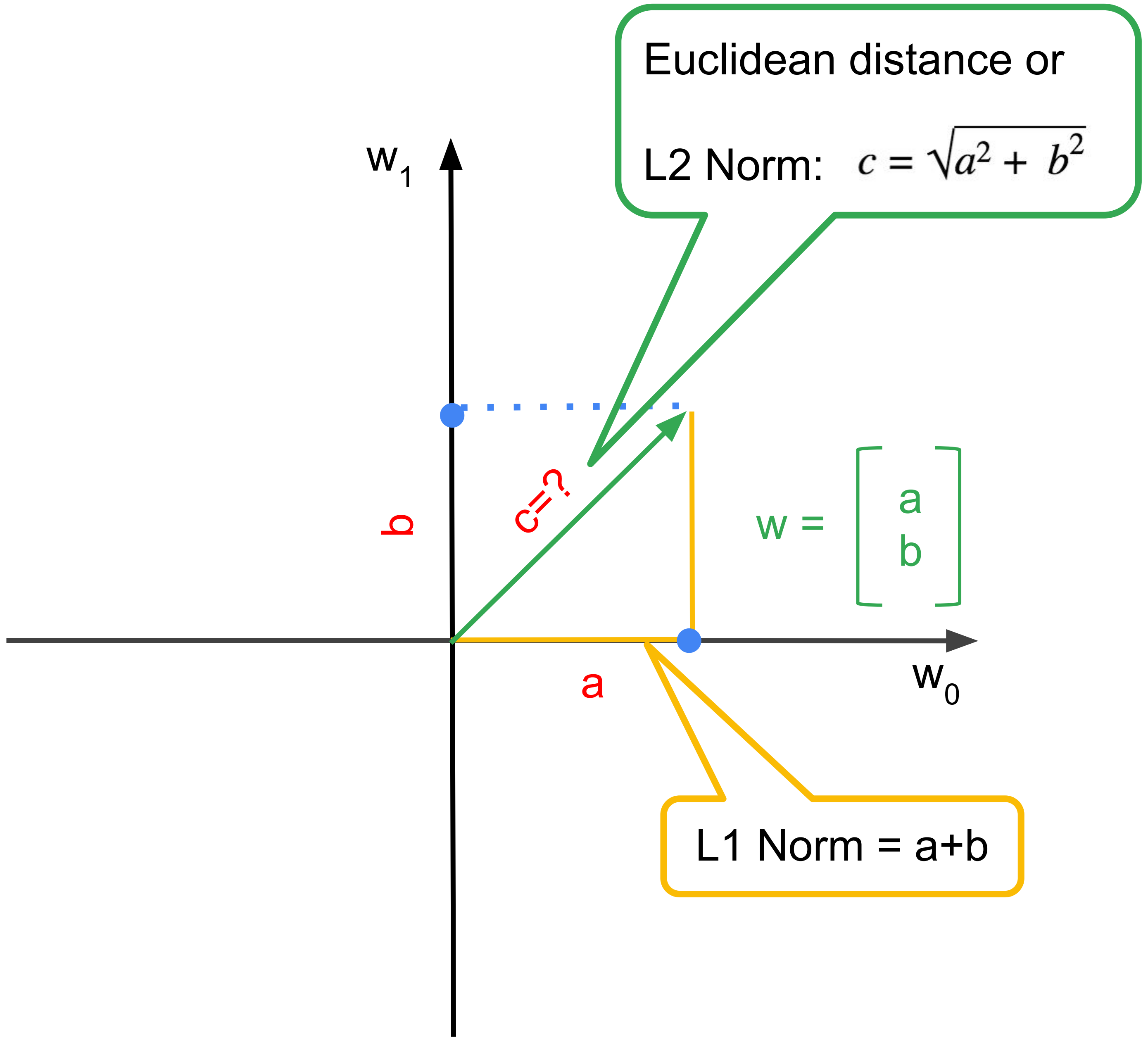
FYI: For your knowledge

L2 vs. L1 Norm

$$w = [w_0, w_1, \ldots, w_n]^T$$

L2 Norm

$$\|w\|_2 = (w_0{}^2 + w_1{}^2 + \ldots + w_n{}^2)^{1/2}$$



FYI: For your knowledge

L2 vs. L1 Norm

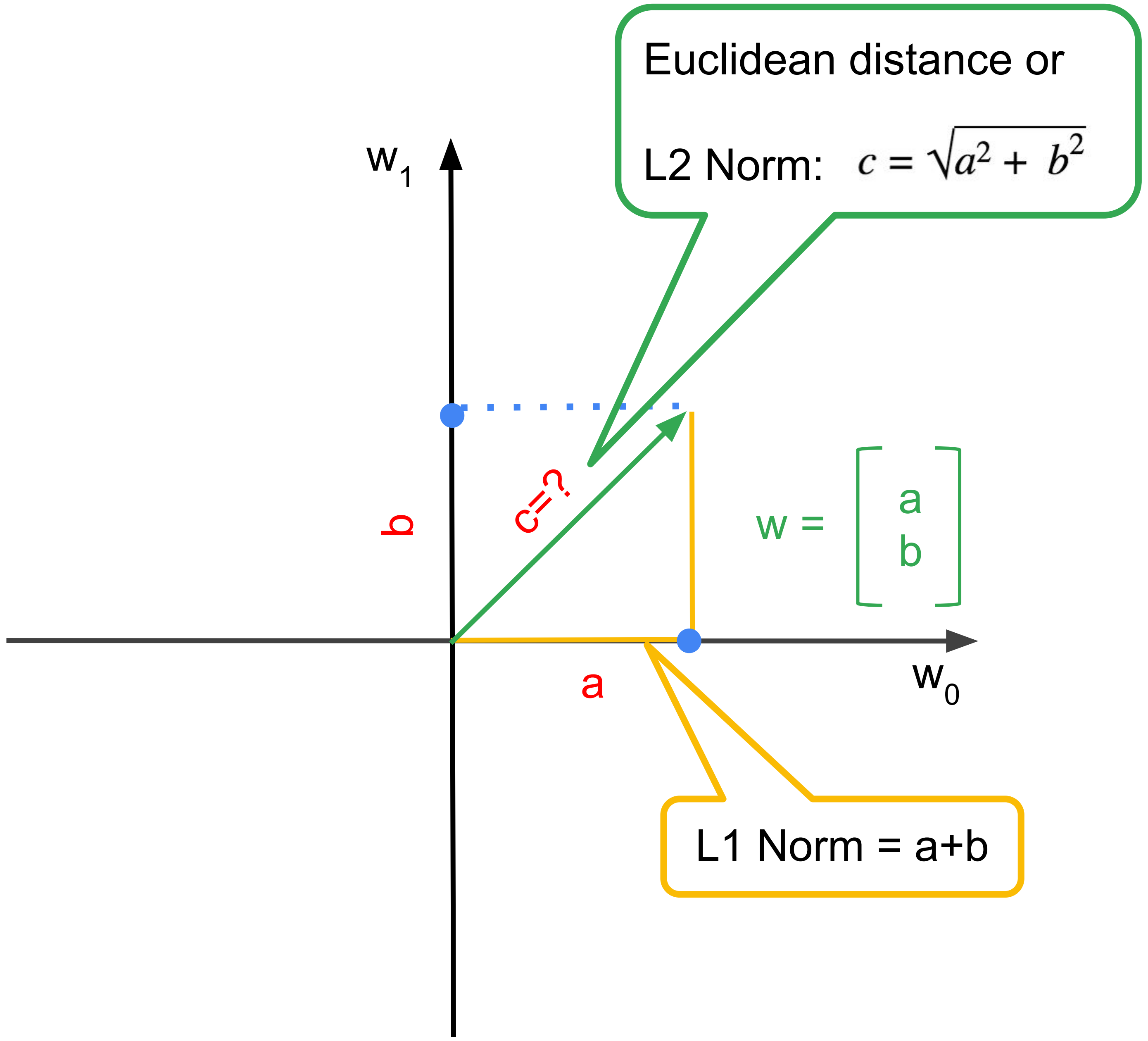
$$w = [w_0, w_1, \ldots, w_n]^T$$

L2 Norm

$$\|w\|_2 = (w_0^2 + w_1^2 + \ldots + w_n^2)^{1/2}$$

L1 Norm

$$\|w\|_1 = (|w_0| + |w_1| + \ldots + |w_n|)$$

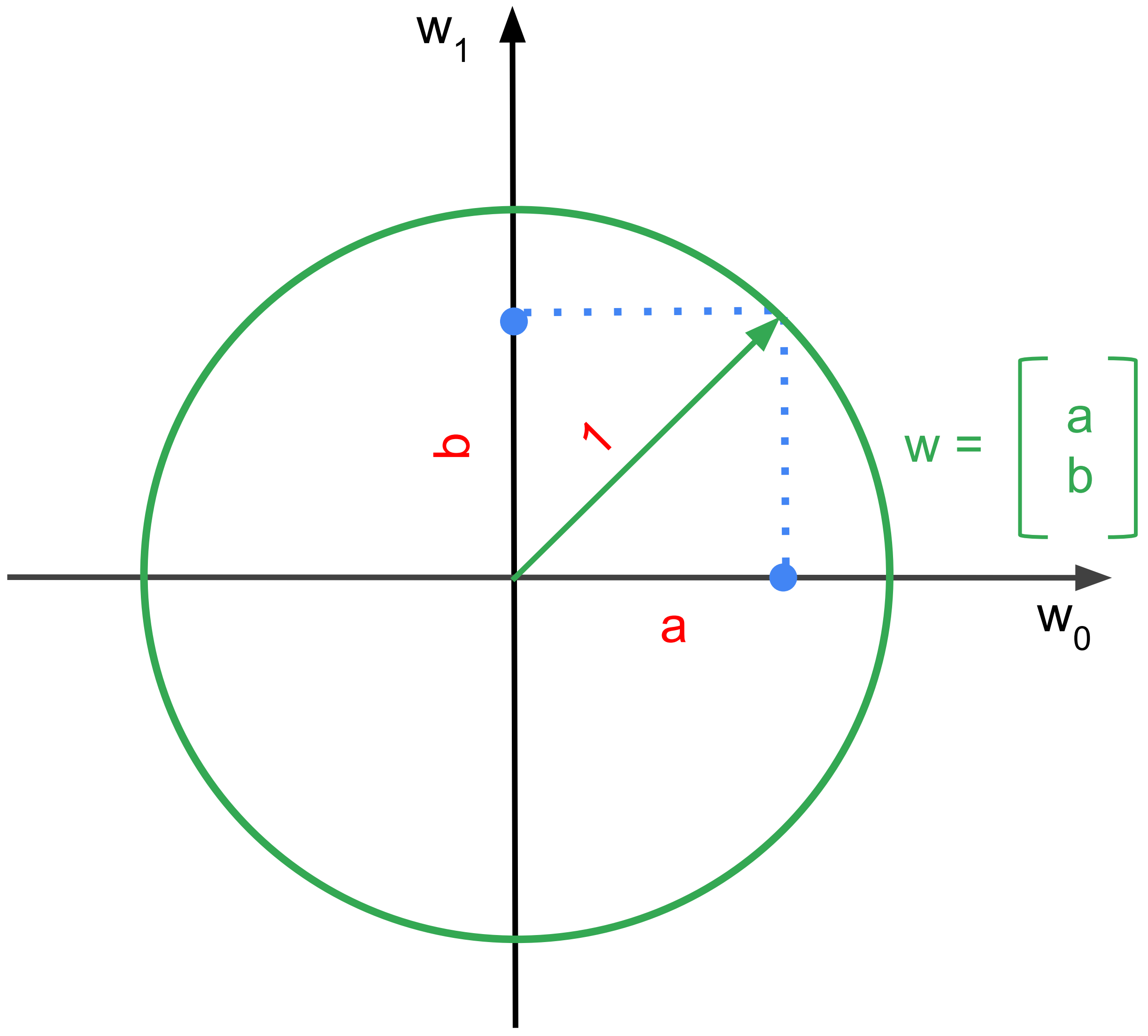


L2 vs. L1 Norm

$$w = [w_0, w_1, \dots, w_n]^T$$

L2 Norm

$$\|w\|_2 = (w_0^2 + w_1^2 + \dots + w_n^2)^{1/2}$$



FYI: For your knowledge

L2 vs. L1 Norm

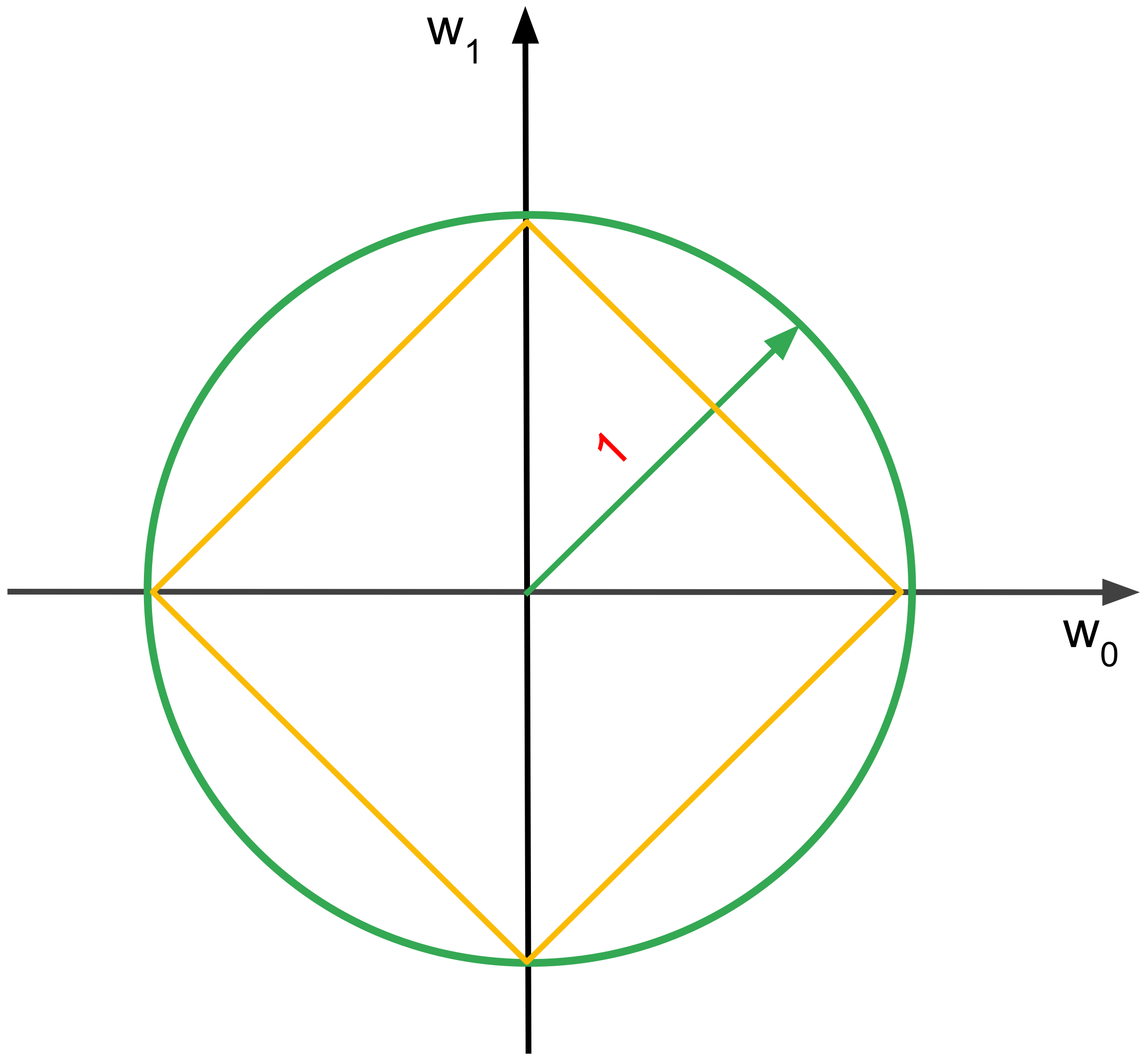
$$w = [w_0, w_1, \dots, w_n]^T$$

L2 Norm

$$\|w\|_2 = (w_0^2 + w_1^2 + \dots + w_n^2)^{1/2}$$

L1 Norm

$$\|w\|_1 = (|w_0| + |w_1| + \dots + |w_n|)$$



FYI: For your knowledge

In L2 regularization, complexity of model is defined by the L2 norm of the weight vector

Aim for low training error

...but balance against complexity

$$L(w, D) \; + \; \lambda \|w\|_2$$

Lambda controls how these are balanced

FYI: For your knowledge

In L1 regularization, complexity of model is defined by the L1 norm of the weight vector

$$L(w, D) + \lambda \boxed{\|w\|_1}$$

L1 regularization can be used as a feature selection mechanism.

Needed For the Exam

Logistic Regression Regularization Quiz

Why is it important to add regularization to logistic regression?

A. Helps stops weights being driven to +/- infinity.
B. Helps logits stay away from asymptotes which can halt training
C. Transforms outputs into a calibrated probability estimate
D. Both A & B
E. Both A & C

Needed For the Exam

Logistic Regression Regularization Quiz

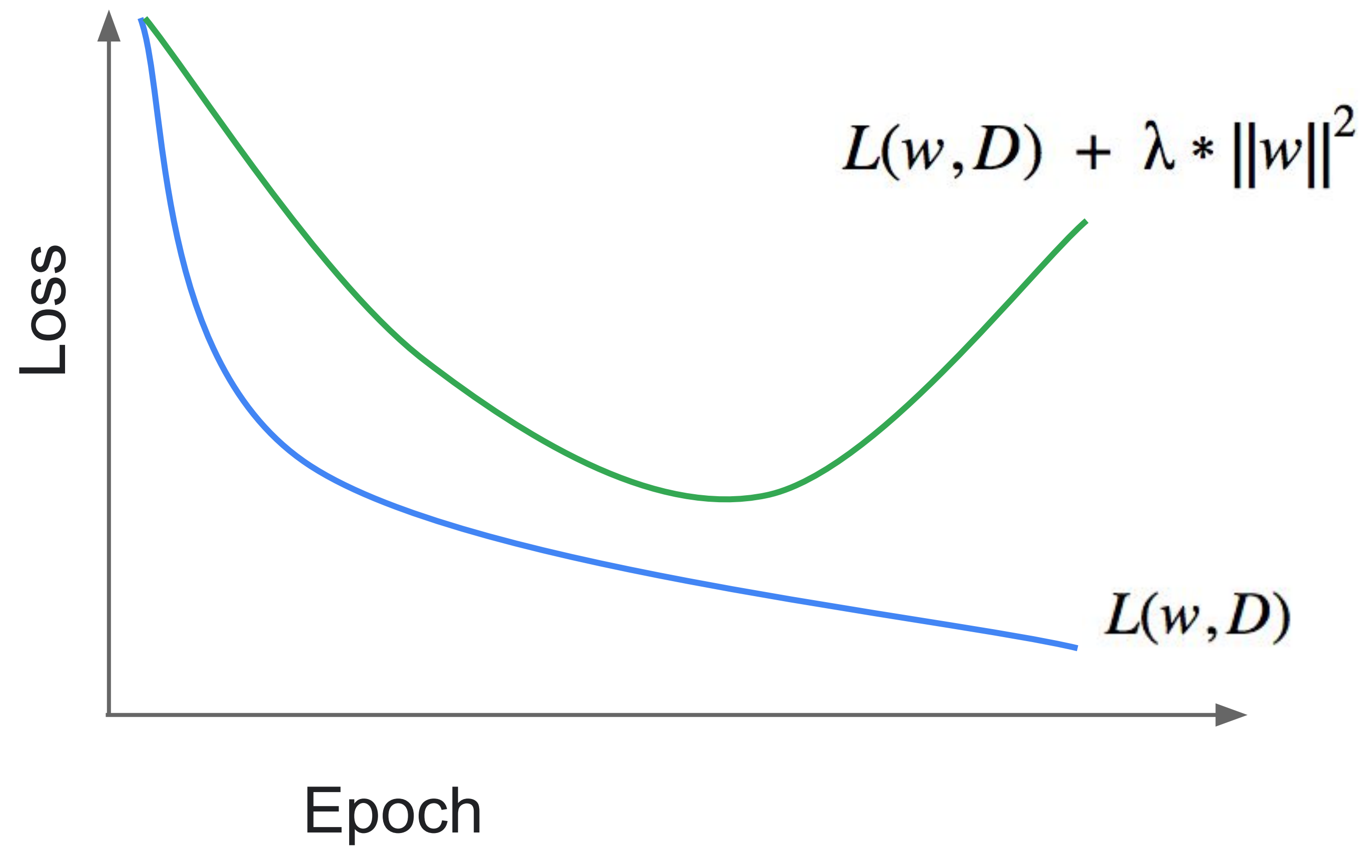
Why is it important to add regularization to logistic regression?

A. Helps stops weights being driven to +/- infinity.
B. Helps logits stay away from asymptotes which can halt training
C. Transforms outputs into a calibrated probability estimate
D. Both A & B
E. Both A & C

Needed For the Exam

Often, we do both L1, L2 regularization and early stopping to counteract overfitting

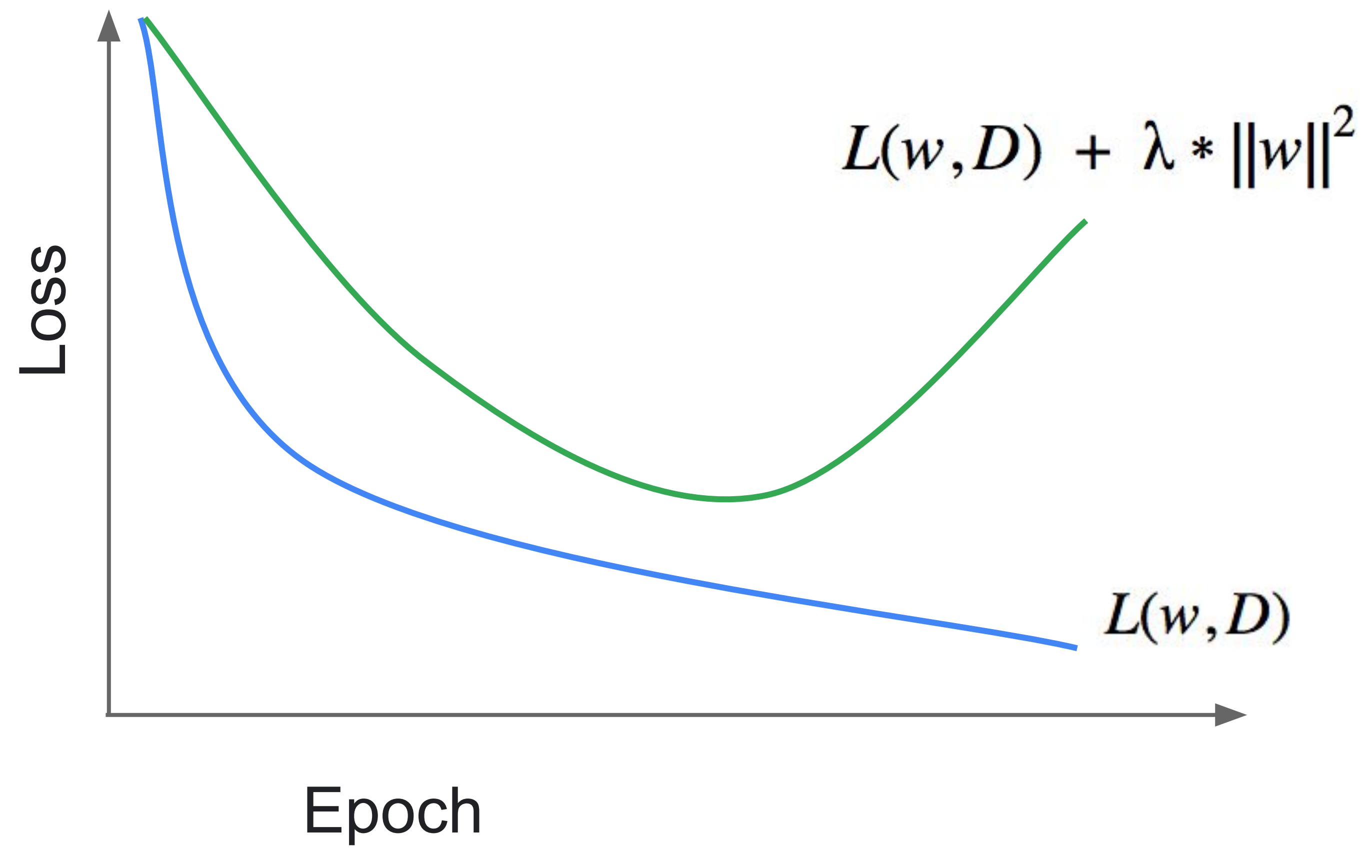
Regularization



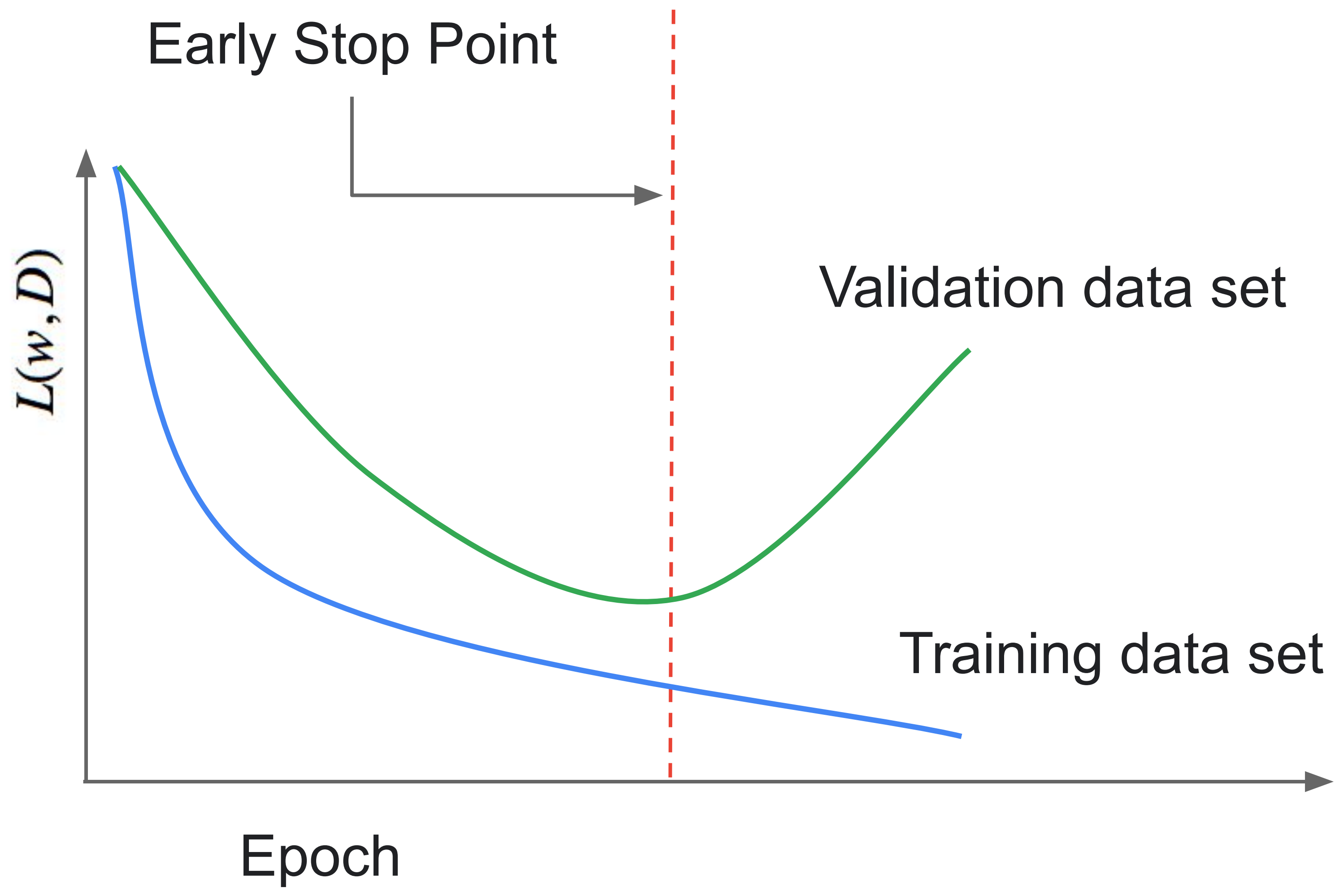
Needed For the Exam

Often, we do both regularization and early stopping to counteract overfitting

Regularization



Early stopping



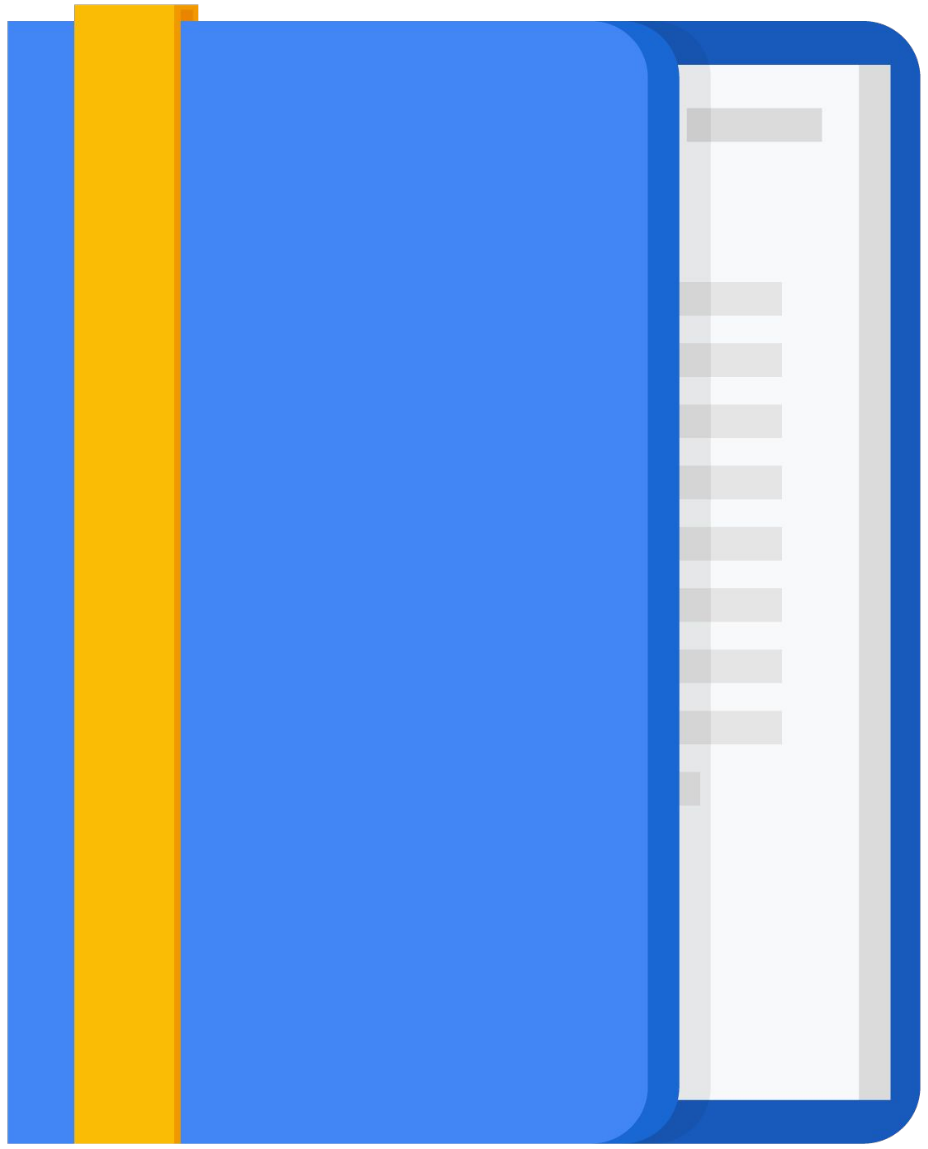
Section Agenda

Generalization in ML

Sampling

Regularizations

ML Model Performance Evaluation

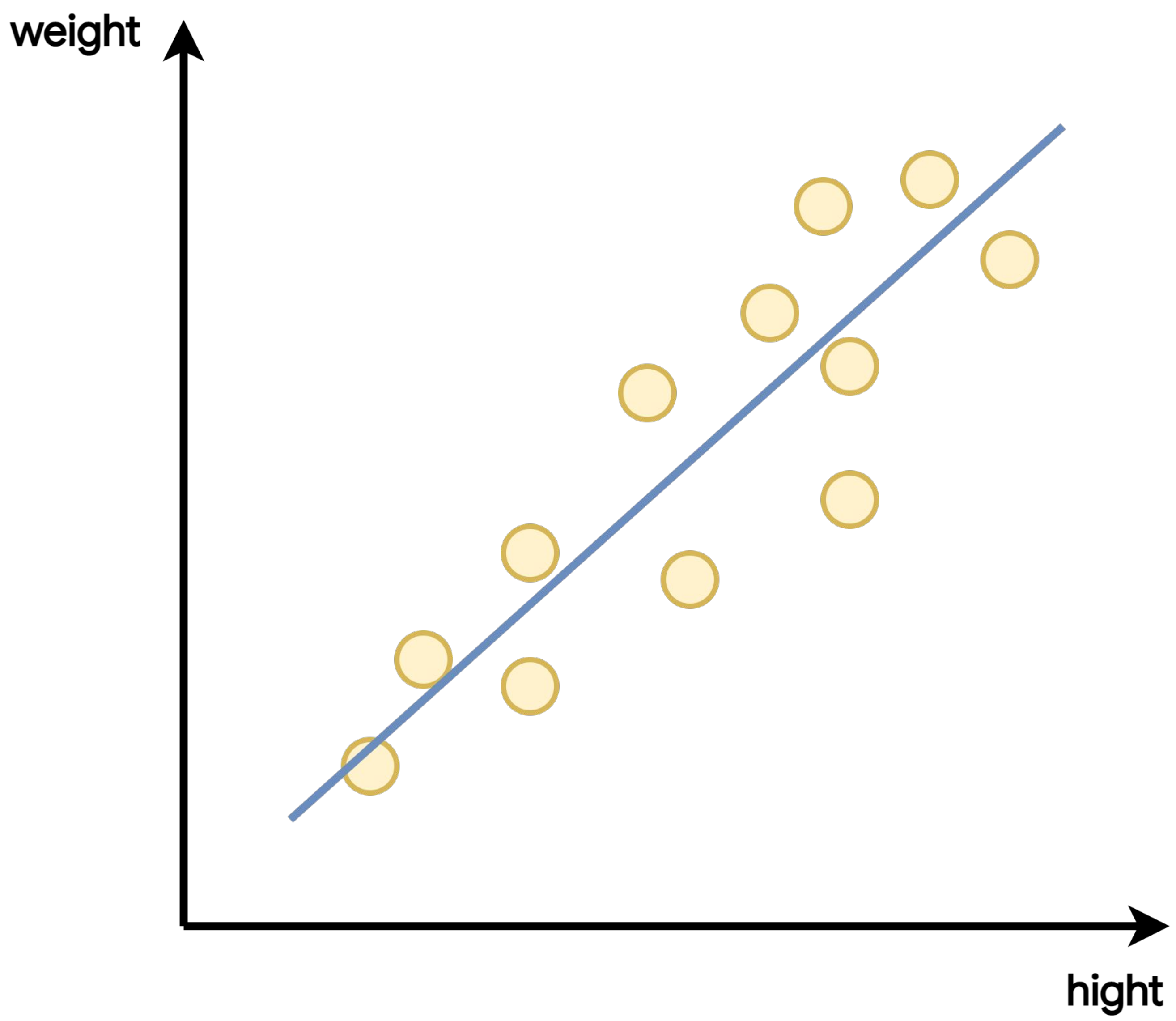


FYI: For your knowledge

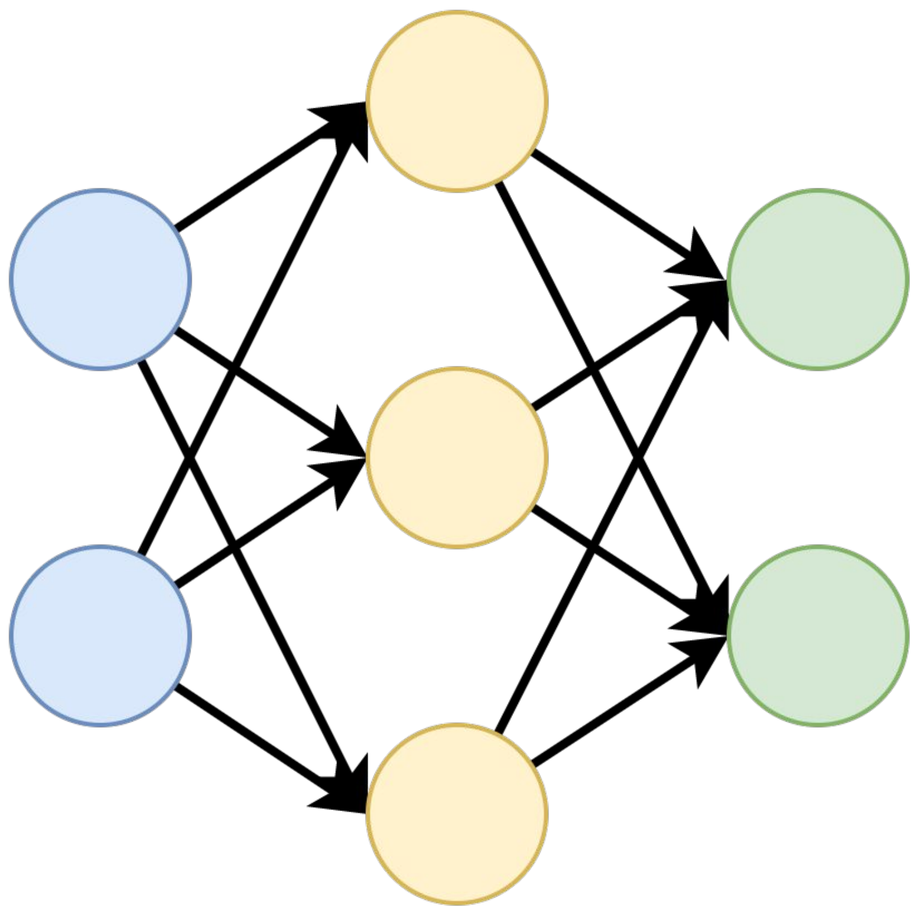
How do we compare algorithms?

FYI: For your knowledge

Is this

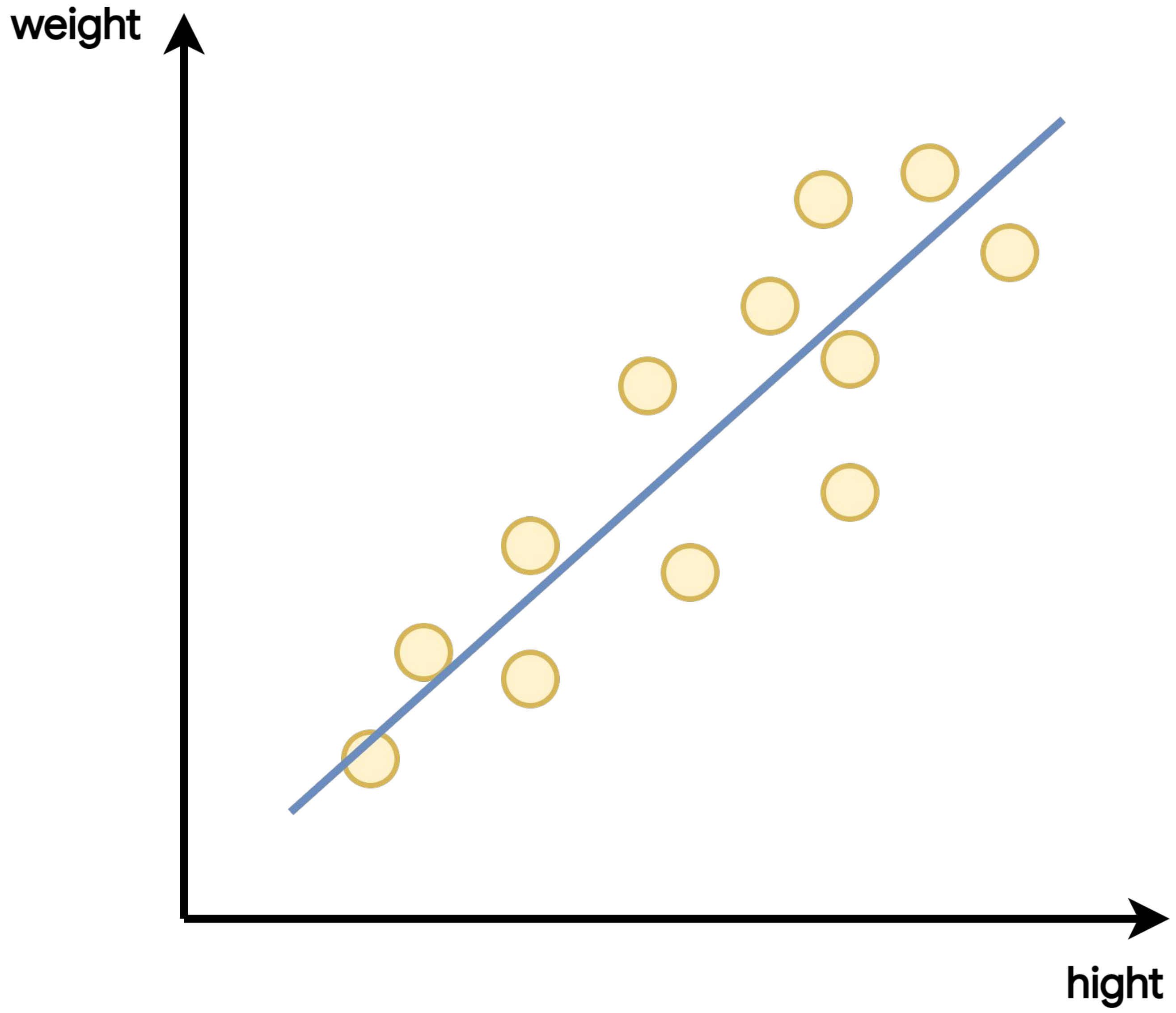


better than

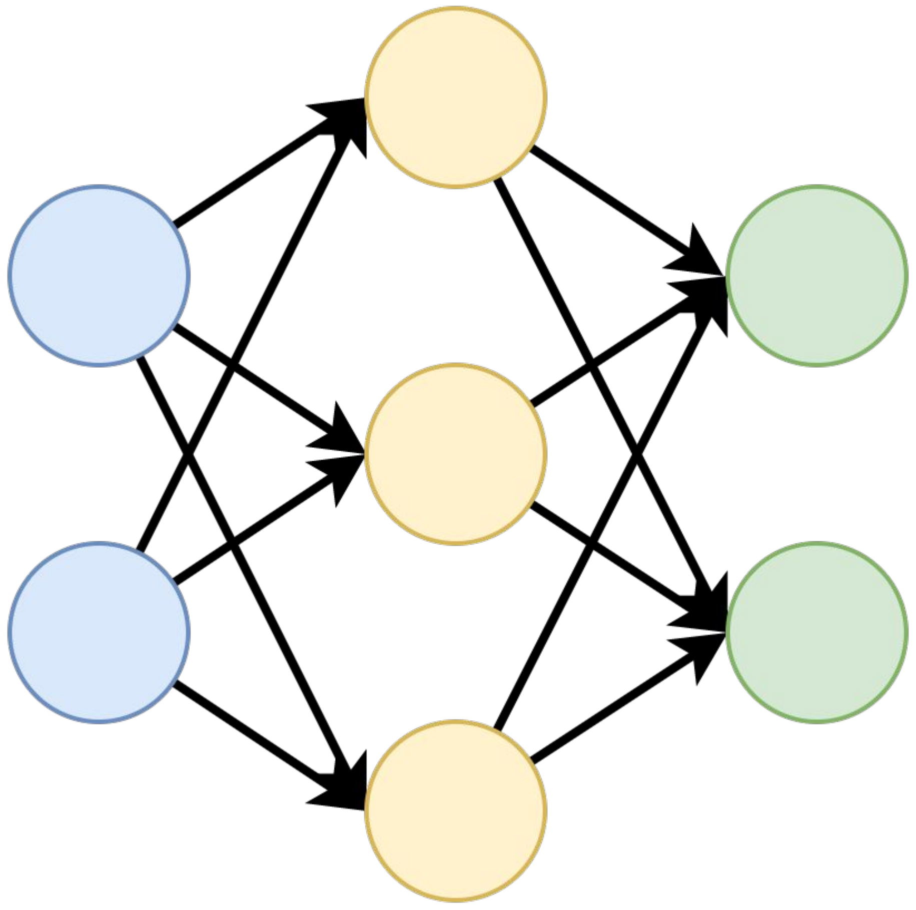


FYI: For your knowledge

Regression problems are easy



< = >



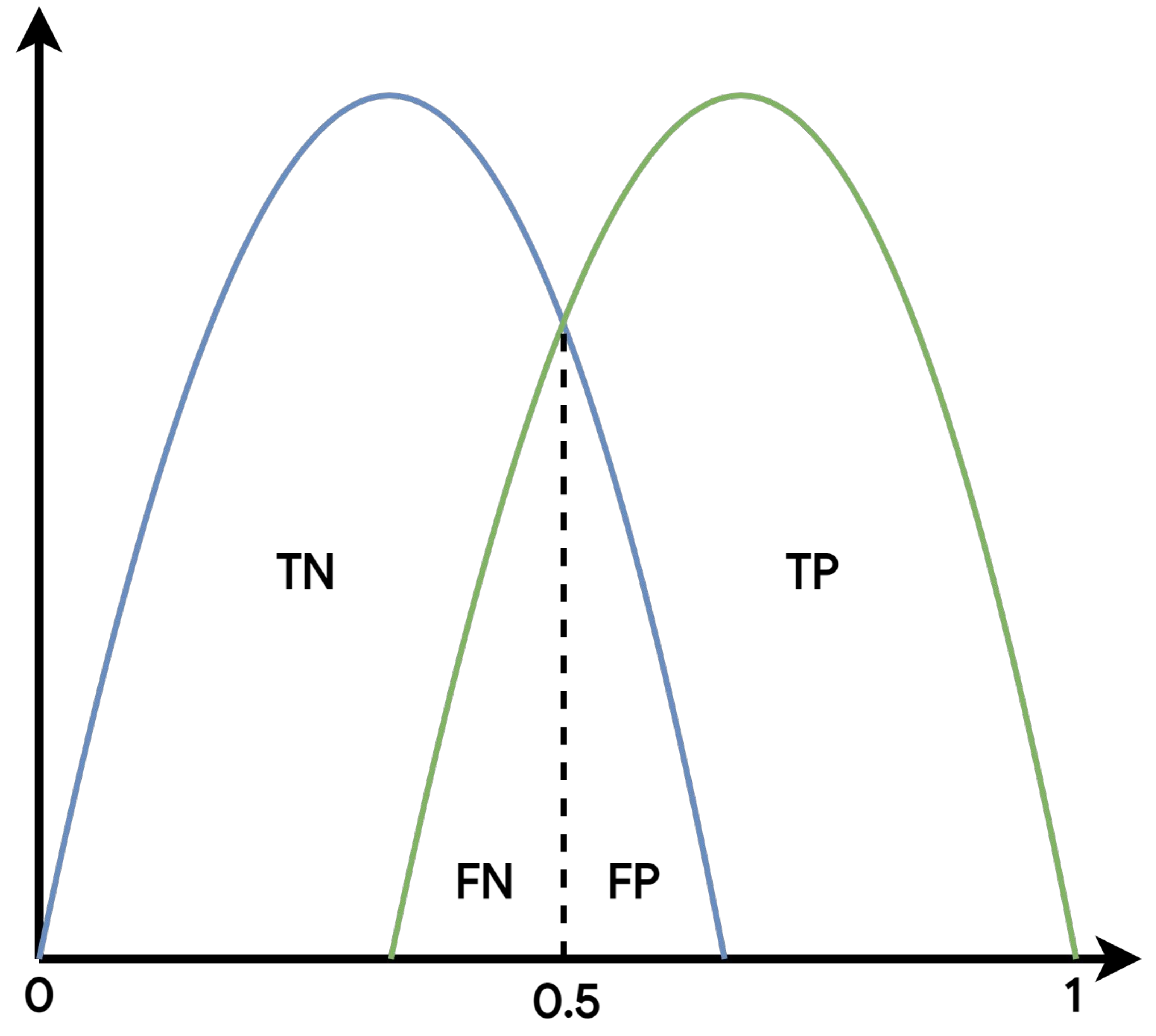
The lower RMSE wins (pss… careful not overfitting)

For classification problems?

Needed For the Exam

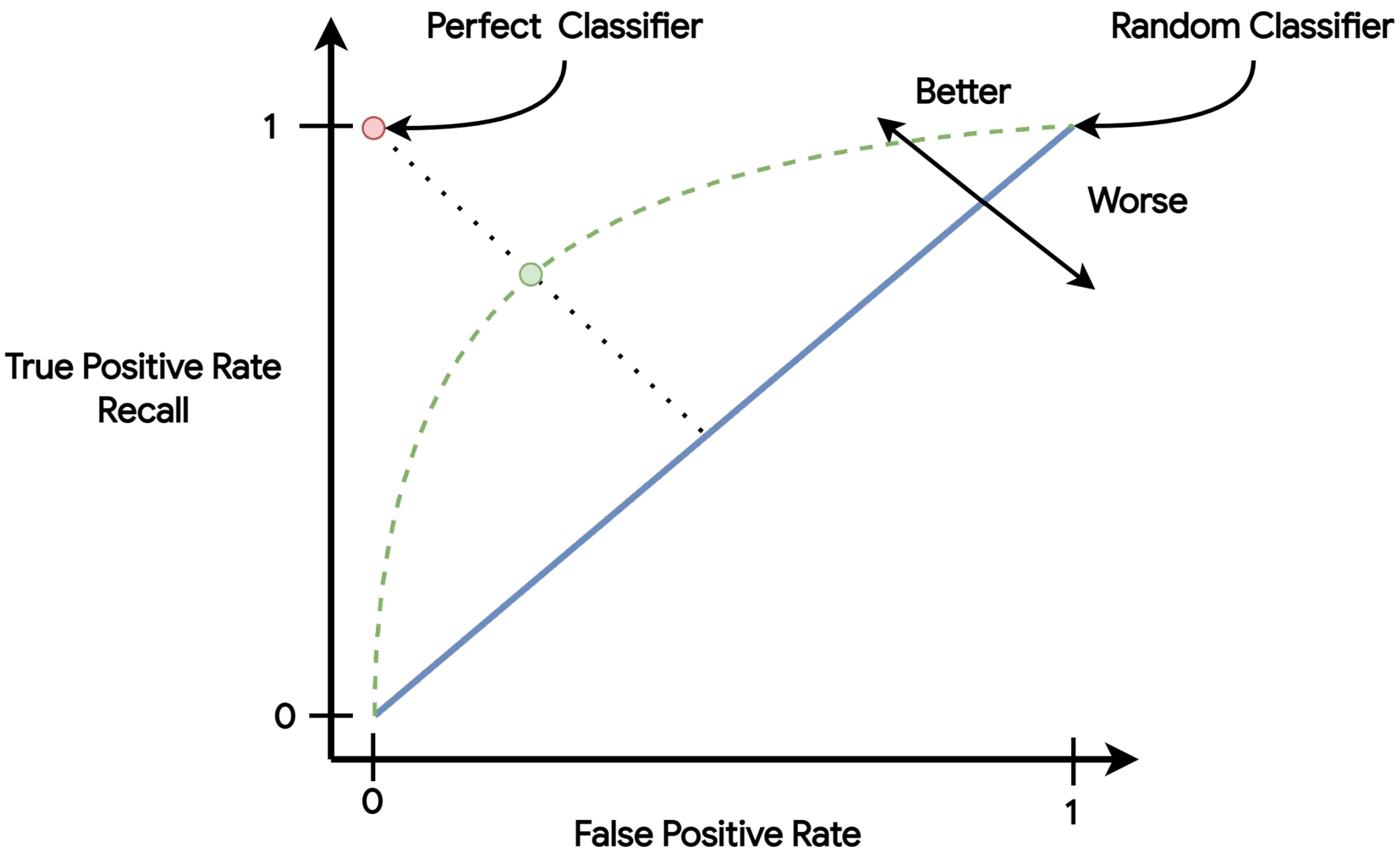
Precision or Recall?

- If you want to reduce False Negative (Type 2 Error) then you should focus on Recall
- If you want to reduce False Positive (Type 1 Error) then you should focus on Precision
- Threshold can be adjusted according with your business needs



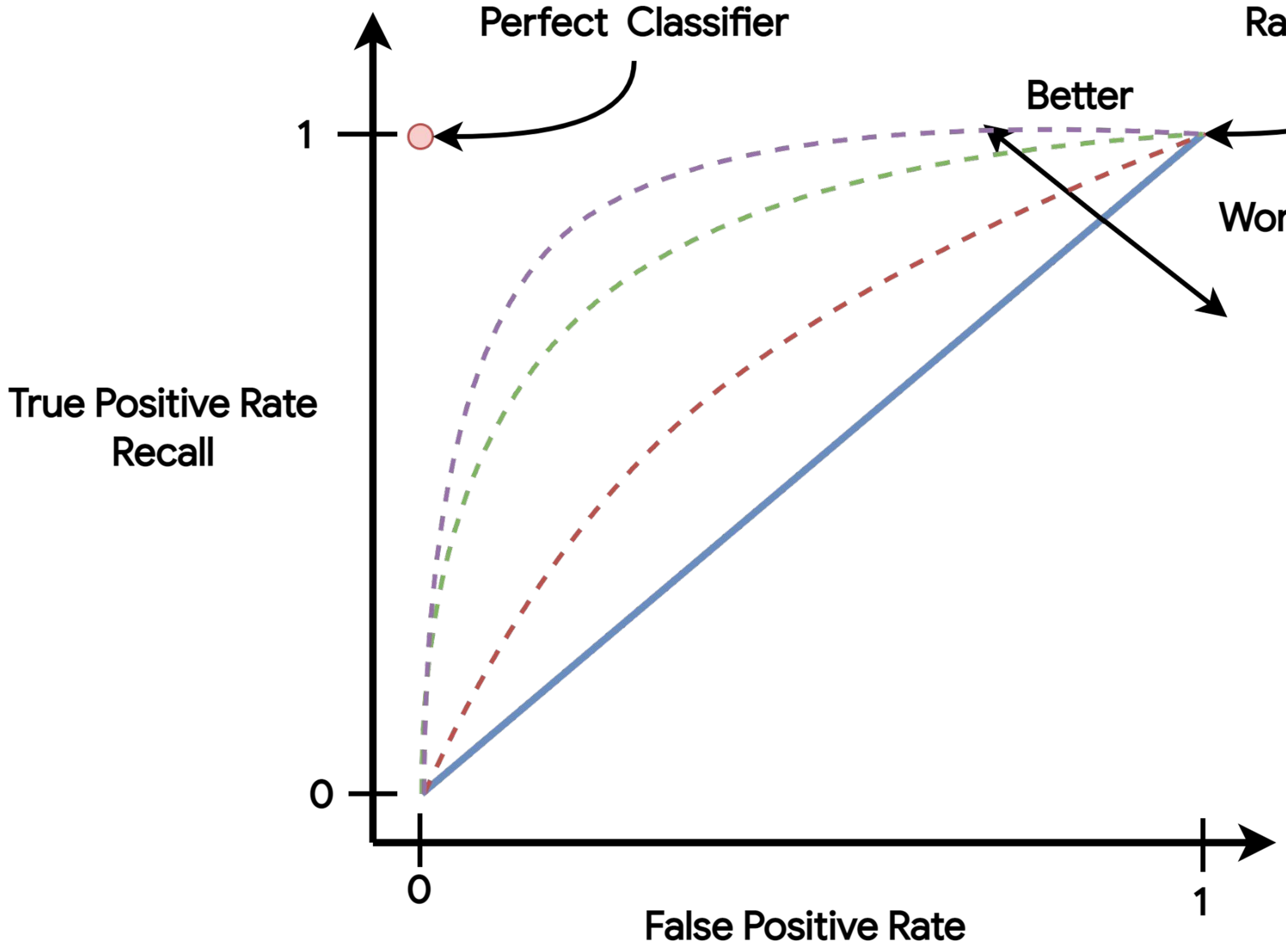
Needed For the Exam

How do I find the best threshold? Let’s ROC



Needed For the Exam

How do I compare classifiers? Area Under Curve (AUC)



End of Session 1

Thank you

Google Cloud

Google Cloud