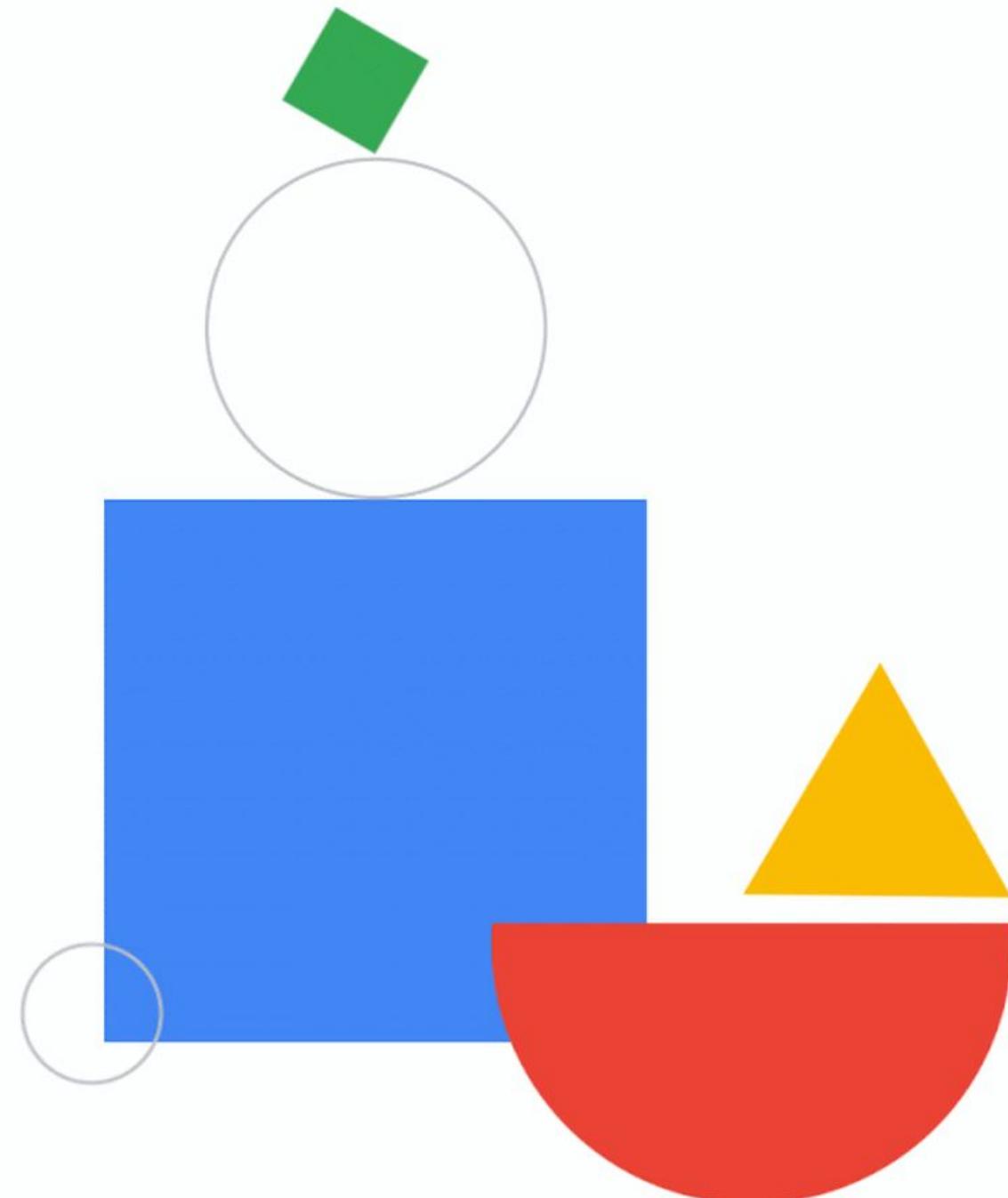


Google Cloud Academy

The Path to Partner Technical Readiness
Professional Machine Learning Engineer Certification

Session 3: Vertex AI & ML Model Metrics on Vertex AI



Instructor: Ben Ahmed



Google Cloud

Authorized Trainer



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

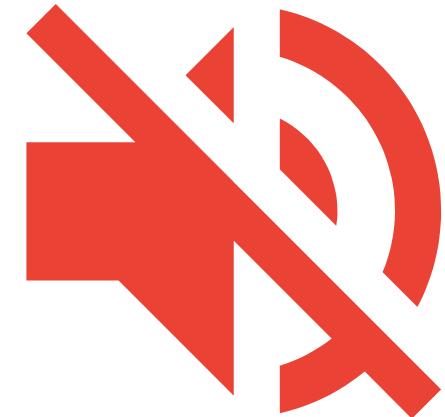
Machine learning Engineer & Data Engineer for Startups, Oil & Gas (BP) & Banking (HSBC)

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.



Course etiquette



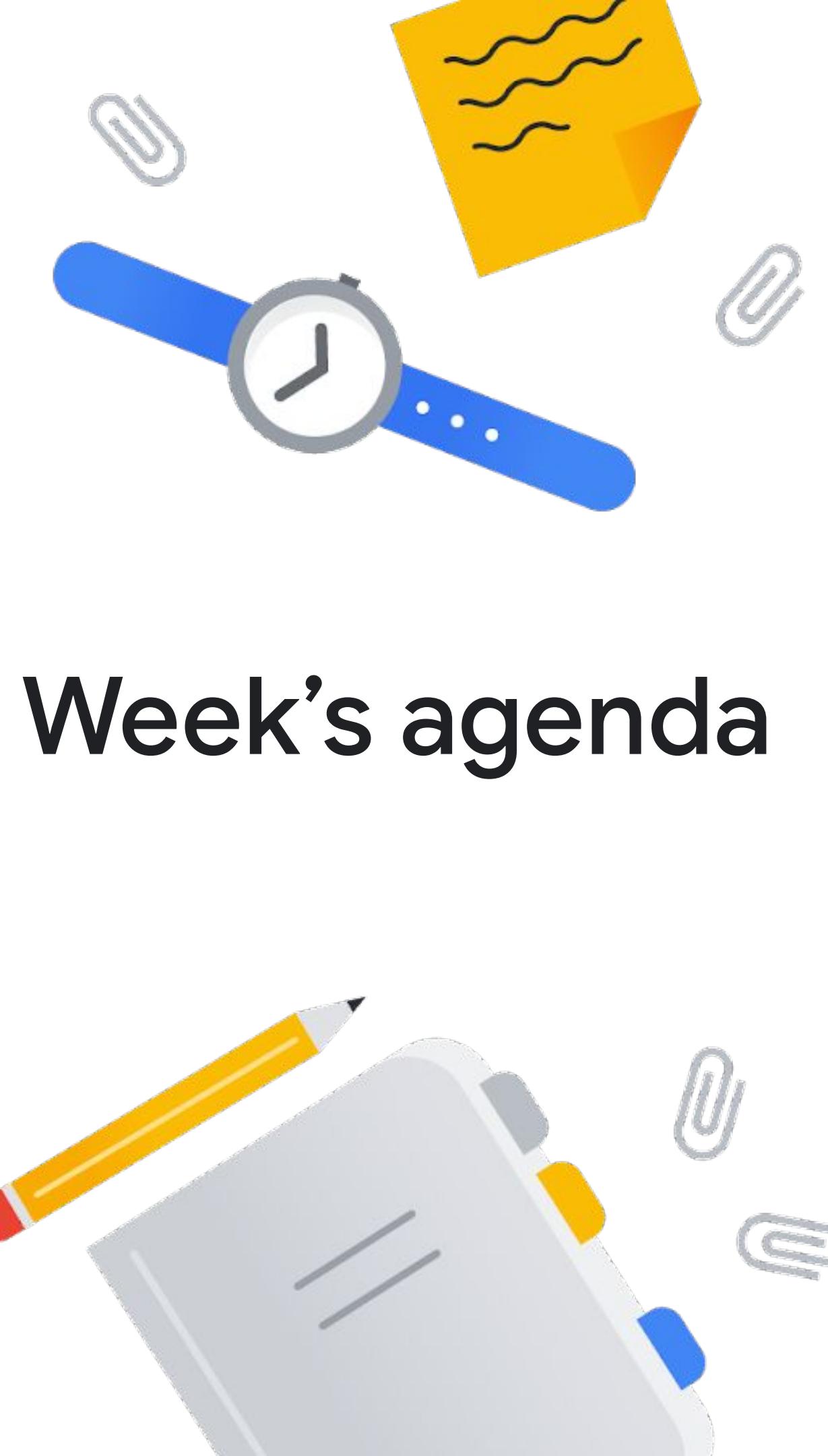
Please silence your phone and take calls outside.



Recording this class is prohibited.



Ask questions via Chat, If time permits



Week's agenda

- 01 [Data Extraction](#)
- 02 [VertexAI Overview](#)
- 03 [ML Models Overview](#)
- 04 [Model Optimization](#)
- 05 [Model Metrics](#)
- 06 [Machine Learning vs Deep Learning](#)

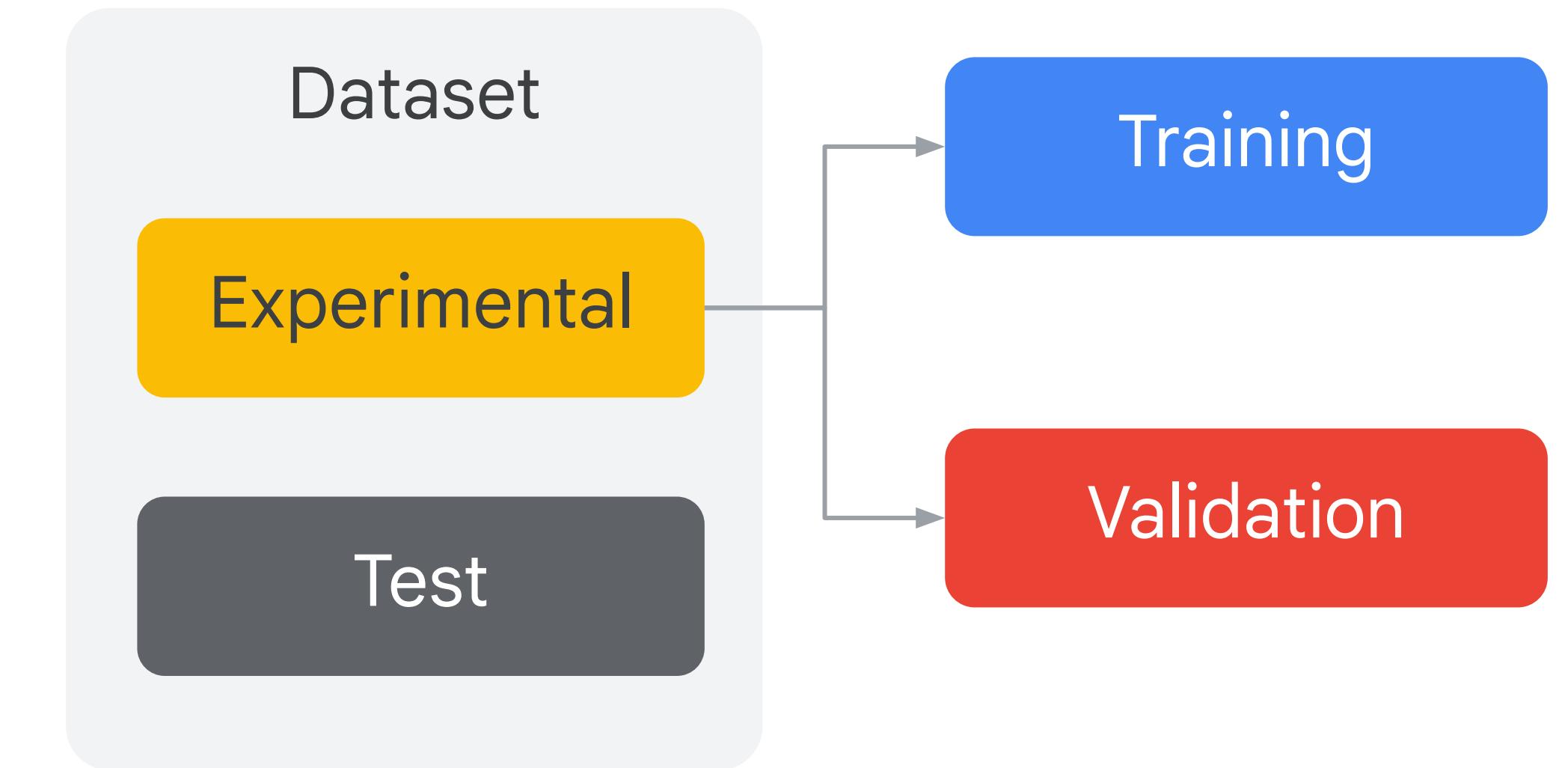


Some Sample Questions

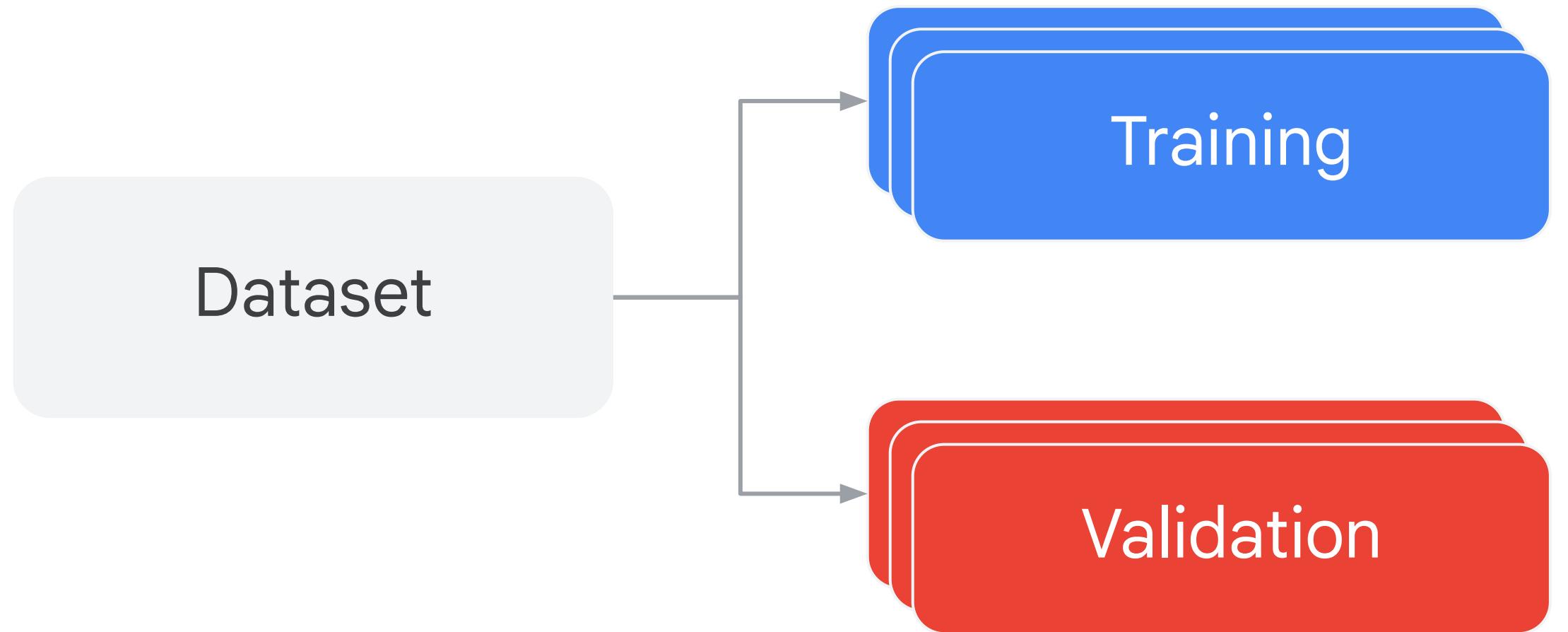


Data Extraction

Evaluate the final
model with
independent
test data

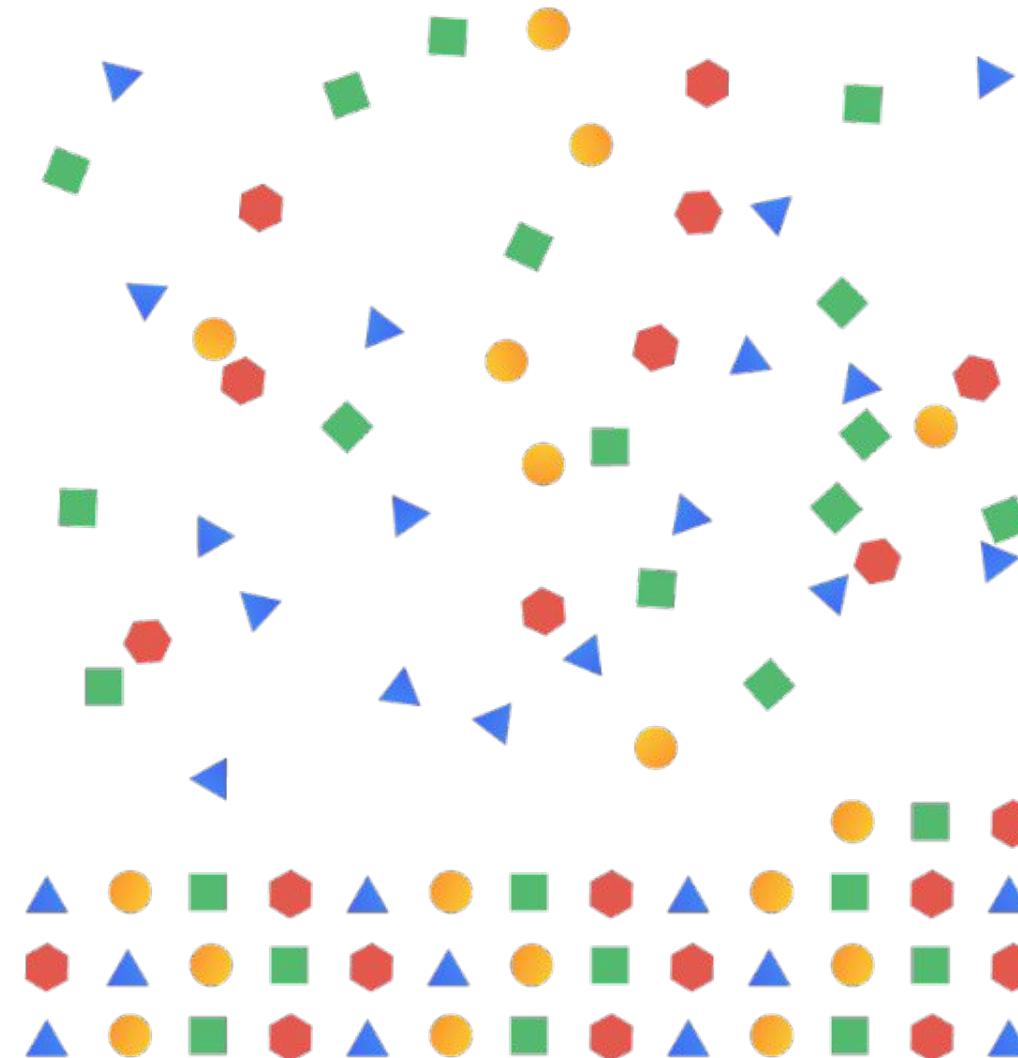


Evaluate the final
model with
cross-validation



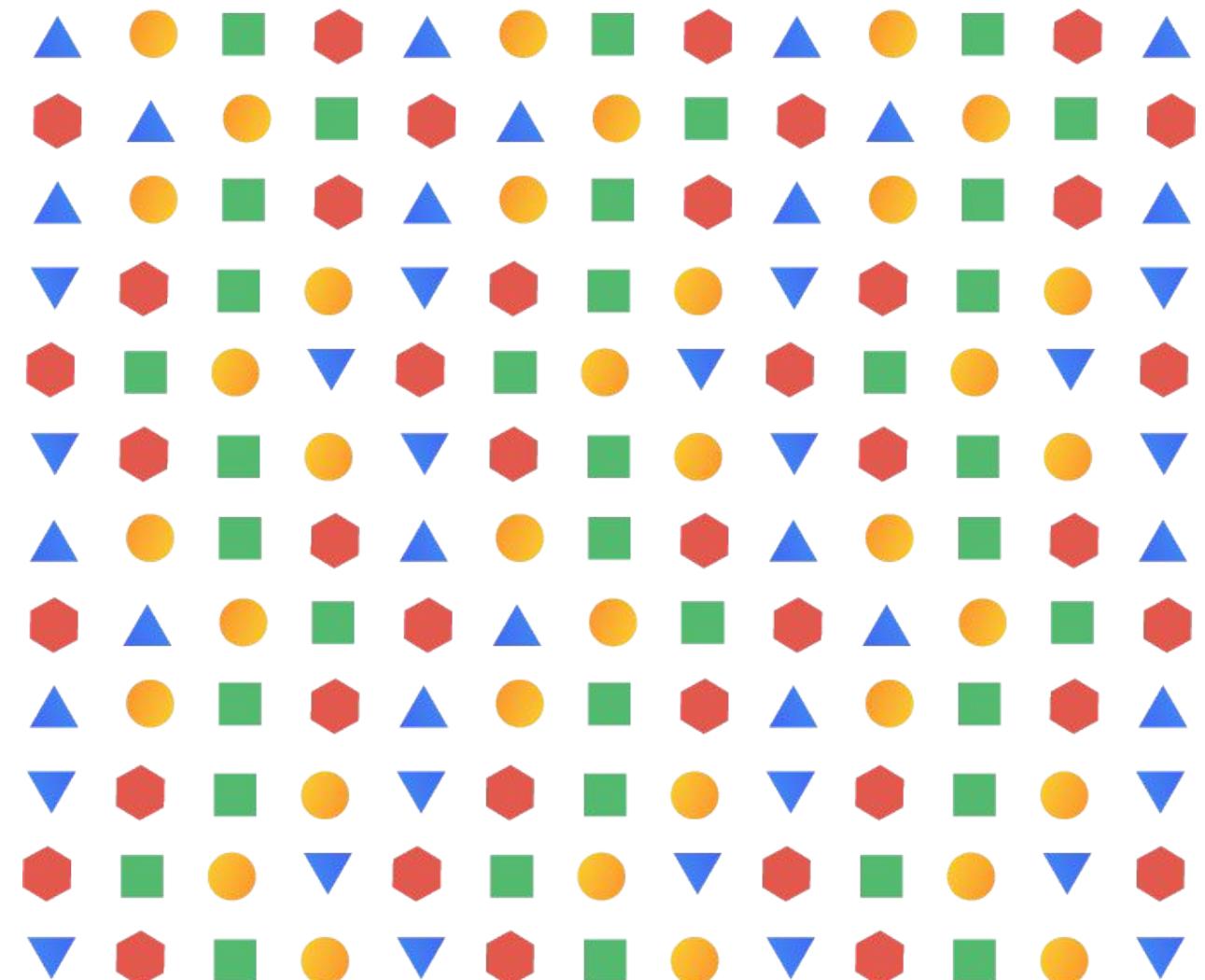
Training set

- The vast majority of your data should be here.
- This is the data your model “sees.”
- The model uses it to learn its parameters.



Validation set

- The validation set is sometimes referred to as the “dev” set, and is also used in the training process.
- It is used to tune the hyperparameters (variables that specify the model’s structure) after the model uses the training data in each iteration of the training process.
- Using the validation set to tune the hyperparameters means the model will generalize better.



Test set

- The test set is used after the training process is complete.
- The test set is a new challenge for the model to get a strong indication of how the model will perform on real-world data.



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

Dataset

Training

Validation

Test

Row	date	airline	departure_airport	departure_schedule	arrival_airport	arrival_delay
1	2019-08-07	TZ	SRQ	1255	IND	-14.0
2	2019-03-05	TZ	SRQ	2117	IND	-9.0
3	2020-04-12	TZ	SRQ	2000	IND	-17.0
4	2021-04-16	TZ	SRQ	1215	IND	-5.0
5	2021-03-20	TZ	SRQ	645	IND	14.0
6	2020-04-06	TZ	SRQ	1235	IND	-8.0

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.

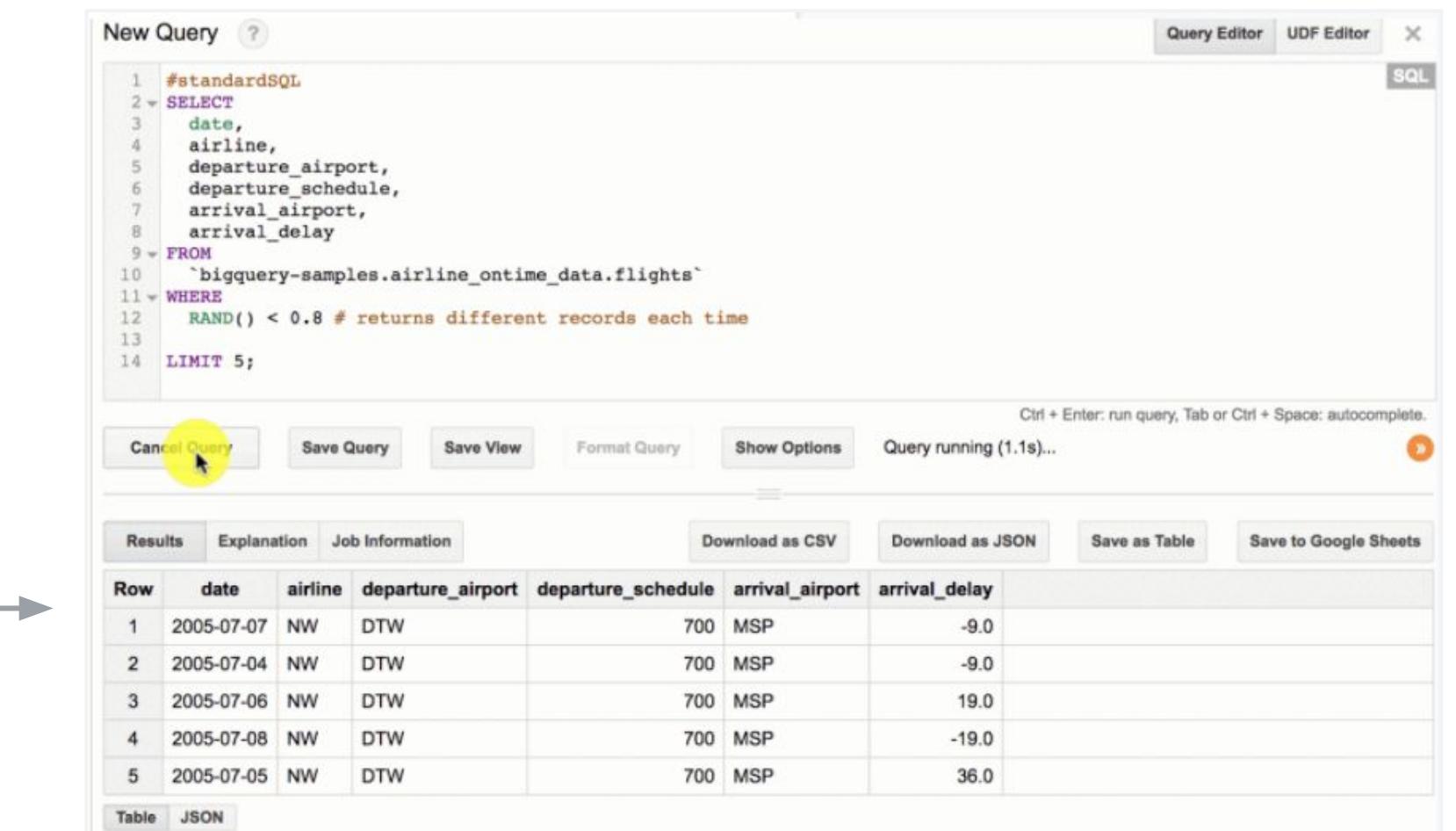


Naive random splitting is not repeatable

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 →



The screenshot shows the BigQuery Query Editor interface. The query editor window has tabs for 'Query Editor' and 'UDF Editor'. The SQL tab is selected, displaying the following code:

```
1 #standardSQL
2 SELECT
3   date,
4   airline,
5   departure_airport,
6   departure_schedule,
7   arrival_airport,
8   arrival_delay
9 FROM
10   `bigquery-samples.airline_ontime_data.flights`
11 WHERE
12   RAND() < 0.8 # returns different records each time
13 LIMIT 5;
```

The status bar at the bottom of the editor indicates 'Query running (1.1s)...'. Below the editor is a results table with the following data:

Row	date	airline	departure_airport	departure_schedule	arrival_airport	arrival_delay	
1	2005-07-07	NW	DTW		700	MSP	-9.0
2	2005-07-04	NW	DTW		700	MSP	-9.0
3	2005-07-06	NW	DTW		700	MSP	19.0
4	2005-07-08	NW	DTW		700	MSP	-19.0
5	2005-07-05	NW	DTW		700	MSP	36.0

Below the table are buttons for 'Table' and 'JSON'.

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.

Carefully choose which field will split your data



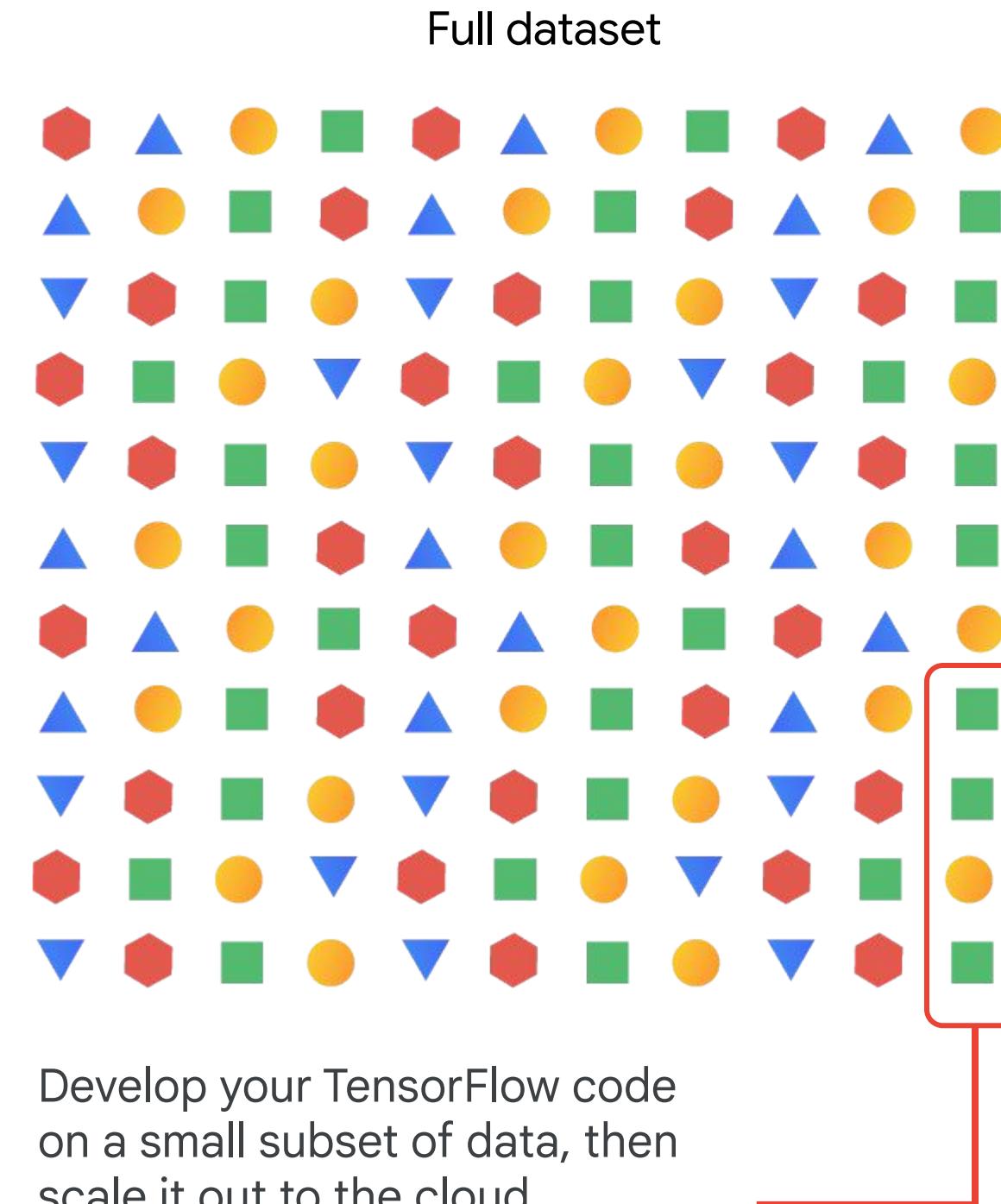
We hypothesize that flight delay depends on the carrier, time of day, weather, and airport characteristics (# of runways, etc.) We want to predict flight delays.

What field should we split our data on?

- Hash on date?
- Hash on airport?
- Hash on carrier name?

Split your data on a field you can afford to lose.

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



Pitfall: Chaining hashes to create subsets won't work

```
#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)),10) < 8
```



Then take 1 in 70 flights.

Take 80% of the dataset? Incorrect!

All records here will also be divisible by 10 (there is no new filtering happening!)

How we want to split our data

All flights (70 million)

1.5% (800,000)

50%
(400,000)

25%
(200,000)

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.



VertexAI: Overview

Course:

Launching into Machine Learning:

https://partner.cloudskillsboost.google/course_templates/8

LAB:

Training AutoML Classification Model - Structured Data:

https://partner.cloudskillsboost.google/course_sessions/2664667/labs/357113

Build and Deploy Machine Learning Solutions on Vertex AI:
<https://partner.cloudskillsboost.google/quests/183>

Google Cloud

**Pause to setup your lab:
(link in the chat)**

We will resume in:

<<8:00->>

Vertex AI > Dashboard

Vertex AI

- Dashboard
- Datasets
- Features
- Labeling tasks
- Workbench
- Pipelines
- Training
- Experiments
- Models
- Endpoints
- Batch predictions
- Marketplace

Dashboard

Get started with Vertex AI

Vertex AI empowers machine learning developers, data scientists, and data engineers to take their projects from ideation to deployment, quickly and cost-effectively. [Learn more](#)

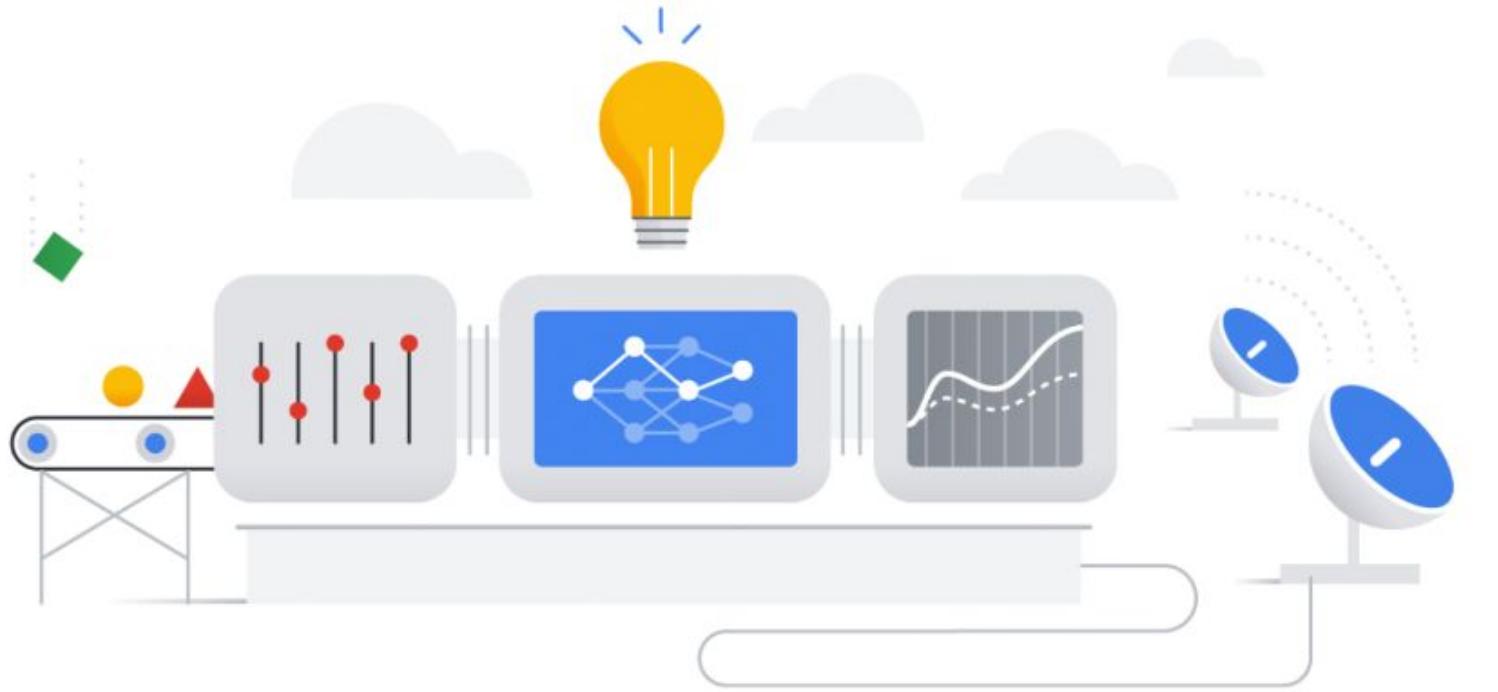
[ENABLE VERTEX AI API](#)

Region: us-central1 (Iowa) [?](#)

■ Prepare your training data
Collect and prepare your data, then import it into a dataset to train a model
[+ CREATE DATASET](#)

! Train your model
Train a best-in-class machine learning model with your dataset. Use Google's AutoML, or bring your own code.
[+ TRAIN NEW MODEL](#)

🔔 Get predictions
After you train a model, you can use it to get predictions, either online as an endpoint or through batch requests
[+ CREATE BATCH PREDICTION](#)



Vertex AI > Datasets

Vertex AI

Datasets

[CREATE](#)

REFRESH

Managed datasets contain data used to train a machine learning model. [Learn more](#)

Region: us-central1 (Iowa)

Filter: Enter a property name

Name	ID	Region	Type	Items	Labels	Last updated	Status	Metadata
taxifare	7644271029228929024	us-central1	Image	0	-	August 27, 2021	Created dataset	⋮

⋮

Features

Labeling tasks

Workbench

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Marketplace

Vertex AI > Datasets > Create > Image classification (Single-label)

Vertex AI

← Create dataset

Dataset name *

taxifare-single-label

Can use up to 128 characters.

Select a data type and objective

First select the type of data your dataset will contain. Then select an objective, which is the outcome that you want to achieve with the trained model. [Learn more about model types](#)

IMAGE TABULAR TEXT VIDEO

Image classification (Single-label)
Predict the one correct label that you want assigned to an image.

Image classification (Multi-label)
Predict all the correct labels that you want assigned to an image.

Image object detection
Predict all the locations of objects that you're interested in.

Image segmentation
Predict per-pixel areas of an image with a label.

Region
us-central1 (Iowa) ▾ ?

Encryption

Use a customer-managed encryption key (CMEK)

▲ SHOW LESS

You can use this dataset for other image-based objectives later by creating an annotation set. [Learn more about annotation sets](#)

CREATE CANCEL

Vertex AI > Datasets > Create > Image classification (Multi-label)

Vertex AI

Dashboard

Datasets

Features

Labeling tasks

Workbench

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Marketplace

Create dataset

Dataset name * Can use up to 128 characters.

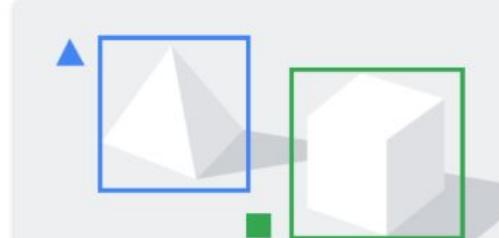
Select a data type and objective

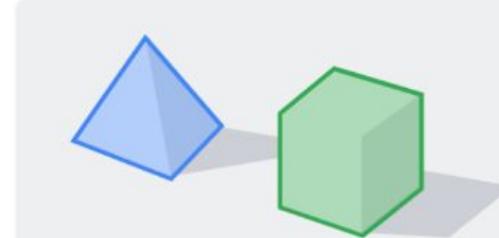
First select the type of data your dataset will contain. Then select an objective, which is the outcome that you want to achieve with the trained model. [Learn more about model types](#)

IMAGE TABULAR TEXT VIDEO


 Image classification (Single-label)
Predict the one correct label that you want assigned to an image.


 Image classification (Multi-label)
Predict all the correct labels that you want assigned to an image.


 Image object detection
Predict all the locations of objects that you're interested in.


 Image segmentation
Predict per-pixel areas of an image with a label.

Region ?

Encryption

Use a customer-managed encryption key (CMEK)

[SHOW LESS](#)

You can use this dataset for other image-based objectives later by creating an annotation set. [Learn more about annotation sets](#)

CREATE CANCEL

Vertex AI > Datasets > Create > Image object detection

Vertex AI Create dataset

Dashboard Dataset name * **axifare-object-detection**
Datasets Can use up to 128 characters.

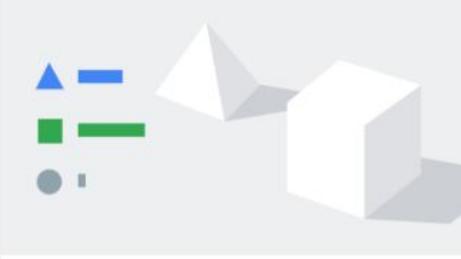
Features

Select a data type and objective

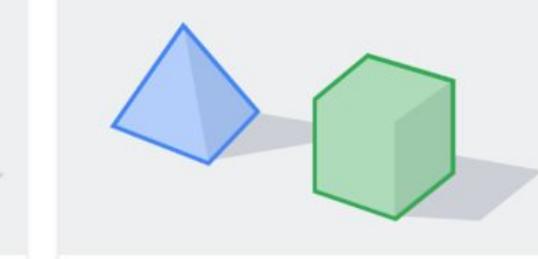
Please select the type of data your dataset will contain. Then select an objective, which is the outcome that you want to achieve with the trained model. [Learn more about model types](#)

IMAGE TABULAR TEXT VIDEO

Pipelines 

Training 

Experiments 

Models 

Endpoints

Batch predictions Region: us-central1 (Iowa) [?](#)

Marketplace

Encryption

Use a customer-managed encryption key (CMEK)

[SHOW LESS](#)

You can use this dataset for other image-based objectives later by creating an annotation set. [Learn more about annotation sets](#)

[CREATE](#) [CANCEL](#)

Vertex AI > Datasets > Create > Image segmentation

Vertex AI [Create dataset](#)

Dataset name * Can use up to 128 characters.

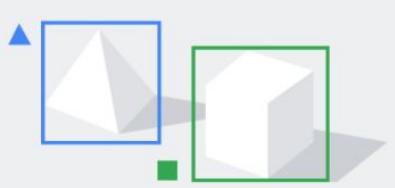
Select a data type and objective

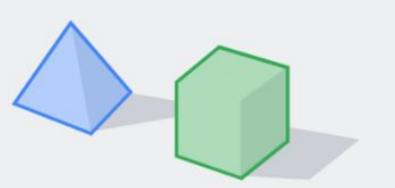
First select the type of data your dataset will contain. Then select an objective, which is the outcome that you want to achieve with the trained model. [Learn more about model types](#)

IMAGE TABULAR TEXT VIDEO

 Image classification (Single-label)
Predict the one correct label that you want assigned to an image.

 Image classification (Multi-label)
Predict all the correct labels that you want assigned to an image.

 Image object detection
Predict all the locations of objects that you're interested in.

 Image segmentation
Predict per-pixel areas of an image with a label.

Info Image segmentation is in Preview. While in Preview, AutoML training for image segmentation is not available. You can still create a custom-trained model.

Select annotation type

Your dataset images must all use the same annotation type. Once imported, annotations can't be added or edited. You can create a labeling task to add labels and annotations to images without them

Bitmap mask
 Polygon
 Polyline

Region

Encryption Use a customer-managed encryption key (CMEK)

[SHOW LESS](#)

You can use this dataset for other image-based objectives later by creating an annotation set. [Learn more about annotation sets](#)

[CREATE](#) [CANCEL](#)

Vertex AI > Features

Vertex AI

Features PREVIEW + CREATE ENTITY TYPE VIEW INGESTION JOBS REFRESH

Dashboard Datasets **Features** Labeling tasks Workbench Pipelines Training Experiments Models Endpoints Batch predictions Marketplace

Featurestores are typically used to store both large volumes of feature data and provide low-latency access to features for online applications. They manage the whole lifecycle of features: from training models to providing low-latency access to features for model inference. [Learn more](#)

Region us-central1 (Iowa)

Filter Enter a property name

Feature	Entity type	Featurestore	Description	Last updated	Created	Labels
film	feature	movie_prediction_20210826142335	film name	Aug 26, 2021, 8:08:11 PM	Aug 26, 2021, 8:08:11 PM	-
average_rating	movies	movie_prediction_20210826142335	The average rating for the movie, range is [1.0-5.0]	Aug 26, 2021, 7:57:08 PM	Aug 26, 2021, 7:57:08 PM	-
genres	movies	movie_prediction_20210826142335	The genres of the movie	Aug 26, 2021, 7:57:08 PM	Aug 26, 2021, 7:57:08 PM	-
subject	movies	movie_prediction_20210826142335	movie subject	Aug 26, 2021, 8:02:47 PM	Aug 26, 2021, 8:02:47 PM	-
title	movies	movie_prediction_20210826142335	The title of the movie	Aug 26, 2021, 8:45:48 PM	Aug 26, 2021, 7:57:08 PM	-
age	users	movie_prediction_20210826142335	User age	Aug 26, 2021, 7:57:02 PM	Aug 26, 2021, 7:57:02 PM	-
gender	users	movie_prediction_20210826142335	User gender	Aug 26, 2021, 7:57:02 PM	Aug 26, 2021, 7:57:02 PM	-
liked_genres	users	movie_prediction_20210826142335	An array of genres that this user liked	Aug 26, 2021, 7:57:02 PM	Aug 26, 2021, 7:57:02 PM	-

Vertex AI > Features > Create Entity Type

Create entity type

Region

us-central1 (Iowa)



Featurestore *

movie_prediction_20210826142335



Entity type name *

feature

Must start with a letter or underscore. Can use letters, numbers, and underscores.

Description

entity creation for movie prediction

CREATE

CANCEL

Create features

Feature names must start with a lowercase letter or underscore. Can use lowercase letters, numbers and underscores.

Feature name *

Feature name 1 *
subject

Value type *

Value type 1 *
STRING

Description

Description 1
movie subject



+ ADD ANOTHER FEATURE

CREATE

CANCEL

Vertex AI > Features > View Ingestion Jobs

[Ingestion jobs](#)

 REFRESH

List of operations importing features from a storage source.

Region
us-central1 (Iowa)  

	Ingestion job	Entity type	Featurestore	Elapsed time	Entities	Feature values	Last updated	Started	Status
	5600770669768867840	movies	movie_prediction_20210826142335	5 min 33 sec	4	12	Aug 26, 2021, 8:09:36 PM	Aug 26, 2021, 8:04:03 PM	Done
	2250092547005218816	users	movie_prediction_20210826142335	5 min 35 sec	7	12	Aug 26, 2021, 8:03:06 PM	Aug 26, 2021, 7:57:31 PM	Done

Vertex AI > Labeling tasks

Vertex AI

Labeling tasks

[CREATE](#) [MANAGE LABELER GROUPS](#) [REFRESH](#)

Dashboard

datasets Datasets

Features

Labeling tasks

Region: us-central1 (Iowa) [?](#)

Filter: Enter a property name [?](#)

Task	Data	Objective	Items	Progress	Labeler group
No rows to display					

Workbench

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Marketplace

Vertex AI > Labeling tasks > Create

New labeling task

- 1 Define labeling task
- 2 Set up labels
- 3 Set up instructions
- 4 Customize labelers

START TASK

CANCEL

Name *

Dataset * ▾

Annotation set * ▾

Objective ▾

Use active learning

Active learning expedites the labeling process by partially labeling your data with human labeler, then applying machine learning techniques to automatically label the rest.

▼ ADVANCED OPTIONS

CONTINUE

Vertex AI > Workbench

The screenshot shows the Vertex AI Workbench interface. On the left is a sidebar with icons for Dashboard, Datasets, Features, Labeling tasks, Workbench (which is selected), Pipelines, and Training. The main area has a header with 'Notebooks' and buttons for 'NEW NOTEBOOK', 'REFRESH', 'START', 'STOP', 'RESET', and 'DELETE'. Below the header are tabs for 'MANAGED NOTEBOOKS' (PREVIEW) and 'USER-MANAGED NOTEBOOKS' (selected). A message box contains two informational messages about notebook service migration and environment upgrades. At the bottom, a note states that notebooks have JupyterLab pre-installed and are configured with GPU-enabled machine learning frameworks, with a link to 'Learn more'.

Vertex AI

Notebooks

+ NEW NOTEBOOK C REFRESH ▶ START ■ STOP ⚡ RESET 🗑 DELETE

MANAGED NOTEBOOKS PREVIEW USER-MANAGED NOTEBOOKS EXECUTIONS PREVIEW SCHEDULES PREVIEW SCHEDULED RUNS

i Notebooks service has been moved under the Vertex AI Workbench service. Please find your Notebooks instances in Workbench under the User-Managed Notebooks tab.

i As of the M80 DLVM release, all environments will include JupyterLab 3.x by default. To continue using an existing environment's JupyterLab 1.x version, disable auto-upgrade (if enabled) and do not manually upgrade the environment to a new environment version. To create new Notebooks with JupyterLab 1.x installed, [see creating specific versions of Notebooks](#).

Notebooks have JupyterLab pre-installed and are configured with GPU-enabled machine learning frameworks. [Learn more](#)

Info panel

DOCUMENTATION LABELS

[Documentation Home](#) [Registering legacy DLVMs](#)

Vertex AI > Workbench > New Instance

The screenshot shows the Vertex AI Workbench interface for creating a new instance. The left sidebar has 'Workbench' selected. The main area displays a list of 'MANAGED NOTEBOOKS' with a red box highlighting the first item, 'Python 3'. The 'PREVIEW' tab is selected in the top right.

Notebooks

MANAGED NOTEBOOKS

Customize...

Python 3
Includes scikit-learn, pandas and more

Python 3 (CUDA Toolkit 11.0)
Optimized for NVIDIA GPUs

TensorFlow Enterprise
Includes Keras, scikit-learn, pandas, NLTK and more

PyTorch 1.9
Includes scikit-learn, pandas, NLTK and more

R 4.1
Includes basic R packages, scikit-learn, pandas, NLTK and more

RAPIDS 0.18 [EXPERIMENTAL]
Optimized for NVIDIA GPUs

Kaggle Python [BETA]
Python image for Kaggle Notebooks, supporting hundreds of machine learning libraries popular on Kaggle

Theia IDE [EXPERIMENTAL]
IDE with notebook support including scikit-learn, pandas, and more

Smart Analytics Frameworks
BigQuery, Apache Beam, Apache Spark, Apache Hive, and more

SCHEDULES **PREVIEW** **SCHEDULED RUNS**

Info panel

DOCUMENTATION **LABELS**

[Documentation Home](#) [Registering legacy DLVMs](#)

Environment

TensorFlow

TensorFlow

NumPy/Sci
learn

Performance issues c

Vertex AI > Pipelines

Vertex AI

Pipelines PREVIEW + CREATE RUN C REFRESH □ CLONE ← COMPARE ■ STOP

Dashboard Pipelines help you to automate, monitor, and govern your machine learning systems by orchestrating your workflow in a serverless manner. [Learn more](#)

Datasets Region us-central1 (Iowa) ▾ ?

Features

Labeling tasks Filter Status : Succeeded X Filter runs X ?

Workbench

Run	Status	Pipeline	Duration	Start time ↓	End time	⋮
hello-world-20210827124434	Succeeded	hello-world	6 min 50 sec	Aug 27, 2021, 6:14:35 PM	Aug 27, 2021, 6:21:25 PM	⋮

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Marketplace

Create pipeline run

Complete the following steps to create a new pipeline run

1 Run details

2 Parameters

SUBMIT

CANCEL

File *

intro_pipeline_job.json



CHOOSE

Choose a pipeline template from your local file system

Pipeline name *

hello-world

Run name *

hello-world-pipeline

Must start with a letter. Can use lowercase letters, numbers, and hyphens.

▼ ADVANCED OPTIONS

CONTINUE

Vertex AI > Training

Vertex AI

Training [+ CREATE](#) [REFRESH](#)

[Dashboard](#) [TRAINING PIPELINES](#) [CUSTOM JOBS](#) [HYPERPARAMETER TUNING JOBS](#)

[Datasets](#) [Features](#) [Labeling tasks](#) [Workbench](#) [Pipelines](#)

[Experiments](#) [Models](#) [Endpoints](#) [Batch predictions](#) [Marketplace](#)

Region: us-central1 (Iowa) [?](#)

Filter: Enter a property name [?](#) [☰](#)

Name	ID	Status	Job type	Model type	Created	Elapsed time
horses-humans-hypertune	4006398270266933248	✓ Finished	Training pipeline	Custom	Oct 30, 2021, 5:00:10 PM	43 min 11 sec
credit_risk_2_2021102774648	6151044078755643392	✓ Finished	Training pipeline	Tabular regression	Oct 27, 2021, 12:51:01 AM	2 hr 27 sec
tidyadvertising_today_2021104214553	642495221644394496	✓ Finished	Training pipeline	Tabular regression	Oct 4, 2021, 2:50:00 PM	1 hr 54 min
water_potability_ETT_10.04.2021_202110455935	3744490202484178944	✓ Finished	Training	Tabular	Oct 3, 2021,	2 hr 2 min

Vertex AI > Training > Create

Train new model

Training method

Model details

3 Compute and pricing

START TRAINING **CANCEL**

Dataset * movie (8 images) ▾ ?

Annotation set * movie_icn ▾ ?

Objective Image classification (Single-label) ▾

Please refer to the pricing guide for more details (and available deployment options) for each method.

AutoML
Train high-quality models with minimal effort and machine learning expertise. Just specify how long you want to train. [Learn more](#)

AutoML Edge
Train a model that can be exported for on-prem/on-device use. Typically has lower accuracy. [Learn more](#)

Custom training (advanced)
Run your TensorFlow, scikit-learn, and XGBoost training applications in the cloud. Train with one of Google Cloud's pre-built containers or use your own. [Learn more](#)

CONTINUE

Vertex AI > Training > Custom Jobs

Training + CREATE REFRESH

TRAINING PIPELINES CUSTOM JOBS HYPERPARAMETER TUNING JOBS

Custom jobs specify how Vertex AI runs your custom training code, including worker pools, machine types, and settings related to your Python training application and custom container. Custom jobs are only used by custom-trained models and not AutoML models. [Learn More](#)

Region
us-west1 (Oregon) ?

Filter Enter a property name ? ☰

Name	ID	Job type	Model type	Status	Created	Elapsed time	⋮
✓ multiworker-cassava-custom-job	7540397966729674752	Custom job	—	Succeeded	Aug 20, 2021, 9:40:54 PM	31 min 9 sec	⋮

Vertex AI > Training > Hyperparameter Tuning Jobs

Training

+ CREATE

REFRESH

TRAINING PIPELINES

CUSTOM JOBS

HYPERPARAMETER TUNING JOBS

Hyperparameter tuning searches for the best combination of hyperparameter values by optimizing metric values across a series of trials. Hyperparameter tuning is only used by custom-trained models and not AutoML models. [Learn More](#)

Region

us-central1 (Iowa)



Filter Enter a property name



Name	ID	Job type	Model type	Status	Created	Elapsed time	
✓ horses-humans-hypertune-hyperparameter-tuning-job	8903618662334726144	Hyperparameter tuning job	—	Succeeded	Aug 17, 2021, 8:06:37 PM	31 min 38 sec	⋮

Vertex AI > Experiments > Experiments

Vertex AI Experiments

EXPERIMENTS PREVIEW STUDIES PREVIEW TENSORBOARD INSTANCES PREVIEW

Region: us-central1 (Iowa) ▾ ?

Name	Tensorboard link	Updated	Created
No rows to display			

Experiments

Models

Endpoints

Batch predictions

Marketplace

Vertex AI > Experiments > Vizier Studies

Vertex AI

- Dashboard
- Datasets
- Features
- Labeling tasks
- Workbench
- Pipelines
- Training
- Experiments
- Models
- Endpoints
- Batch predictions
- Marketplace

Studies

EXPERIMENTS PREVIEW VIZIER STUDIES TENSORBOARD INSTANCES PREVIEW

Vertex Vizier is an optimization service that helps you tune hyperparameters in complex machine learning models. [Learn more](#)

Region — us-central1 (Iowa) ▾ ?

Filter Enter property name or value

Study name	ID	Objective	Created ↓
------------	----	-----------	-----------



Vertex AI > Experiments > Tensorboard Instances

Vertex AI

Tensorboard instances [+ CREATE](#)

EXPERIMENTS PREVIEW VIZIER STUDIES **TENSORBOARD INSTANCES PREVIEW**

Vertex TensorBoard instances are used to store experiments. Instances belong to a region and any experiments they store will belong to the same one. [Learn more](#)

Region us-central1 (Iowa) ▾ ?

Name	ID	Description	Updated	Created
credit_risk_tb	5604976026867204096	Credit risk Tensorboard Instance	November 2, 2021	November 2, 2021
tb-cloud-training-demos	7573400907748999168	—	June 25, 2021	June 25, 2021

- Dashboard
- Datasets
- Features
- Labeling tasks
- Workbench
- Pipelines
- Training
- Experiments**
- Models
- Endpoints
- Batch predictions
- Marketplace

Vertex Tensorboard

Allow Data Scientists and ML Researchers to easily collaborate by making it seamless to track, compare, and share their experiments.



Familiar experience: Experience matches the OSS TensorBoard tool users already know and love



No Setup: Readily available TensorBoard that can be used with no installation or additional deployment



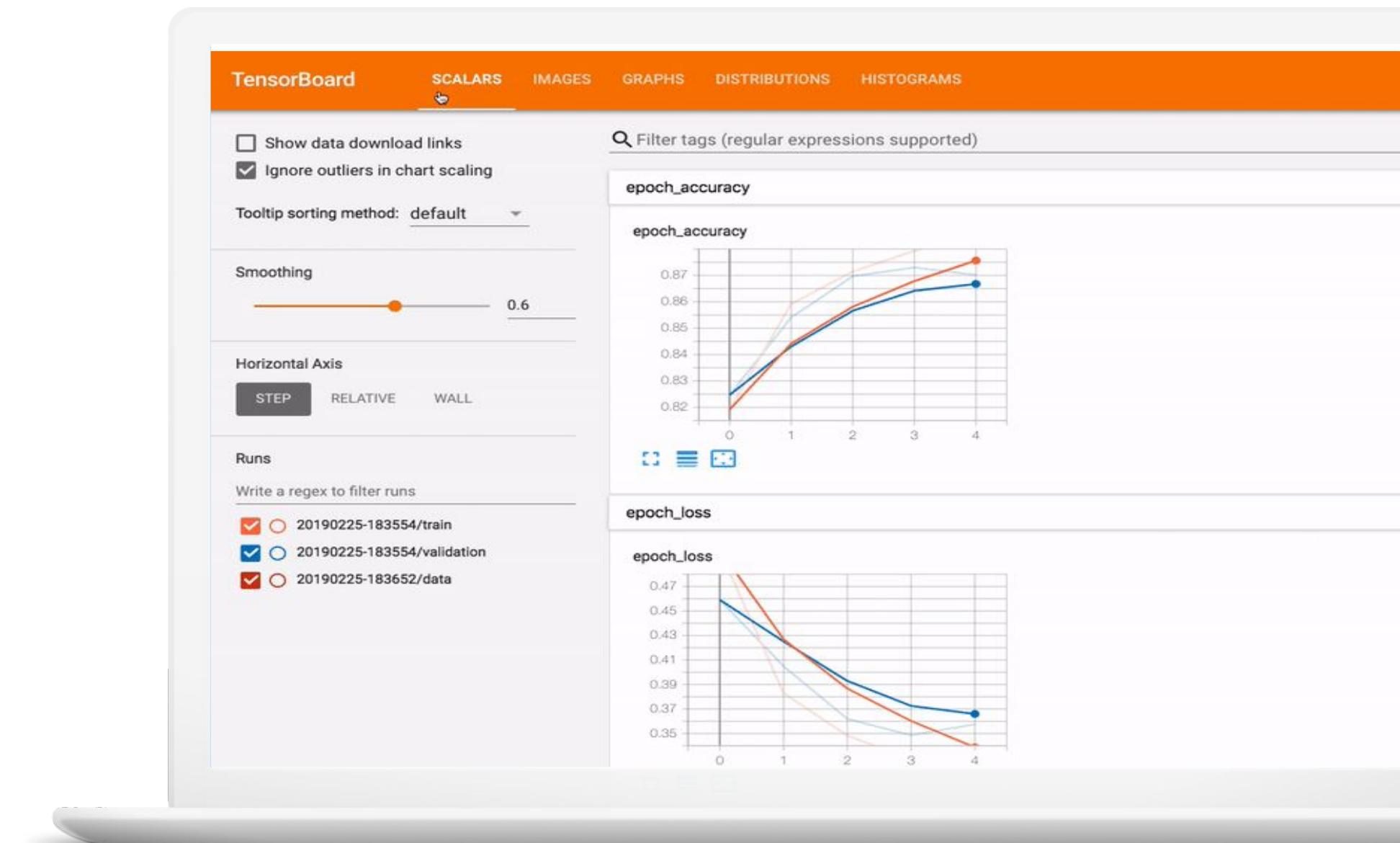
Easy Collaboration: Share results and insights with peers to accelerate ML research and development



Search & Compare: Quickly find specific experiments (e.g. based on hyperparameters) and compare them



Enterprise Ready: Secure, cost-effective, scalable and integrated with the rest of GCP



Vertex AI > Models

Vertex AI

Models [CREATE](#) [IMPORT](#)

Dashboard

Datasets

Features

Labeling tasks

Workbench

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Marketplace

tidyadvertising_today_2021104214553

5363356147640172544

tidyadvertising_today

1

Region

us-central1 (Iowa)

Filter Enter a property name

Name	ID	Status	Data	Endpoints
credit_risk_2_2021102774648	1151237052793094144	Ready	credit_risk_2	0
pre-trained_credit_risk_10-15-2021_v2_gs	8083108110106886144	Ready	-	0
pre-trained_credit_risk_10-15-2021	4236471078379061248	Ready	-	0
tidyadvertising_today_2021104214553	5363356147640172544	Ready	tidyadvertising_today	1

Vertex AI > Models > Create

Train new model

Training method

Model details

3 Compute and pricing

START TRAINING CANCEL

Dataset *
movie (8 images) ?

Annotation set *
movie_icn ?

Objective
Image classification (Single-label)

Please refer to the pricing guide for more details (and available deployment options) for each method.

AutoML

Train high-quality models with minimal effort and machine learning expertise. Just specify how long you want to train. [Learn more](#)

AutoML Edge

Train a model that can be exported for on-prem/on-device use. Typically has lower accuracy. [Learn more](#)

Custom training (advanced)

Run your TensorFlow, scikit-learn, and XGBoost training applications in the cloud. Train with one of Google Cloud's pre-built containers or use your own. [Learn more](#)

CONTINUE

Vertex AI > Models > Import

Import model

- Name and region
- 2 Model settings

IMPORT

CANCEL

- Import model artifacts into a new pre-built container
View the list of [supported runtimes](#) including TensorFlow, scikit-learn and XGBoost versions
- Import an existing custom container
Build a custom Docker container. Must be stored in [Container Registry](#) or [Artifact Registry](#)

Pre-built container settings

In order to run in a pre-built container, your code needs to be in Python 3.7

Model framework *

Model framework version *

Accelerator type *

Model artifact location (Cloud storage path) * gs:// cloud-ai-platform-f267098a-366f-478f-831d-bfa8d2ab9a5a [BROWSE](#)

Path to the Cloud Storage directory where the exported model file is stored (not the path to the model file itself). The model name must be one of: saved_model.pb, model.pkl, model.joblib, or model.bst, depending on which library you used.

Predict schemata

Optional. [Learn more about the predict schemata](#)

[BROWSE](#)

Cloud Storage location to a YAML file that defines the format of a single instance used in prediction and explanation requests.

[BROWSE](#)

Cloud Storage location to a YAML file that defines the prediction and explanation parameters.

[BROWSE](#)

Cloud Storage location to a YAML file that defines the format of a single prediction or explanation.

Vertex AI > Endpoints

Vertex AI

Endpoints [+ CREATE ENDPOINT](#)

Datasets

Features

Labeling tasks

Workbench

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Metadata

Marketplace

Endpoints are machine learning models made available for online prediction requests. Endpoints are useful for timely predictions from many users (for example, in response to an application request). You can also request batch predictions if you don't need immediate results.

To create an endpoint, you need at least one machine learning model. [Learn more](#)

Region [?](#)

Filter Enter a property name

<input type="checkbox"/>	Name	ID	Status	Models	Region	Monitoring	Most recent alerts	Last updated	API
<input type="checkbox"/>	tidyadvertising_endpoint	5842584887676108800	✓ Active	1	us-central1	Enabled	0 alerts	Nov 2, 2021, 9:48:20 AM	Sample request
<input type="checkbox"/>	my_usahousing_10.02.2021	563495311188688896	✓ Active	1	us-central1	Enabled	0 alerts	Nov 2, 2021, 9:26:26 AM	Sample request
<input type="checkbox"/>	credit_risk	1241146317619593216	✓ Active	1	us-central1	Enabled	0 alerts	Nov 2, 2021, 9:21:52 AM	Sample request
<input type="checkbox"/>	hello_endpoint	2976791392761675776	✓ Active	1	us-central1	Disabled	—	Aug 6, 2021, 8:36:27 PM	Sample request

Vertex AI > Endpoints > Create Endpoint

New endpoint

1 Define your endpoint

2 Model settings

CREATE CANCEL

Endpoint name * movie-endpoint ?

Location

Region us-central1 (Iowa) ▼ ?

Access

Determines how your endpoint can be accessed. By default, endpoints are available for prediction serving through a REST API. Endpoint access can't be changed after the endpoint is created.

Standard
Makes the endpoint available for prediction serving through a REST API. AutoML and custom-trained models can be added to standard endpoints.

Private
Create a private connection to this endpoint using a VPC network and [private services access](#). Only custom-trained and tabular models can be added to private endpoints.
[Learn more](#)

▼ ADVANCED OPTIONS

CONTINUE

Vertex AI > Batch Predictions

Vertex AI Batch predictions + CREATE REFRESH

Datasets Batch prediction intakes a group of prediction requests and outputs the results to a specified location. Use batch prediction when you don't require an immediate response and want to process accumulated data with a single request. [Learn more](#)

Features

Labeling tasks Region: us-central1 (Iowa) ?

Workbench

Pipelines Filter: Enter a property name ? III

Name	Model	Objective	Last updated	Status	Actions
cifar10_batch-20210827070640	cifar10-20210827070640	Custom	August 27, 2021	Done	⋮

Training

Experiments

Models

Endpoints

Batch predictions Selected

Metadata

Marketplace

Vertex AI > Batch Predictions > Create

New batch prediction

Batch prediction name *

batch-prediction

Model name *

cifar10-20210827120639

Select source

File on Cloud Storage (JSONL, CSV, TFRecord, and gzipped TFRecord)

Files on Cloud Storage (file list). [Learn more](#)

File on Cloud Storage (JSONL, CSV, TFRecord, and gzipped TFRecord)

If you are using a CSV file, your first row must be a header row containing column names

[More info on data formats](#)

 gs:// Source path *

BROWSE

Select a Cloud Storage location

Prediction results will be stored in the selected Cloud Storage bucket

Output format

JSONL

Destination path *

gs:// cloud-ai-platform-f267098a-366f-478f-831d-bfa8d2ab9a5a

BROWSE

Destination bucket must be standard storage class and located in us-central1 (single region only)

Number of compute nodes *

3

Enter one or more compute nodes to run this batch prediction job. The more nodes assigned to run this batch prediction job, the faster the batch prediction will complete.
[Learn more about pricing](#)

Machine type *

n1-standard-2, 2 vCPUs, 7.5 GiB memory



Accelerator type

ADVANCED OPTIONS

CREATE

CANCEL

Vertex AI > Batch Predictions > Create

New batch prediction

Batch prediction name *

batch-prediction

Model name *

cifar10-20210827120639

Select source

File on Cloud Storage (JSONL, CSV, TFRecord, and gzipped TFRecord)

Files on Cloud Storage (file list). [Learn more](#)

File on Cloud Storage (JSONL, CSV, TFRecord, and gzipped TFRecord)

If you are using a CSV file, your first row must be a header row containing column names

[More info on data formats](#)

 gs:// Source path *

BROWSE

Select a Cloud Storage location

Prediction results will be stored in the selected Cloud Storage bucket

Output format

JSONL

Destination path *

gs:// cloud-ai-platform-f267098a-366f-478f-831d-bfa8d2ab9a5a

BROWSE

Destination bucket must be standard storage class and located in us-central1 (single region only)

Number of compute nodes *

3

Enter one or more compute nodes to run this batch prediction job. The more nodes assigned to run this batch prediction job, the faster the batch prediction will complete.
[Learn more about pricing](#)

Machine type *

n1-standard-2, 2 vCPUs, 7.5 GiB memory



Accelerator type

ADVANCED OPTIONS

CREATE

CANCEL

Vertex AI > Metadata

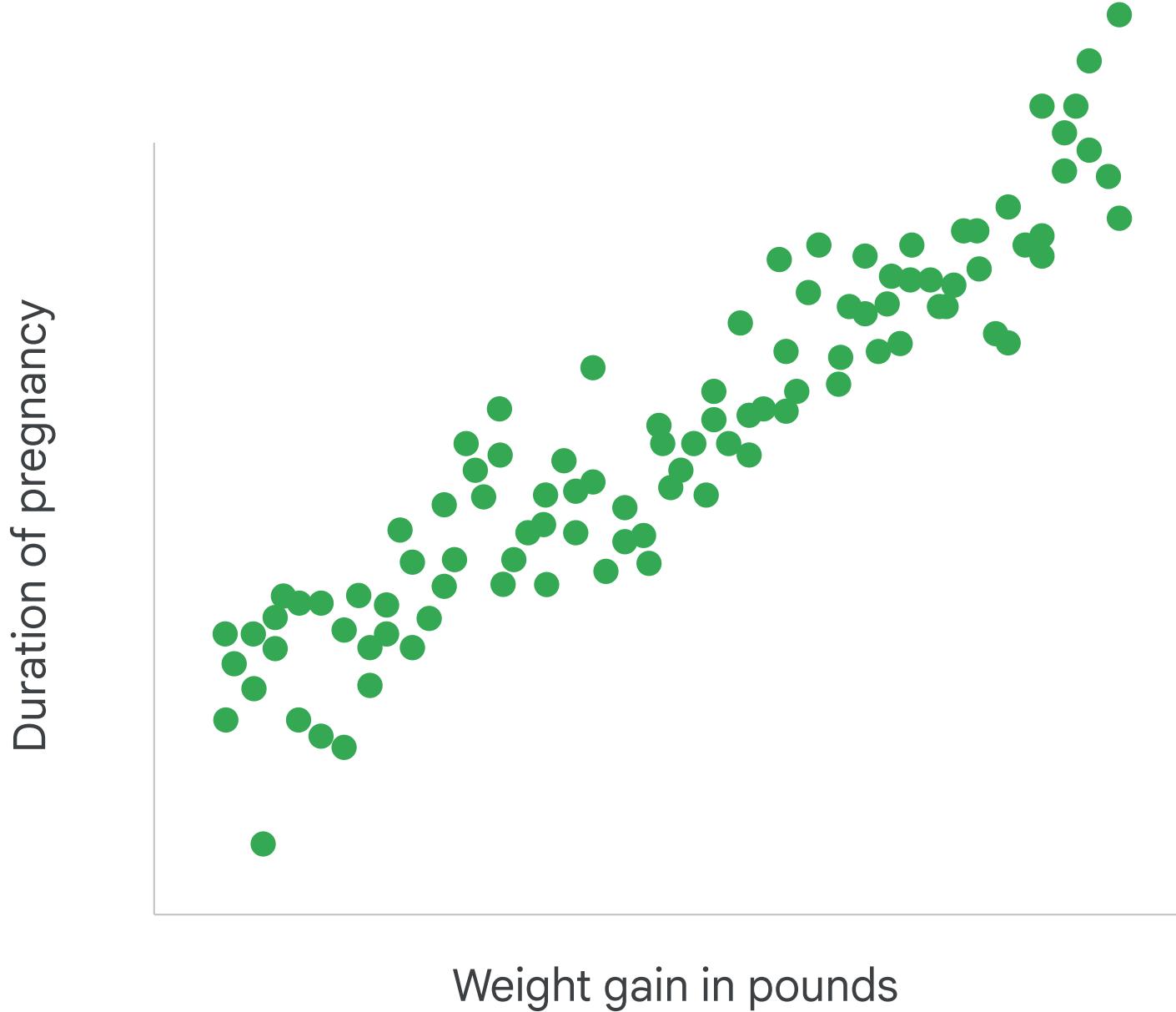
Vertex AI	Metadata																				
 Datasets	ML Metadata is a service that provides capabilities for managing the lifecycle of metadata consumed and produced by machine-learning (ML) workflows. Learn more																				
 Features																					
 Labeling tasks																					
 Workbench																					
 Pipelines	Region us-central1 (Iowa) ▾ ?																				
 Training	Metadata Store default Usage 848KB As of Nov 4, 2021																				
 Experiments	Filter Enter a property name																				
 Models	<table border="1"><thead><tr><th>Artifact name</th><th>Type</th><th>URI</th><th>Created</th><th>Last modified</th></tr></thead><tbody><tr><td><input type="checkbox"/> chicago-taxi-tips-classifier-v01-experiment-run-gcp-20210626020446-metrics</td><td>system.Metrics</td><td></td><td>Jun 25, 2021, 7:04:47 PM</td><td>Jun 25, 2021, 7:06:56 PM</td></tr><tr><td><input type="checkbox"/> chicago-taxi-tips-classifier-v01-experiment-run-gcp-20210626012119-metrics</td><td>system.Metrics</td><td></td><td>Jun 25, 2021, 6:21:20 PM</td><td>Jun 25, 2021, 6:53:39 PM</td></tr><tr><td><input type="checkbox"/> chicago-taxi-tips-classifier-v01-experiment-run-local-20210626010056-metrics</td><td>system.Metrics</td><td></td><td>Jun 25, 2021, 6:00:57 PM</td><td>Jun 25, 2021, 6:13:18 PM</td></tr></tbody></table>	Artifact name	Type	URI	Created	Last modified	<input type="checkbox"/> chicago-taxi-tips-classifier-v01-experiment-run-gcp-20210626020446-metrics	system.Metrics		Jun 25, 2021, 7:04:47 PM	Jun 25, 2021, 7:06:56 PM	<input type="checkbox"/> chicago-taxi-tips-classifier-v01-experiment-run-gcp-20210626012119-metrics	system.Metrics		Jun 25, 2021, 6:21:20 PM	Jun 25, 2021, 6:53:39 PM	<input type="checkbox"/> chicago-taxi-tips-classifier-v01-experiment-run-local-20210626010056-metrics	system.Metrics		Jun 25, 2021, 6:00:57 PM	Jun 25, 2021, 6:13:18 PM
Artifact name	Type	URI	Created	Last modified																	
<input type="checkbox"/> chicago-taxi-tips-classifier-v01-experiment-run-gcp-20210626020446-metrics	system.Metrics		Jun 25, 2021, 7:04:47 PM	Jun 25, 2021, 7:06:56 PM																	
<input type="checkbox"/> chicago-taxi-tips-classifier-v01-experiment-run-gcp-20210626012119-metrics	system.Metrics		Jun 25, 2021, 6:21:20 PM	Jun 25, 2021, 6:53:39 PM																	
<input type="checkbox"/> chicago-taxi-tips-classifier-v01-experiment-run-local-20210626010056-metrics	system.Metrics		Jun 25, 2021, 6:00:57 PM	Jun 25, 2021, 6:13:18 PM																	
 Endpoints																					
 Batch predictions																					
 Metadata																					
 Marketplace																					



Model Generalization

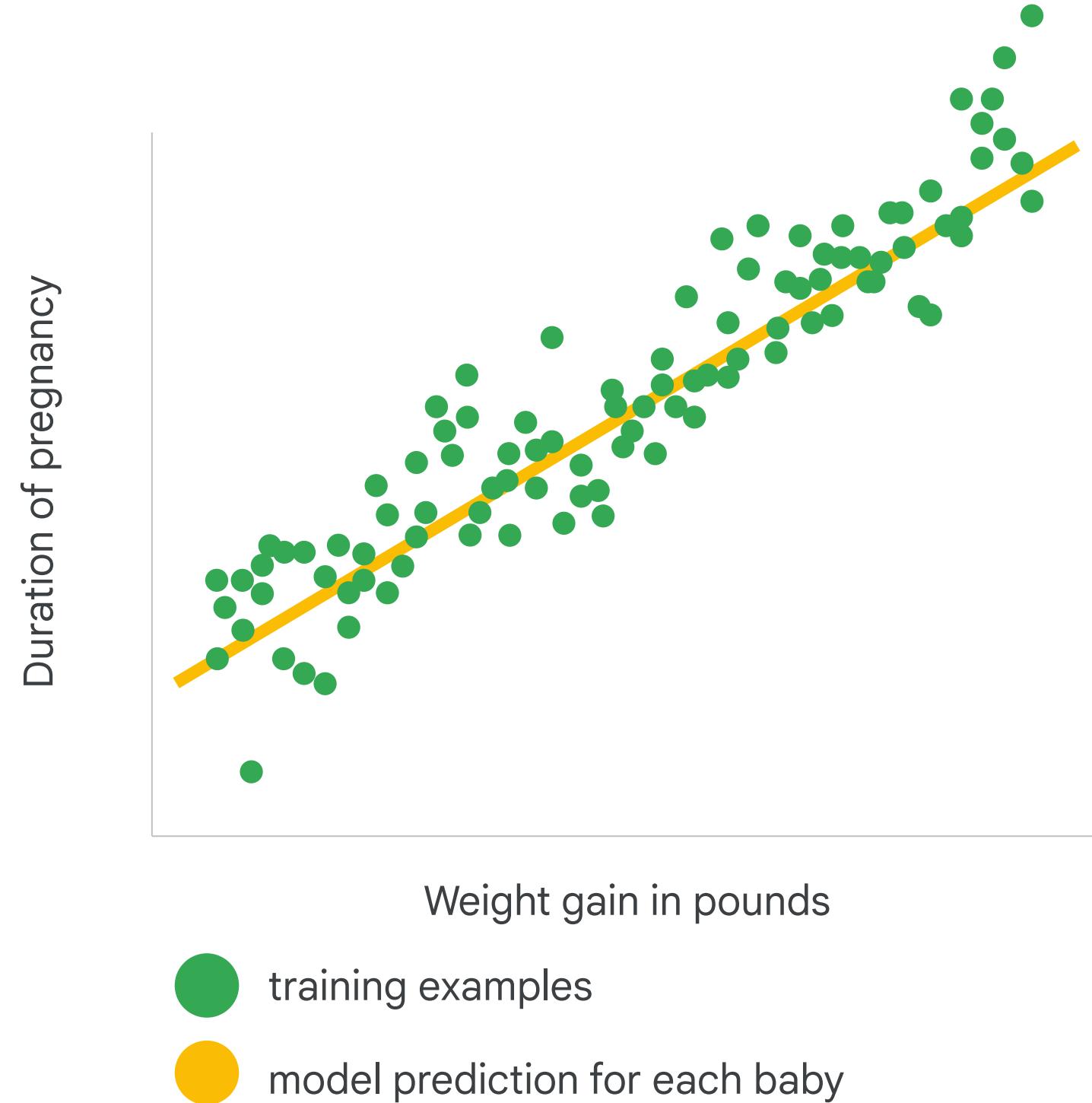
Suppose we want to predict duration of pregnancy based on mother's weight gain in pounds

What is the error measure to optimize?



Model 1 is a linear
model using linear
regression

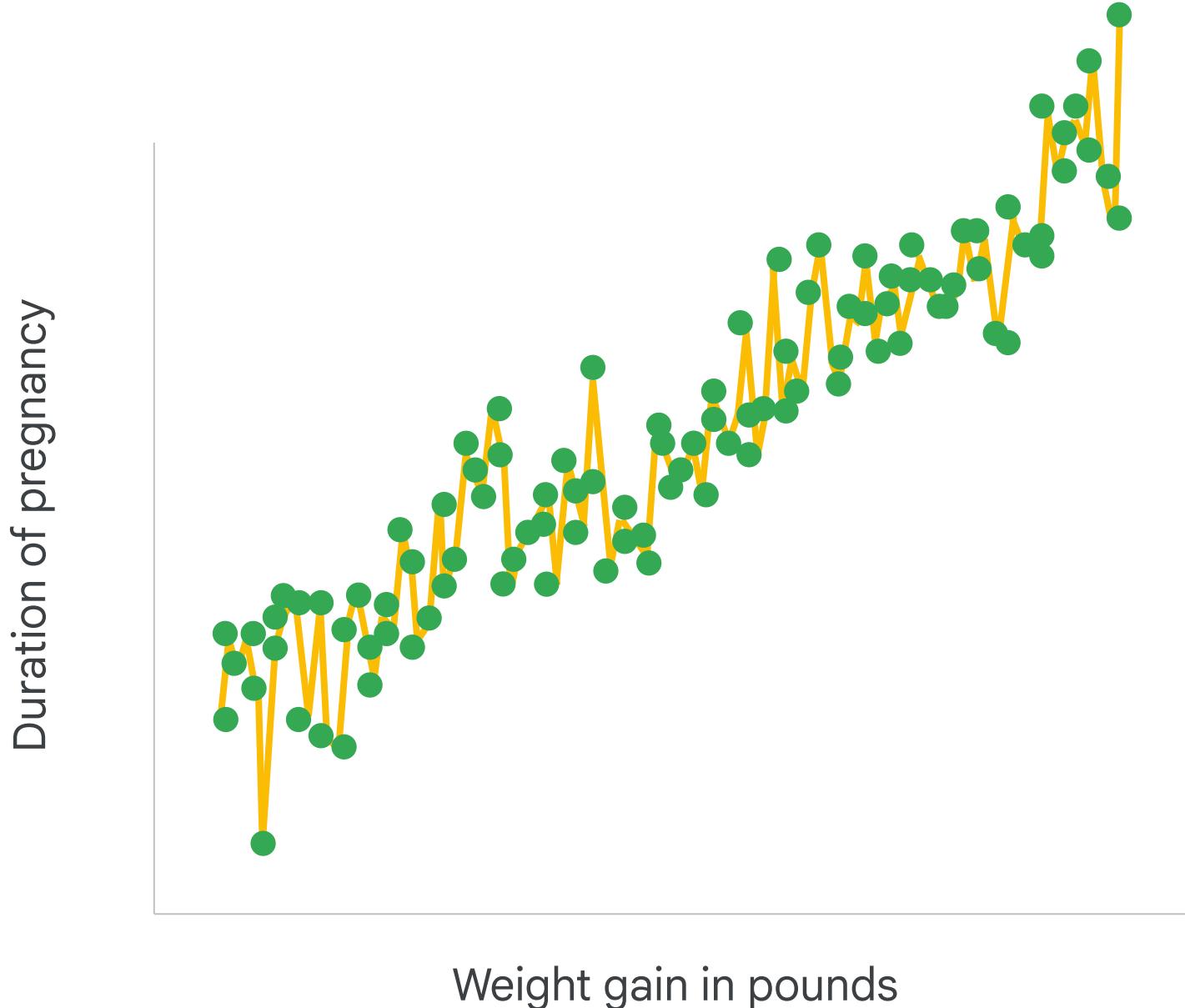
RMSE = 2.224



Model 2 has more
free parameters

RMSE = 0

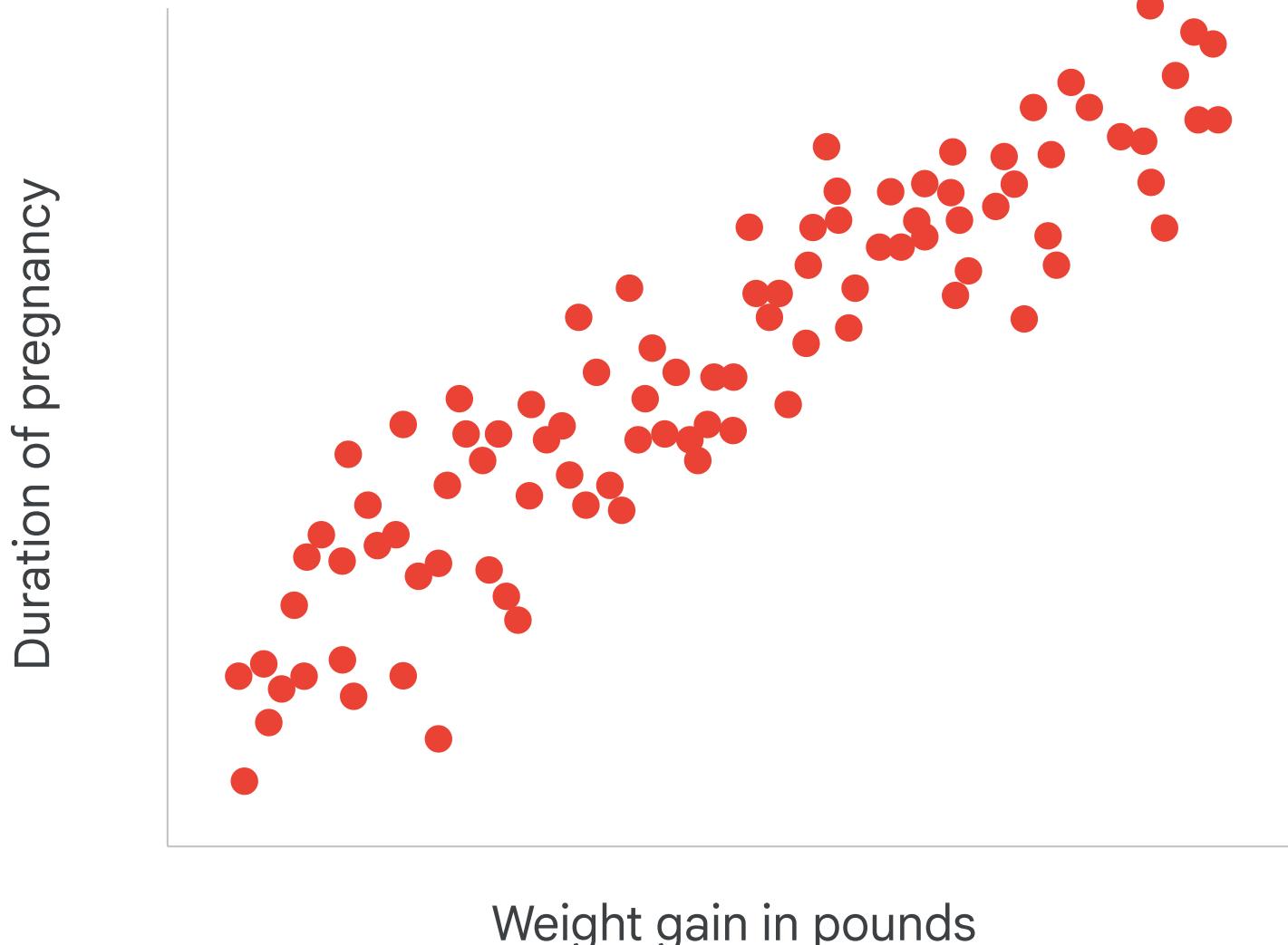
Which model is better?
How can you tell?



Does the model generalize to new data?

Need data that were not
used in training.

New data the model hasn't seen before.



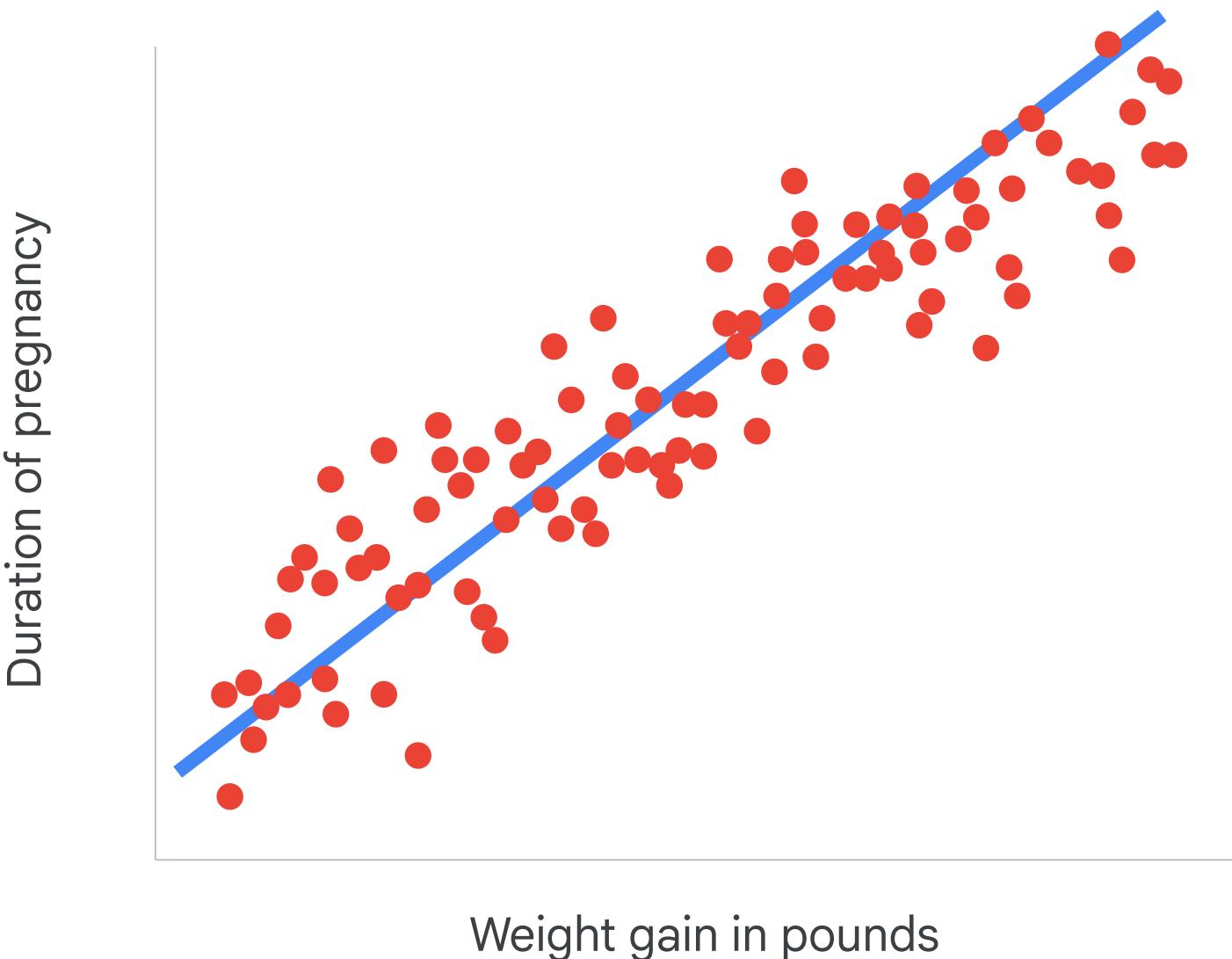
Model 1
generalizes well

Old RMSE = 2.224

New RMSE = 2.198

Pretty similar = good

New data the model hasn't seen before.



Model 2 **does not**
generalize well

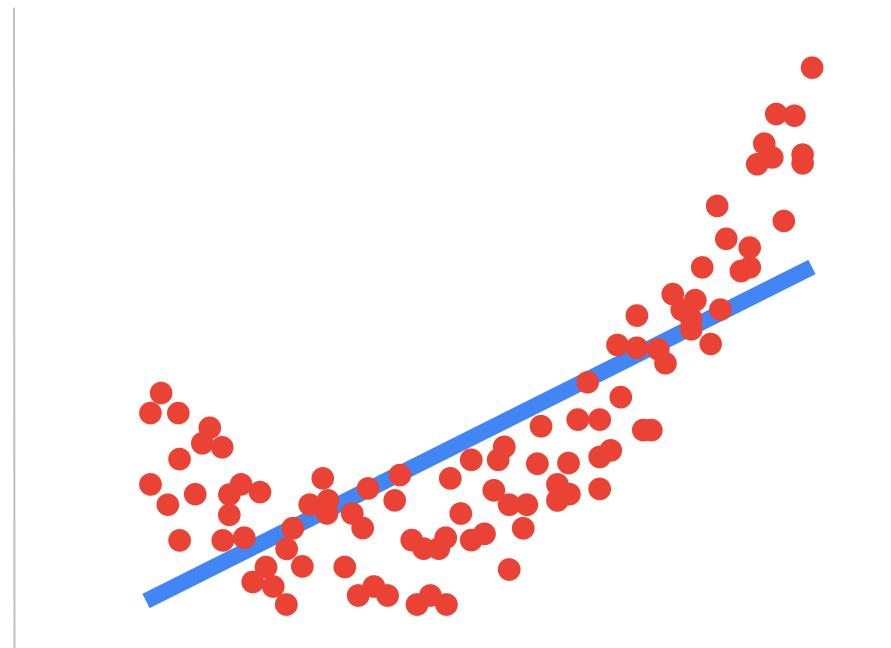
Old RMSE = 0

New RMSE = 3.2

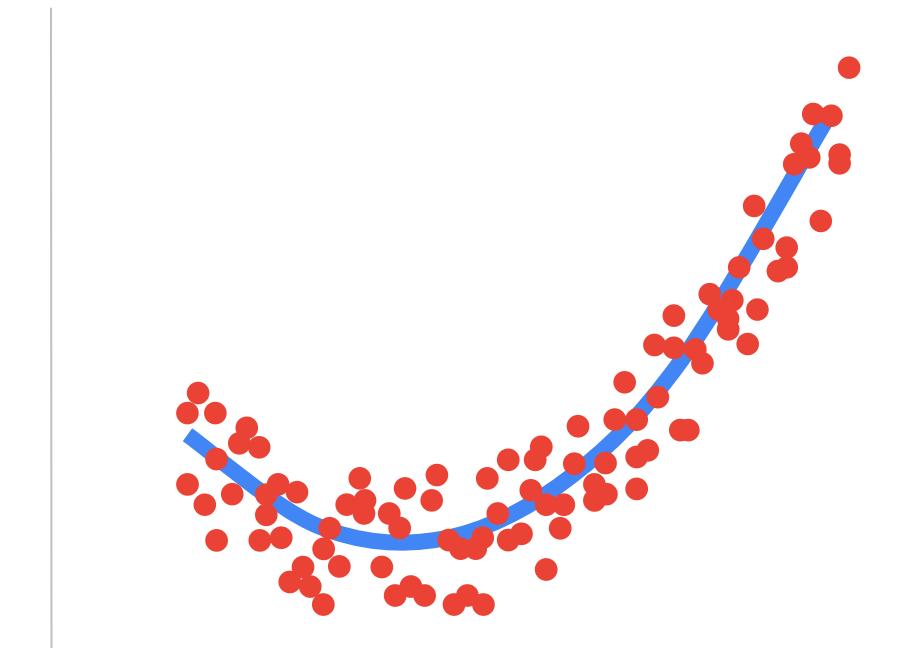
This is a red flag.



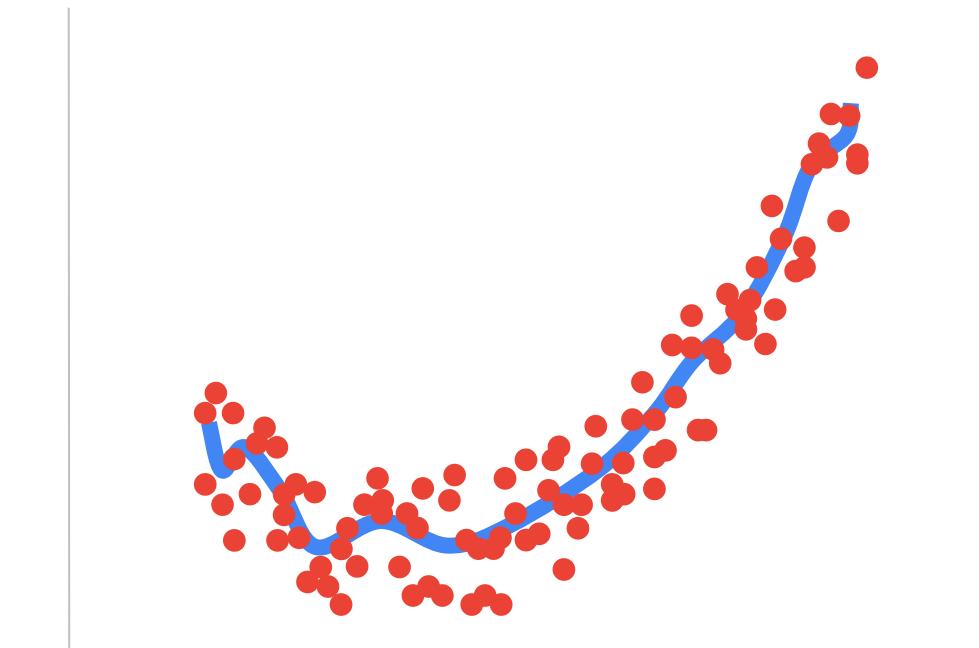
Beware of overfitting as you increase model complexity



Underfit

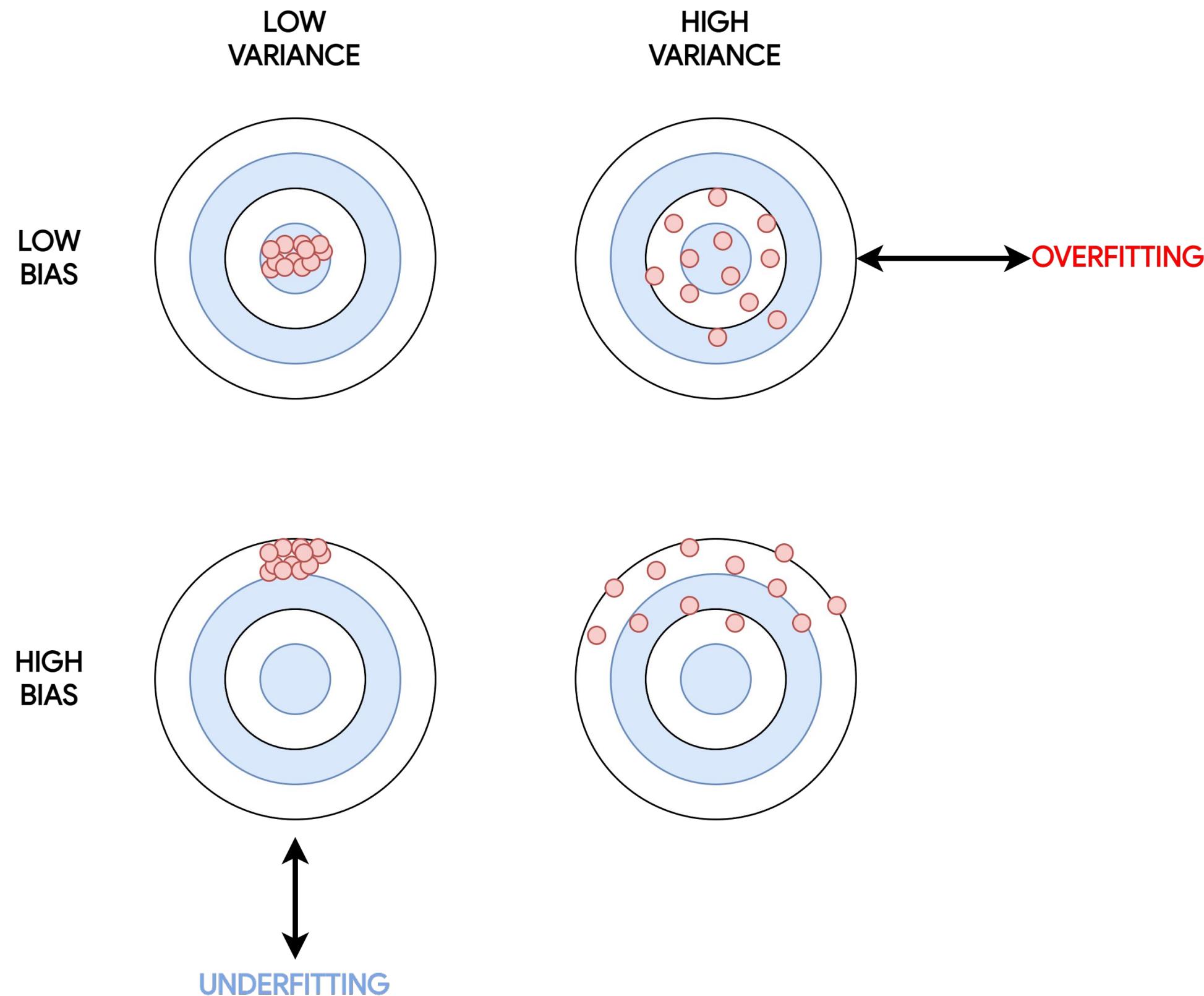


Fit

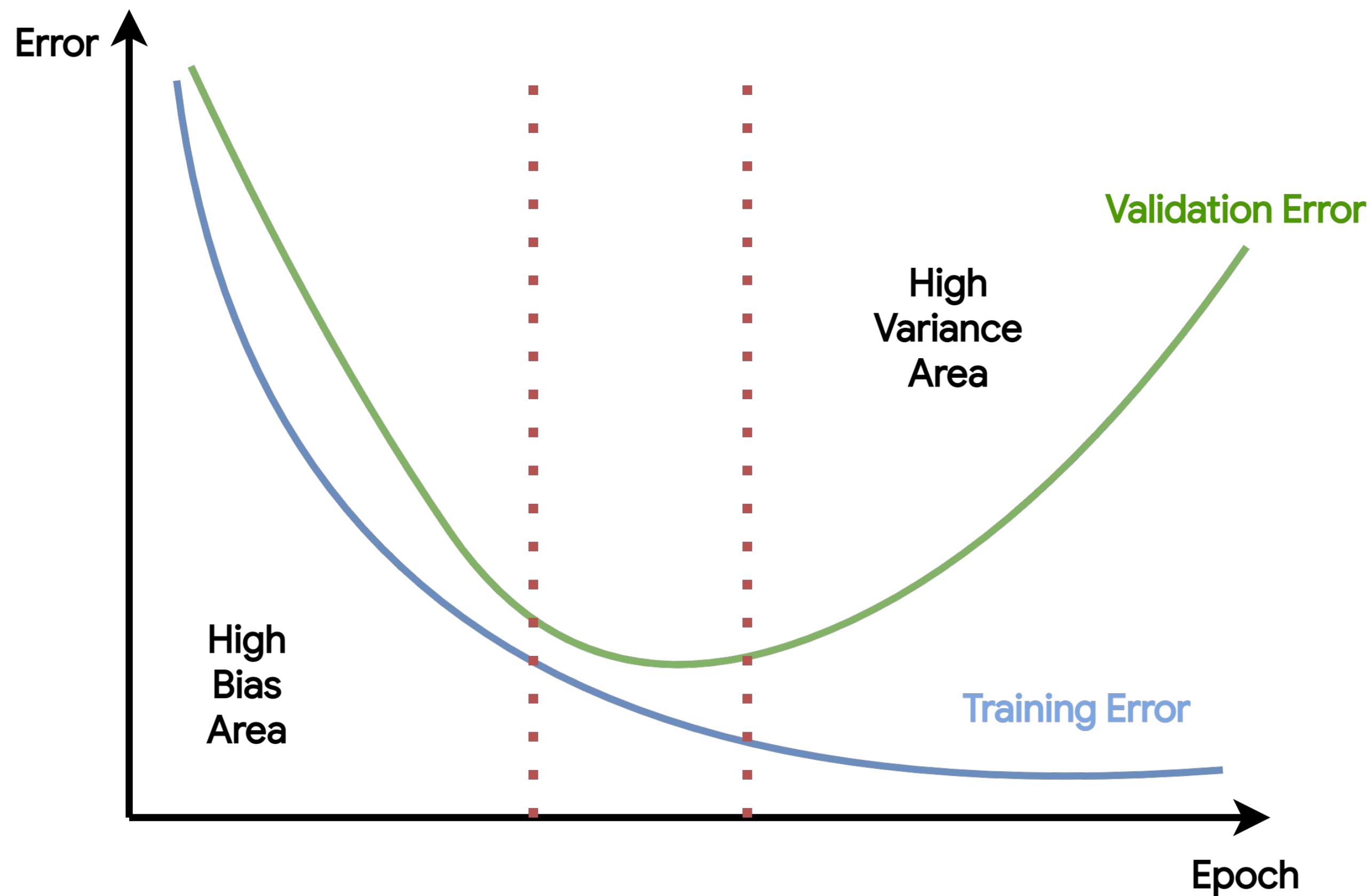


Overfit

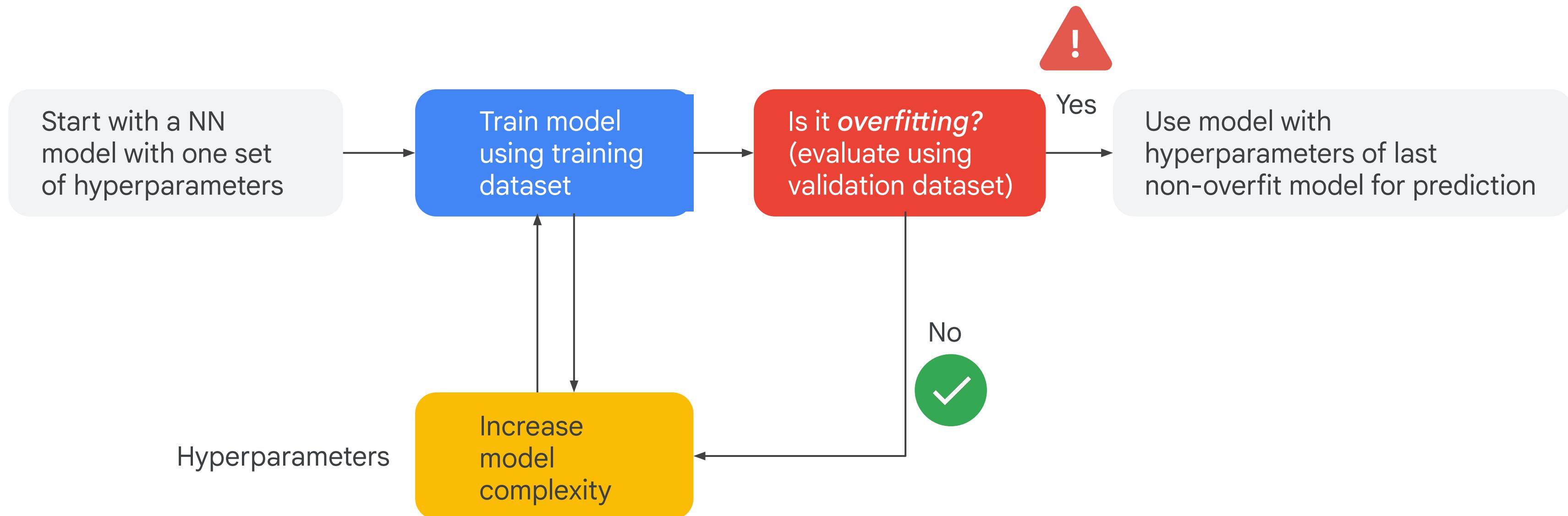
Bias VS Variance



Bias VS Variance



You can use the validation dataset to experiment with model complexity



How can we measure
model complexity?

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.

L1 regularization:

Creating a sparse model with some feature weights being exactly zero, effectively performing feature selection

L2 regularization:

prevent overfitting and distribute the weights more evenly among features

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

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

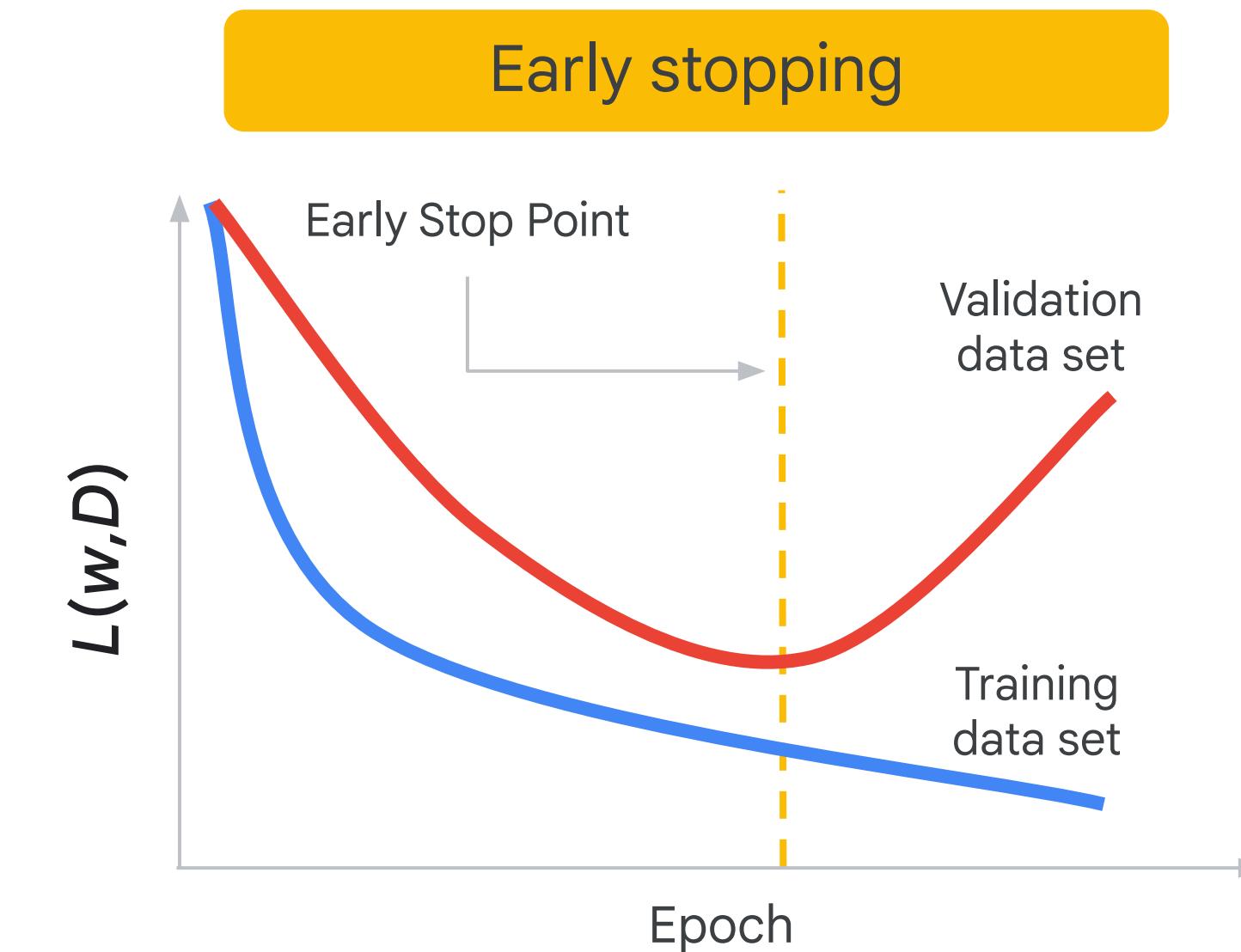
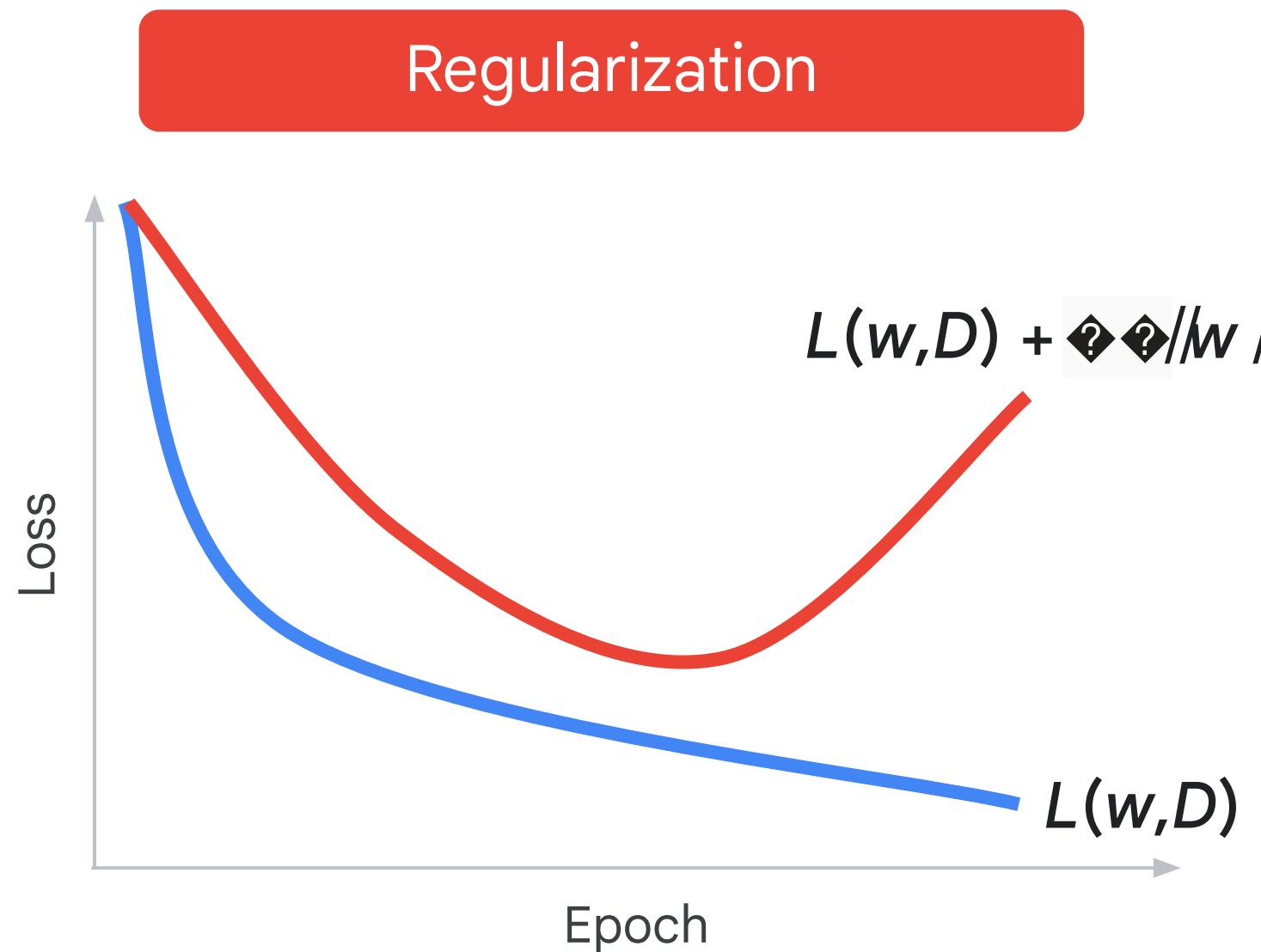
Lambda controls how
these are balanced

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

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

L1 regularization can be used as a
feature selection mechanism.

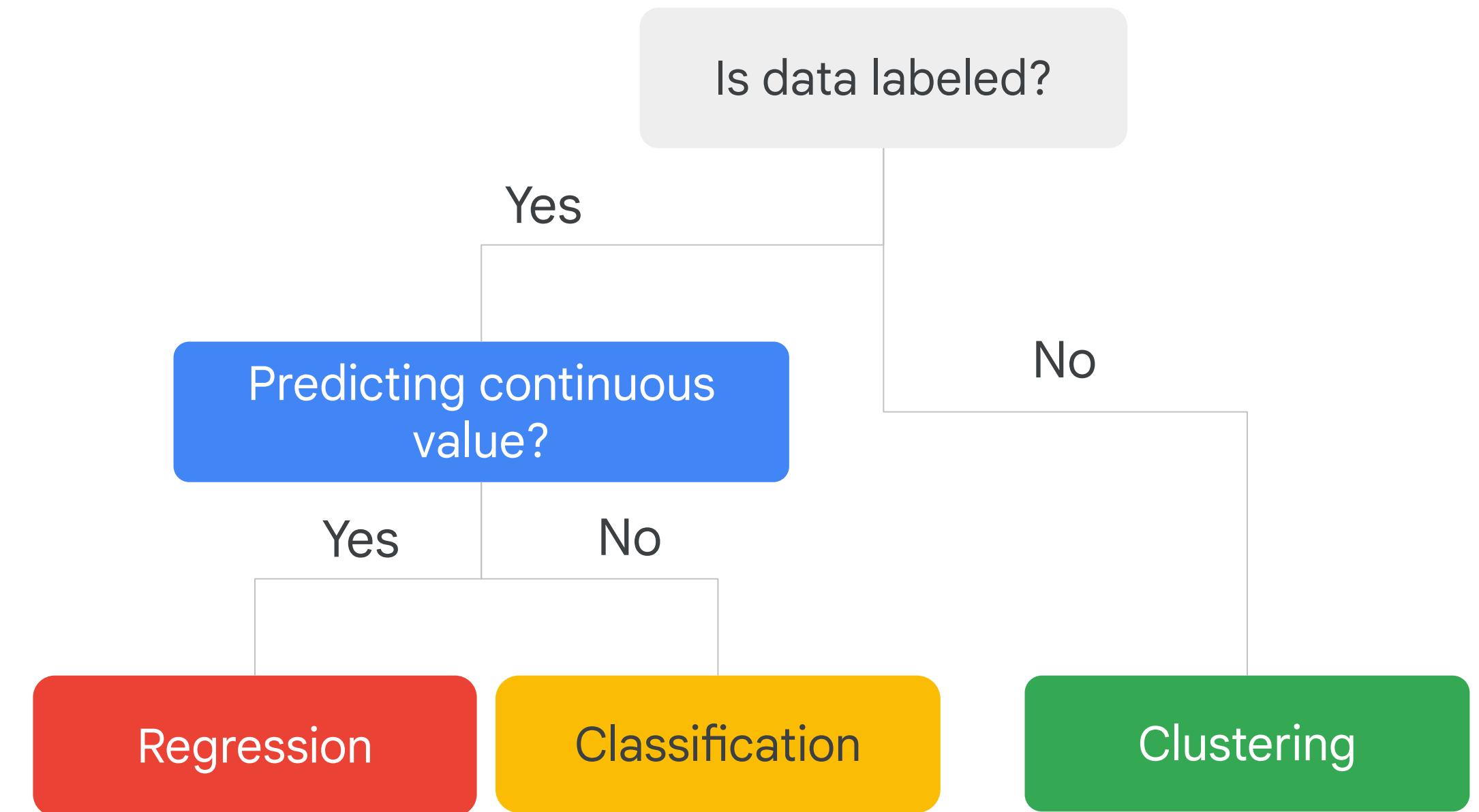
Often we do both regularization and early stopping to counteract overfitting





ML Models Overview

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

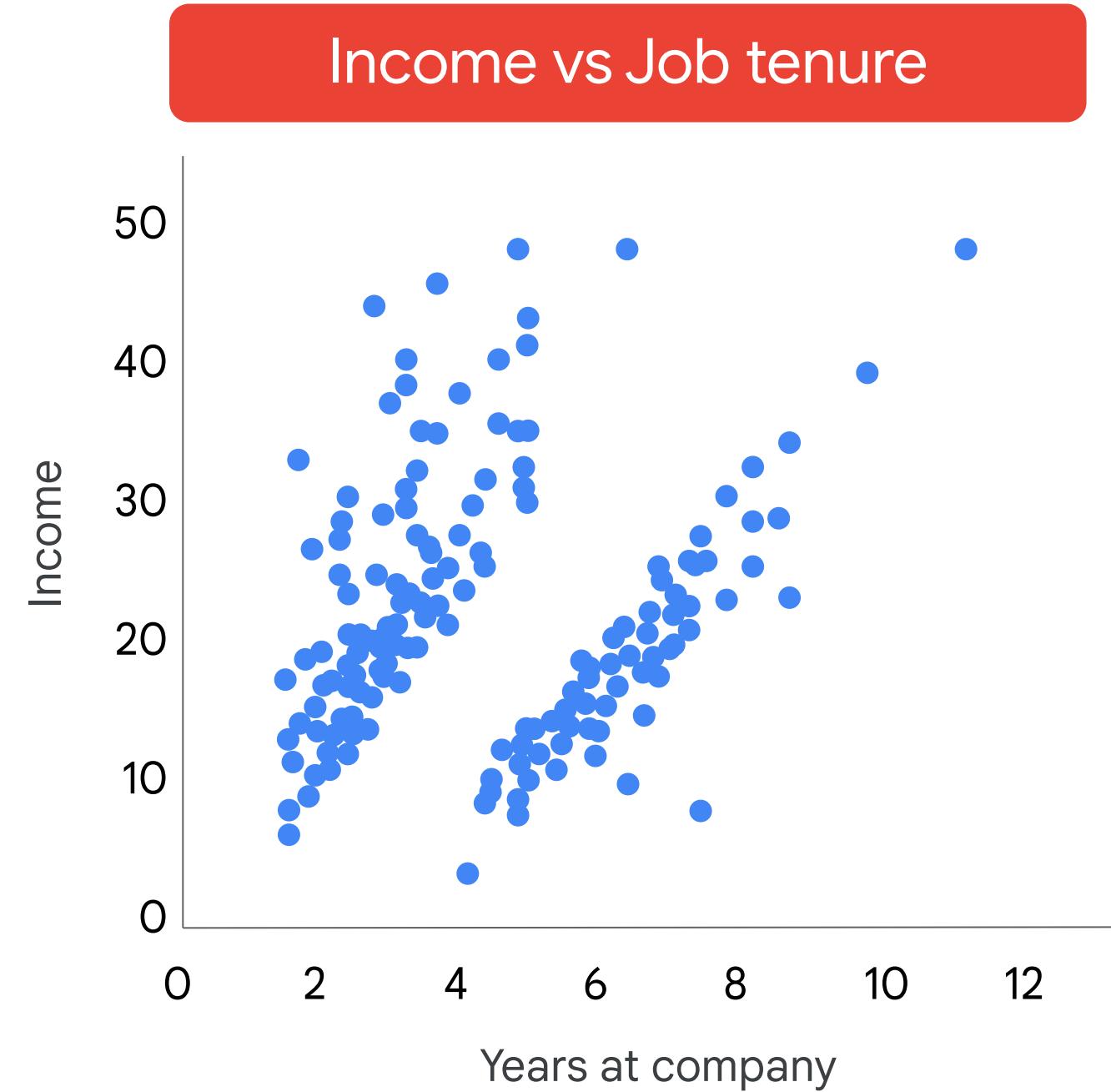


**Unsupervised and
supervised learning
are the two types of
ML algorithms**

In unsupervised learning,
data is not labeled.

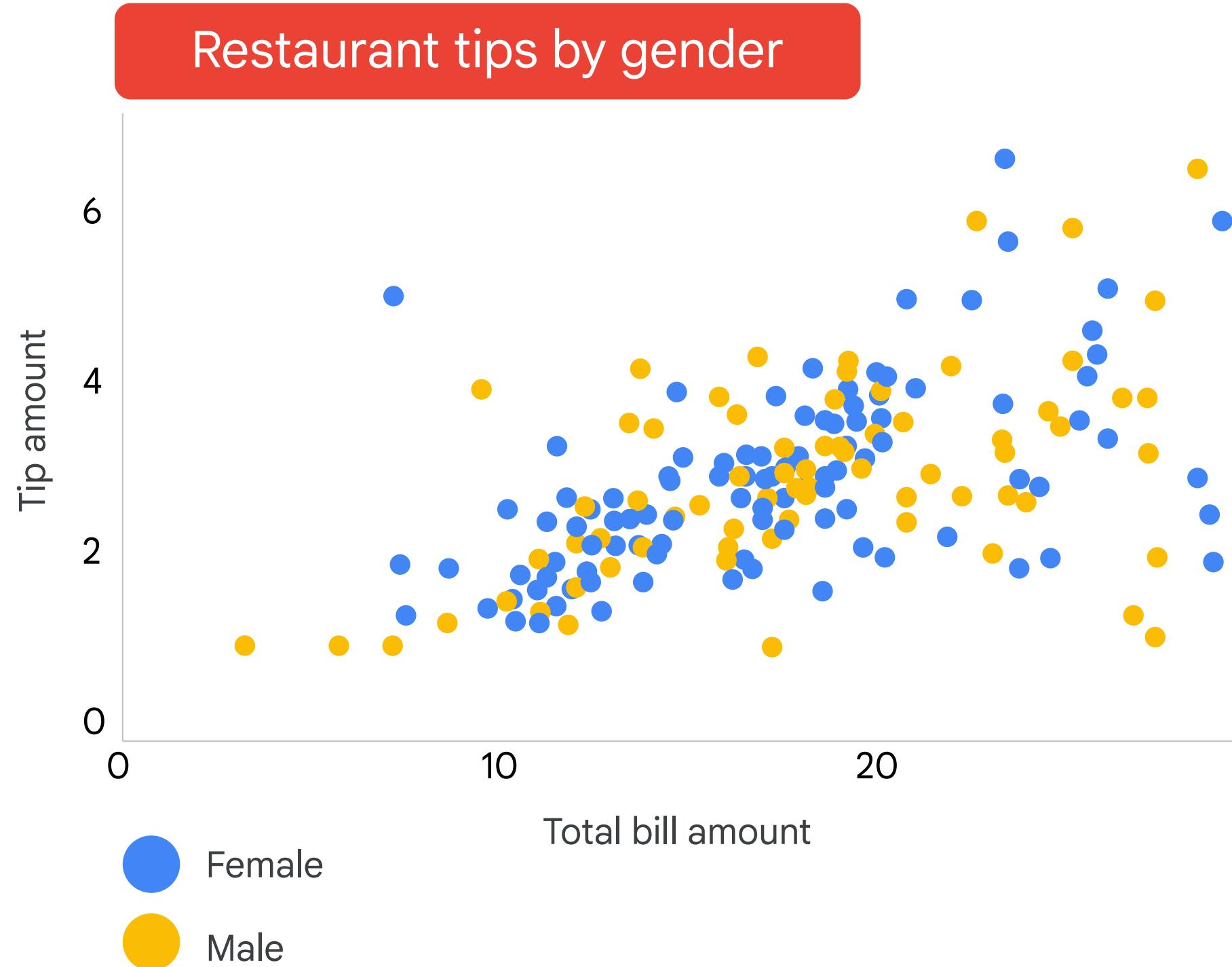
Example Model: Clustering

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



**Supervised learning
implies the data is
already labeled**

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



Supervised ML model types:

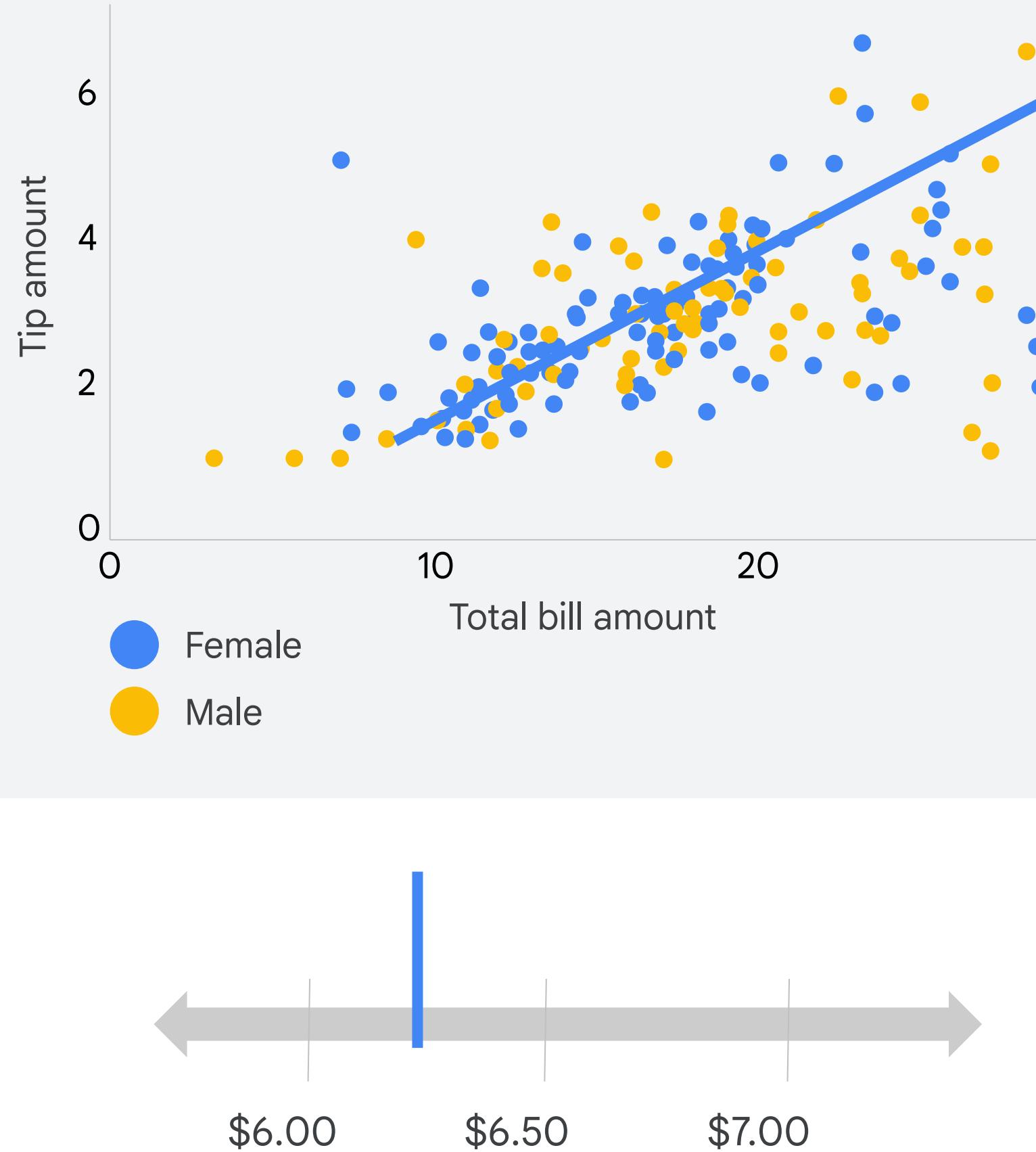
Regression and classification

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

**Option 1
Regression model**
Predict the tip amount

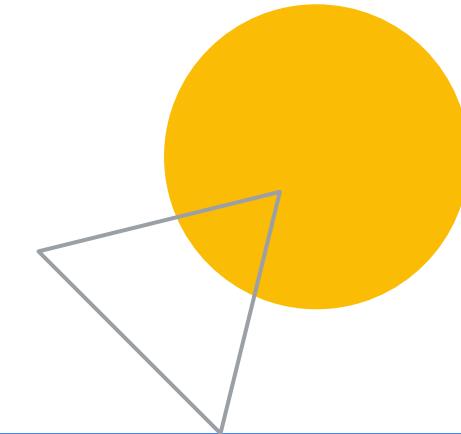
**Option 2
Classification model**
Predict the sex of the customer

Restaurant tips by gender



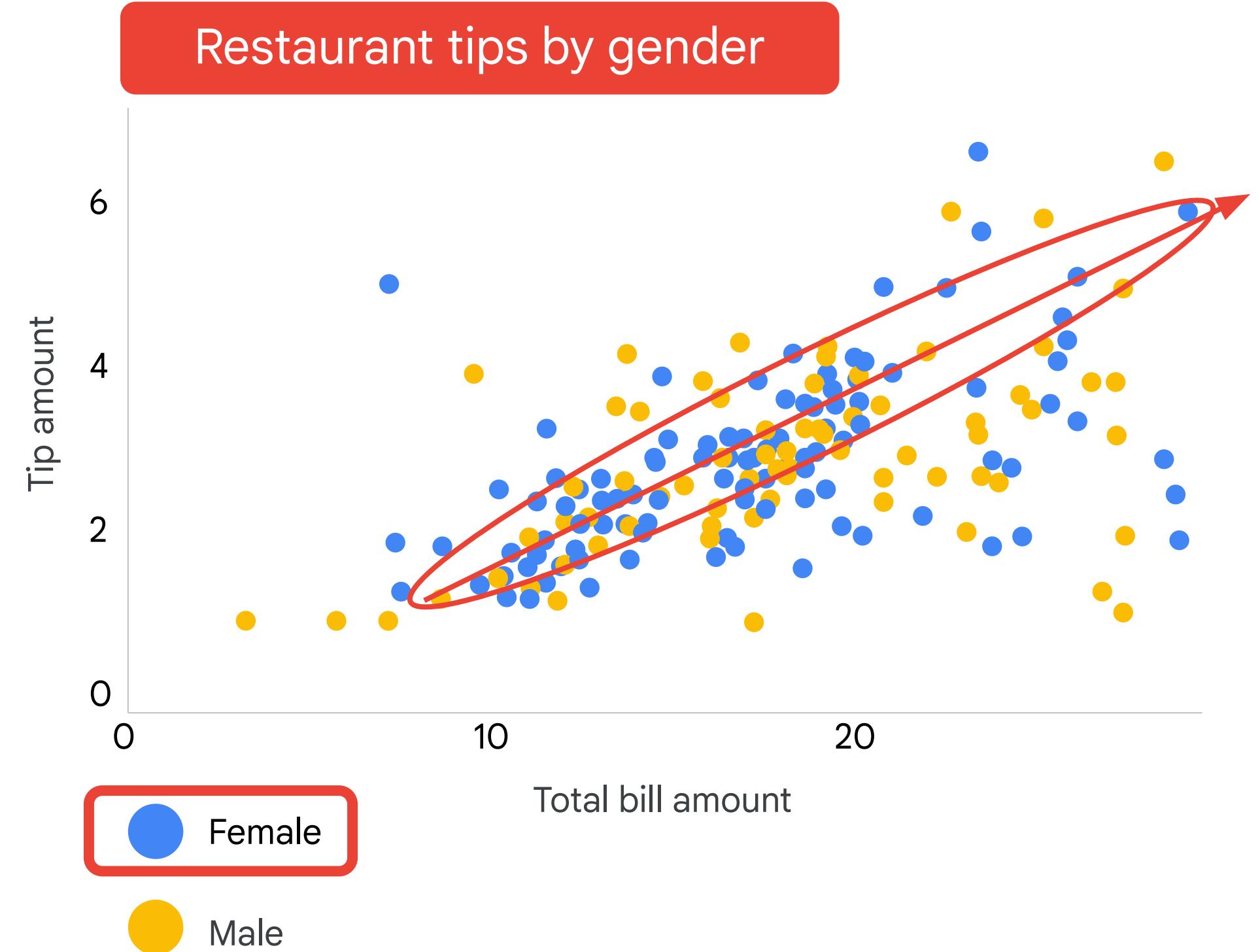
Use regression for
predicting continuous
label values

Regression model (linear) predicts
tip amount of \$6.25.

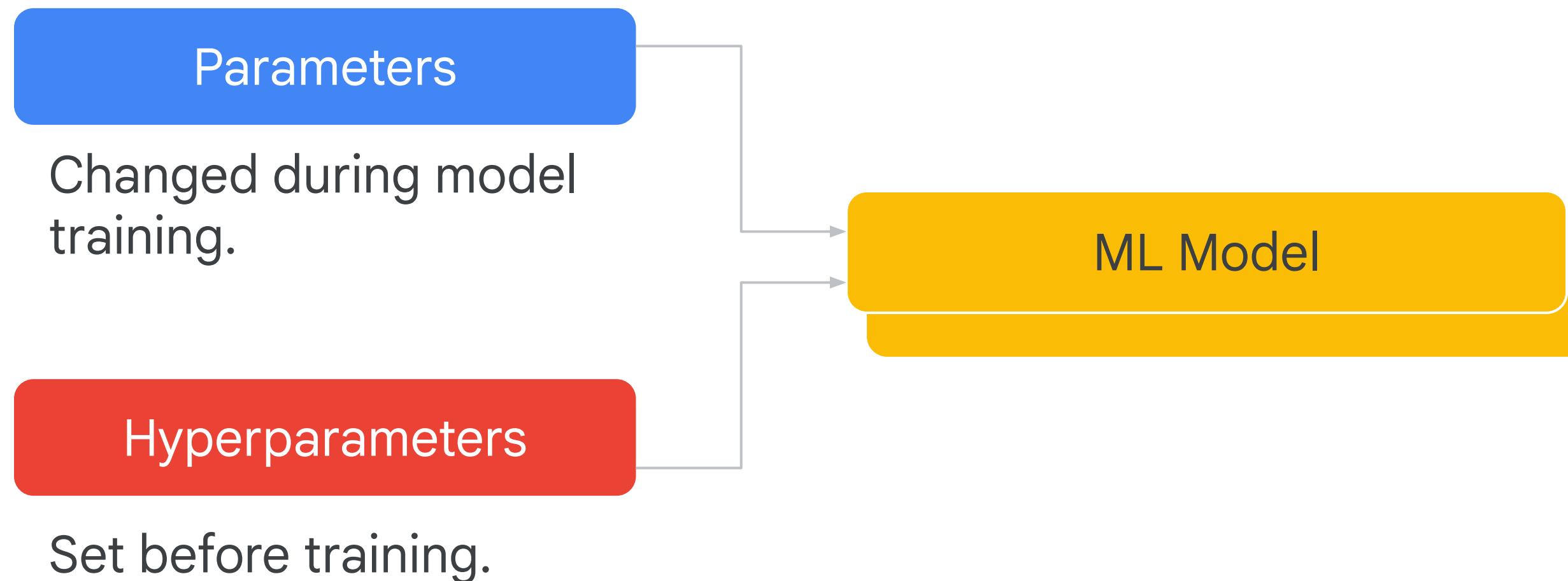


Use classification for predicting categorical label values

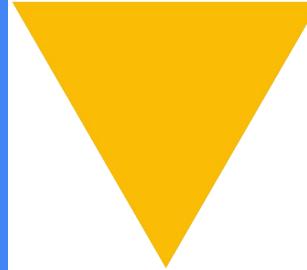
Classification model (binary) predicts category of female.



ML models are mathematical functions with parameters and hyperparameters

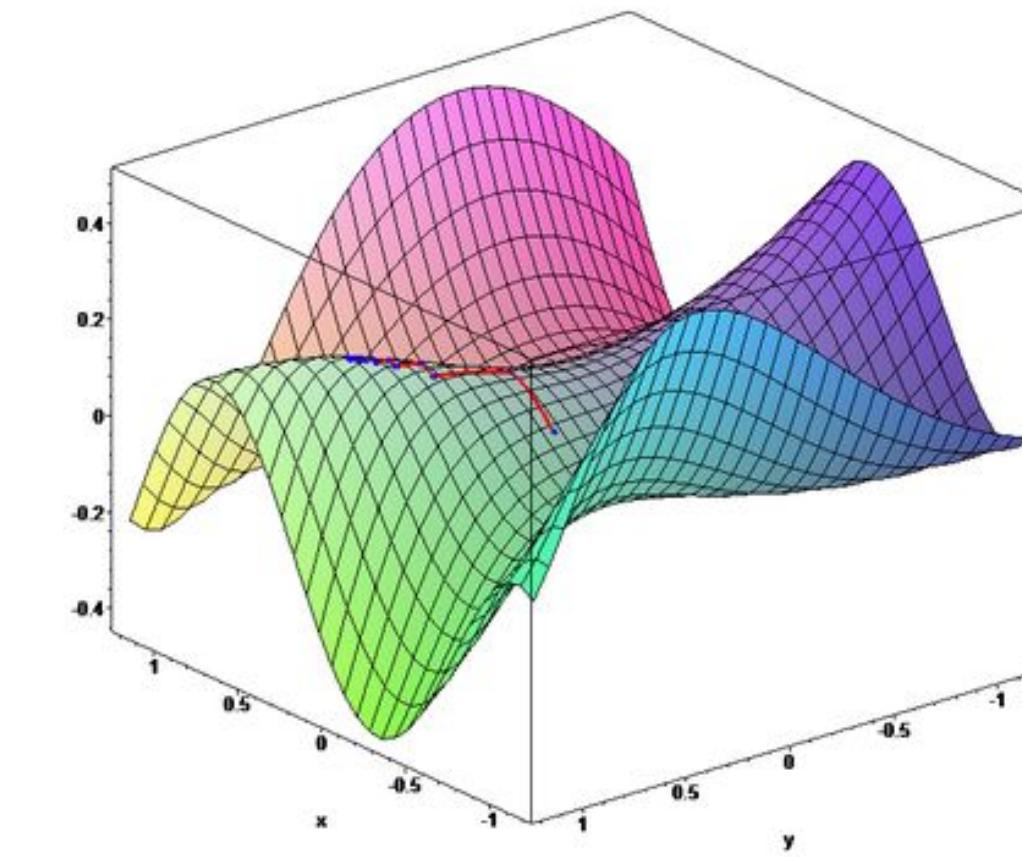
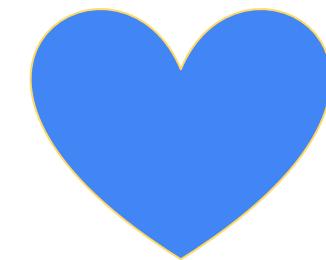


105



Model Optimization

Stochastic Gradient Descent



Optimizers in Tensor Flow... Stochastic Gradient Descent (SGD)

`tf.keras.optimizers.SGD`

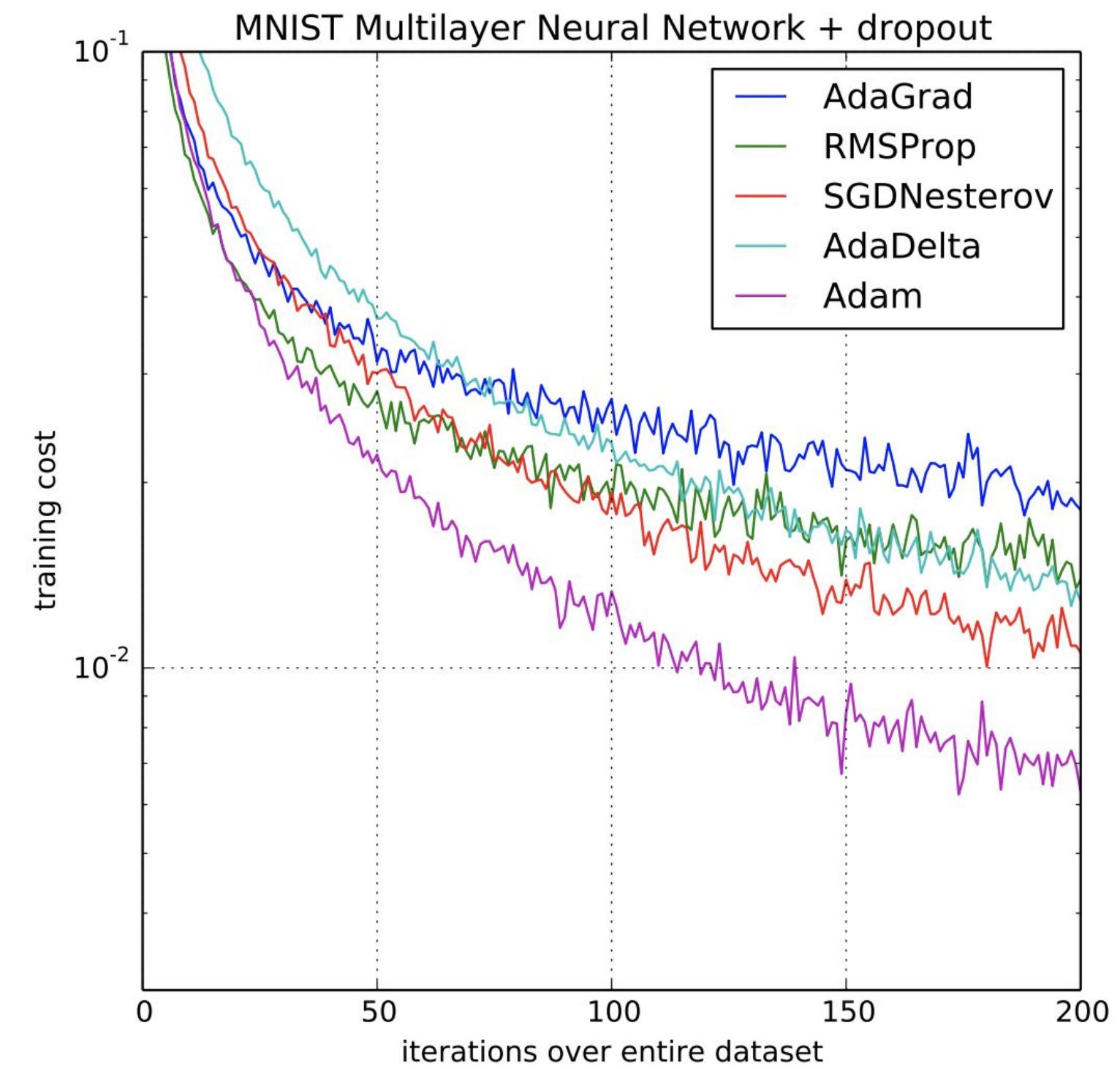
- Implements Gradient Descent with Momentum
- Parameters can be
 - Learning Rate => $w(i+1) = w(i) - LR^*G$
 - Momentum => $velocity(i+1) = velocity(i)^*momentum - LR^*G$
 $w(i+1) = w(i) + velocity(i+1)$
- Only one parameter is updated at each time.

What if we want to update multiple parameter w_1, w_2, \dots
at each time

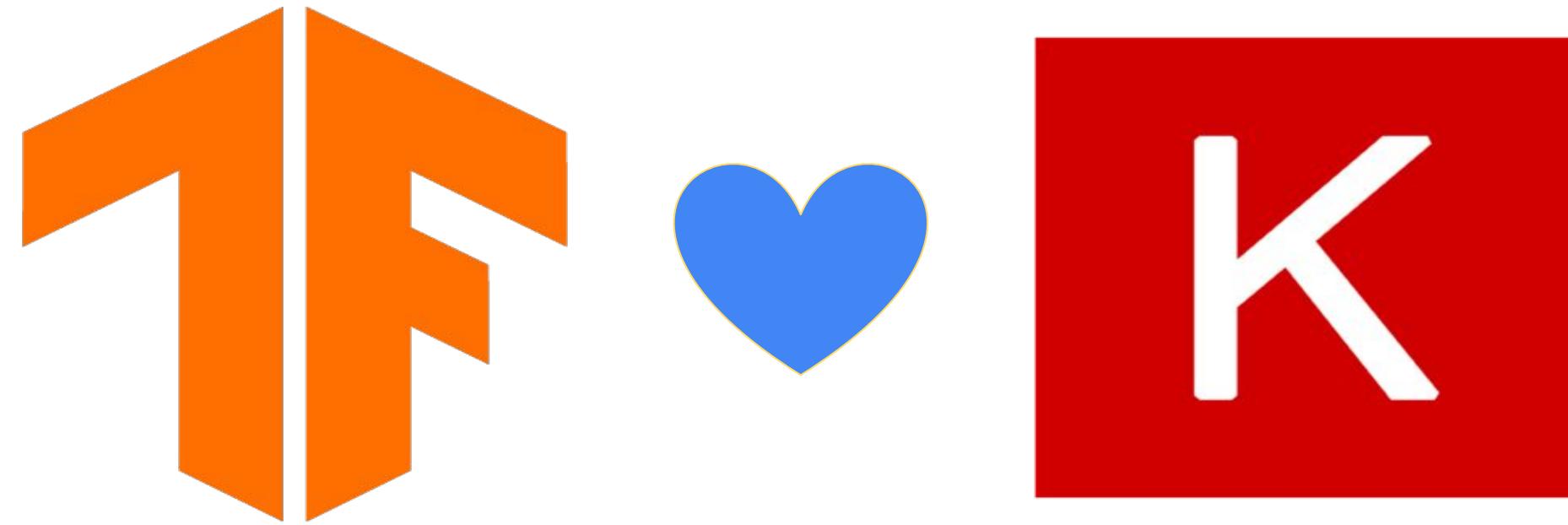
Optimizers in Tensor Flow... Adaptives Gradient Optimizers

tf.keras.optimizers.Adam

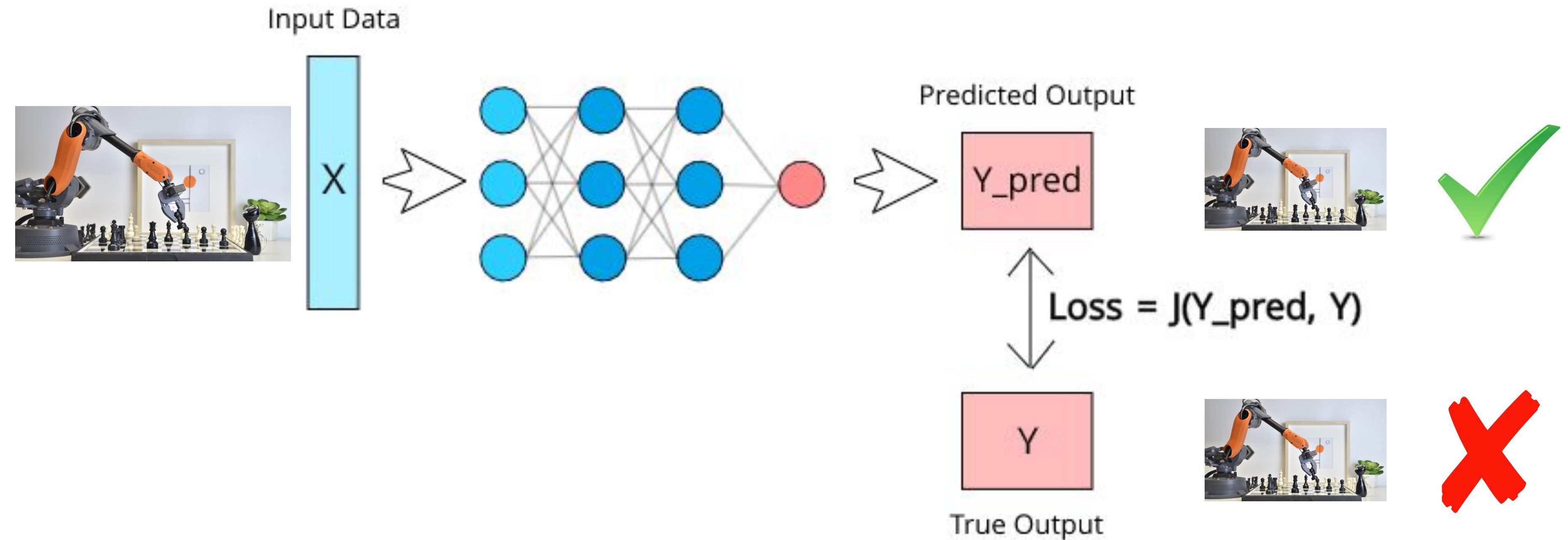
- **Adam** adds momentum to **Adadelta** thus improving convergence speed for every parameters
- **Adam** is commonly used optimizer for training process of large dataset because its reduced training cost



Loss functions in TensorFlow



Loss functions in TensorFlow



Loss functions in Tensor Flow... Categorical (1)

`tf.keras.losses.BinaryCrossentropy`

- Compute cross-entropy loss for binary (0,1) classification problems

```
# Example 1: (batch_size = 1, number of samples = 4)
```

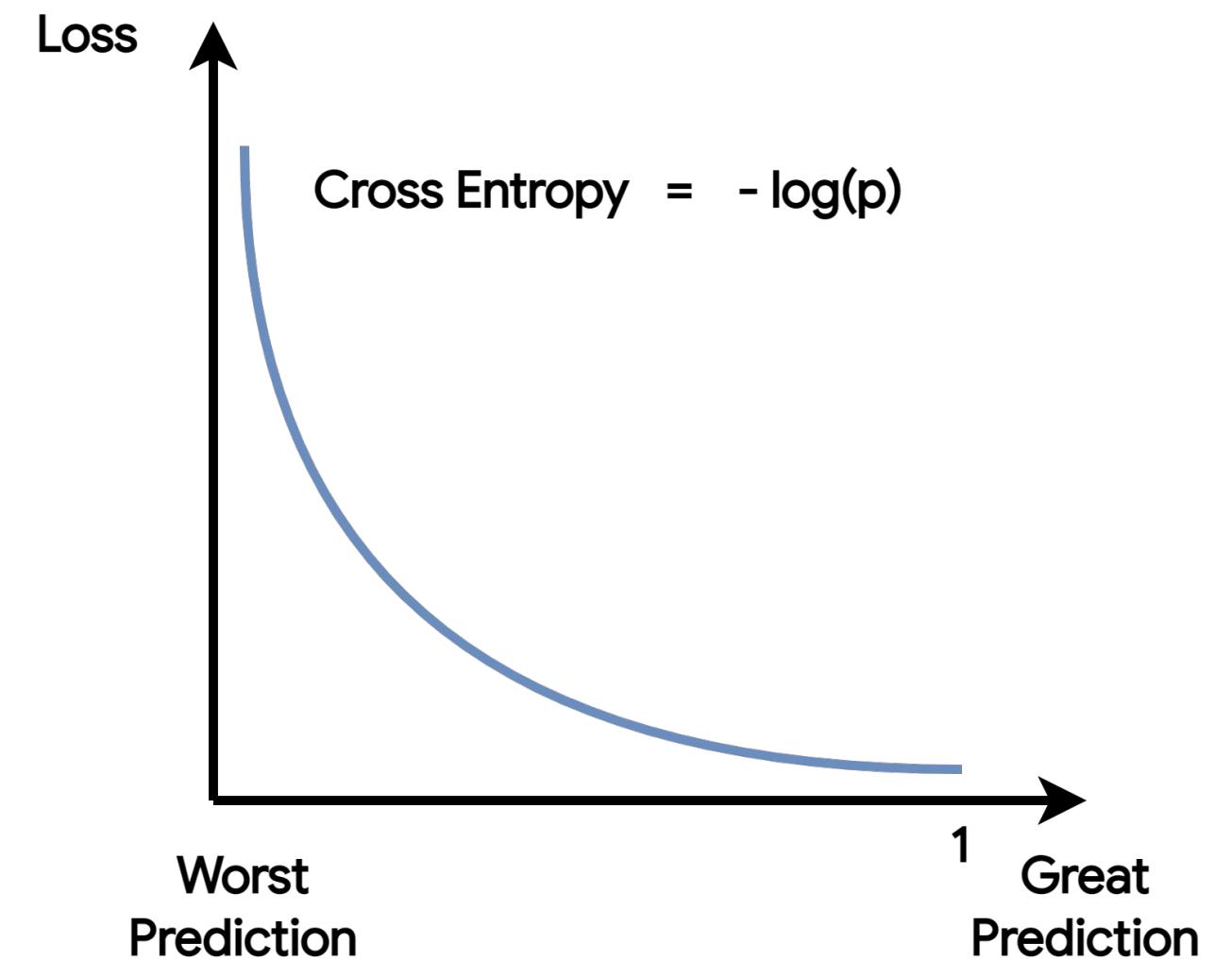
```
y_true = [0, 1, 0, 0]
```

```
y_pred = [-18.6, 0.51, 2.94, -12.8]
```

```
bce = tf.keras.losses.BinaryCrossentropy(from_logits=True)
```

```
bce(y_true, y_pred).numpy()
```

```
0.865
```



Loss functions in Tensor Flow... Categorical (2)

`tf.keras.losses.CategoricalCrossentropy`

- Compute cross-entropy loss when there are with 2 or more label classes.
- **Labels** need to be one-hot encoded otherwise use `SparseCategoricalCrossentropy`

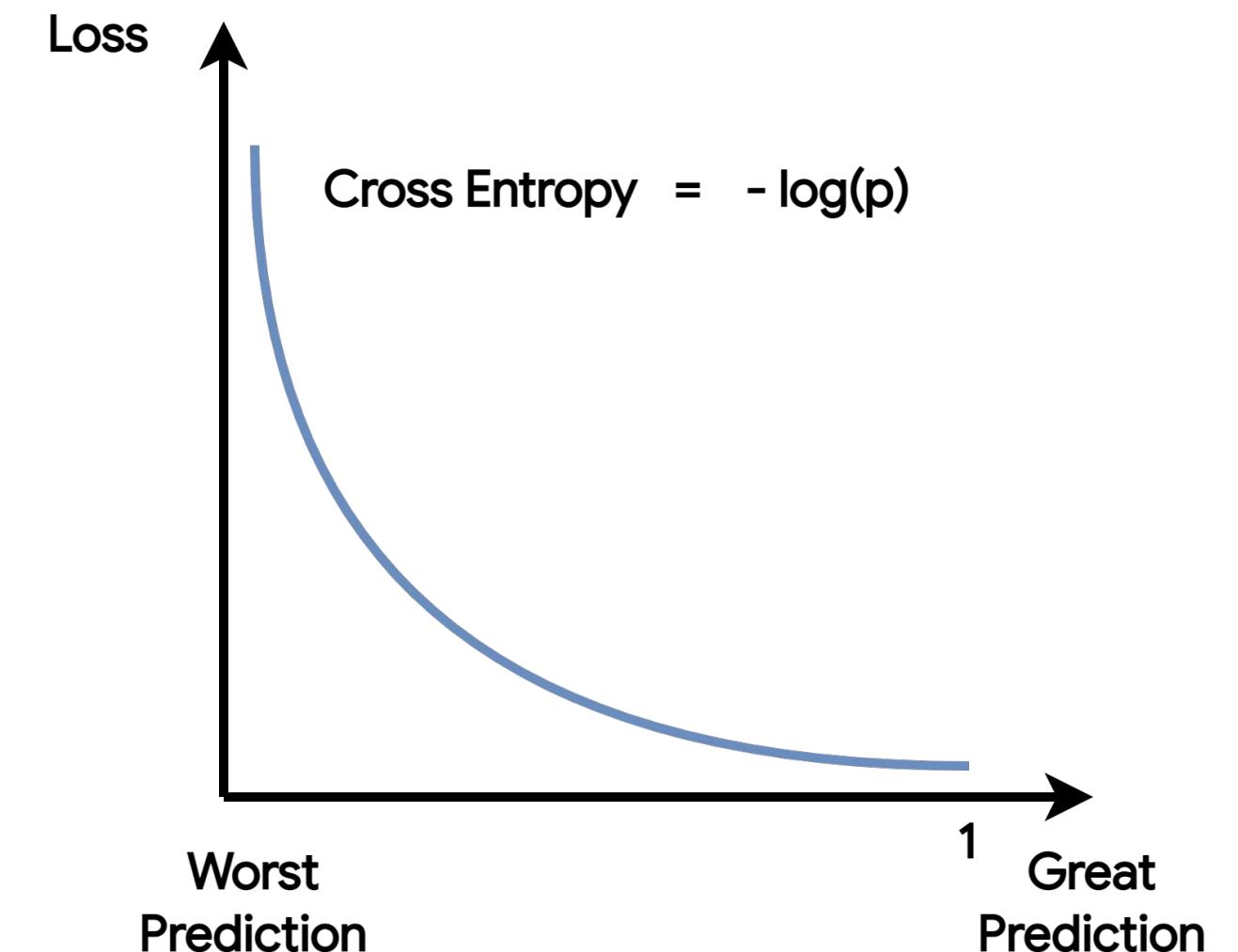
`y_true = [[0, 1, 0], [0, 0, 1]]`

`y_pred = [[0.05, 0.95, 0], [0.1, 0.8, 0.1]]`

`cce = tf.keras.losses.CategoricalCrossentropy()`

`cce(y_true, y_pred).numpy()`

1.117



Loss functions in Tensor Flow... Categorical (3)

`tf.keras.losses.SparseCategoricalCrossentropy`

- Similar to the `CategoricalCrossentropy` but **labels** are expected to be integers

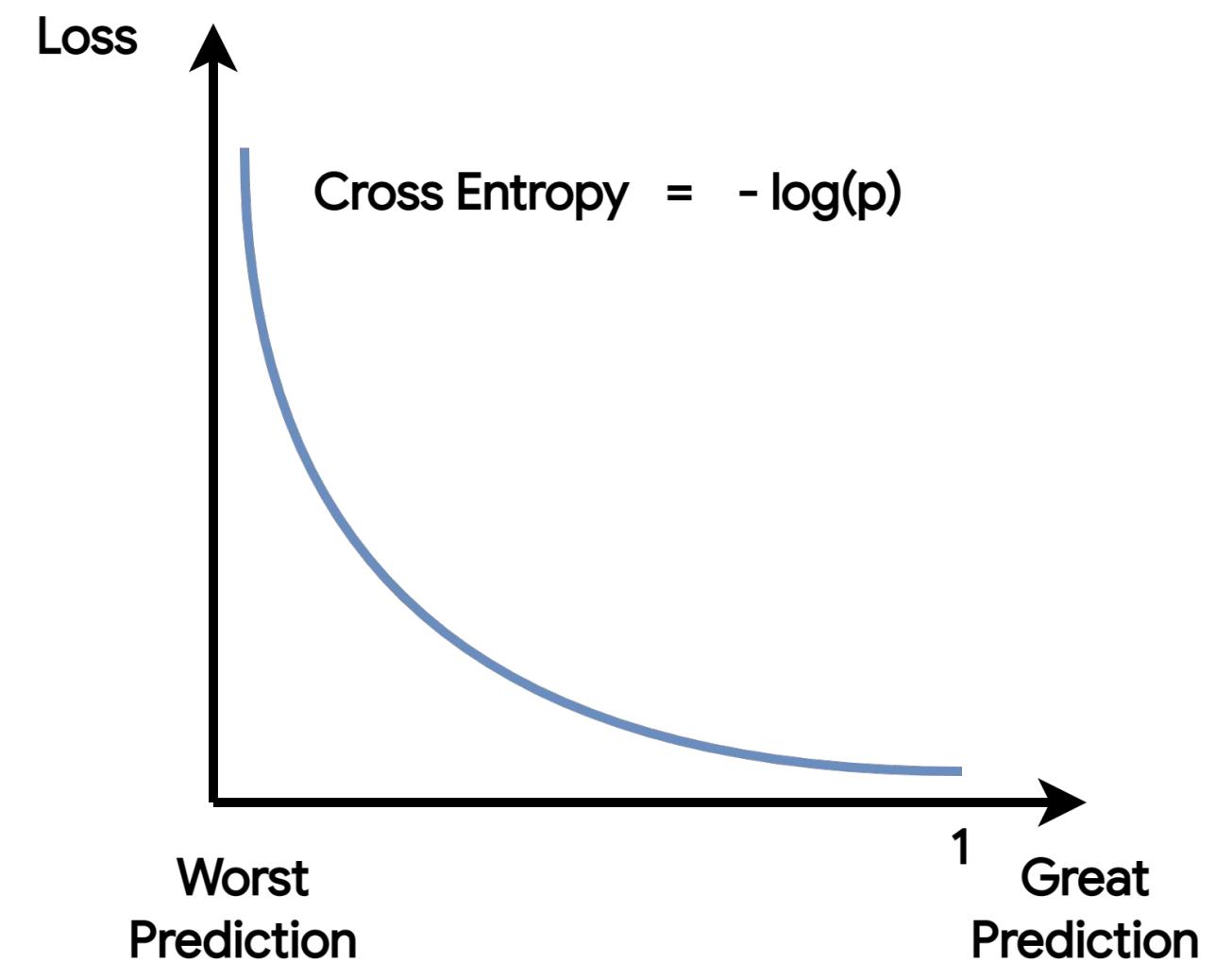
`y_true = [1, 2]`

`y_pred = [[0.05, 0.95, 0], [0.1, 0.8, 0.1]]`

`scce = tf.keras.losses.SparseCategoricalCrossentropy()`

`scce(y_true, y_pred).numpy()`

1.117



Compose a loss function by calculating errors

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

Error = actual (true) - predicted value

Compute the errors:

+0.70

+1.10

+0.65

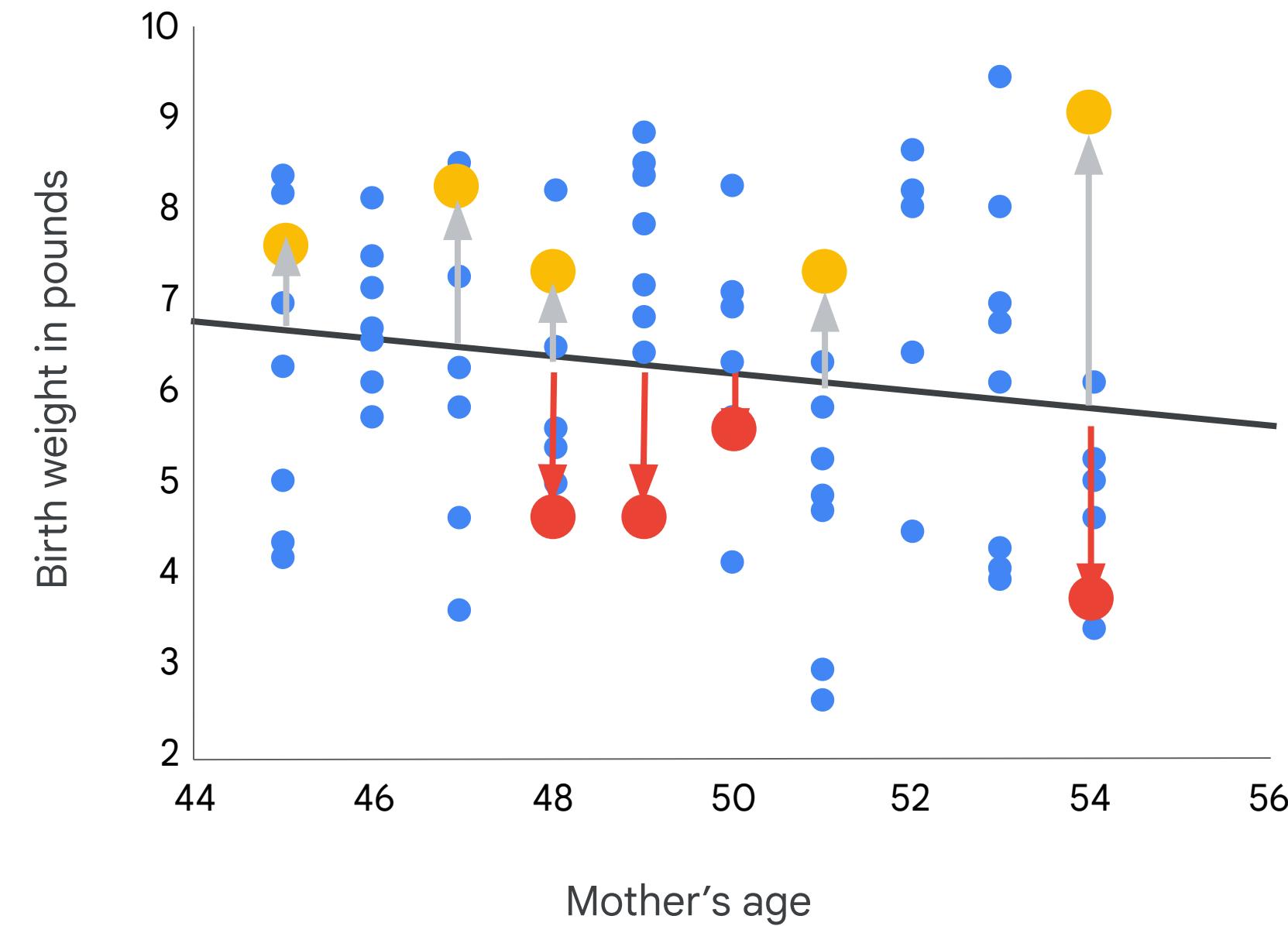
-1.20

-1.15

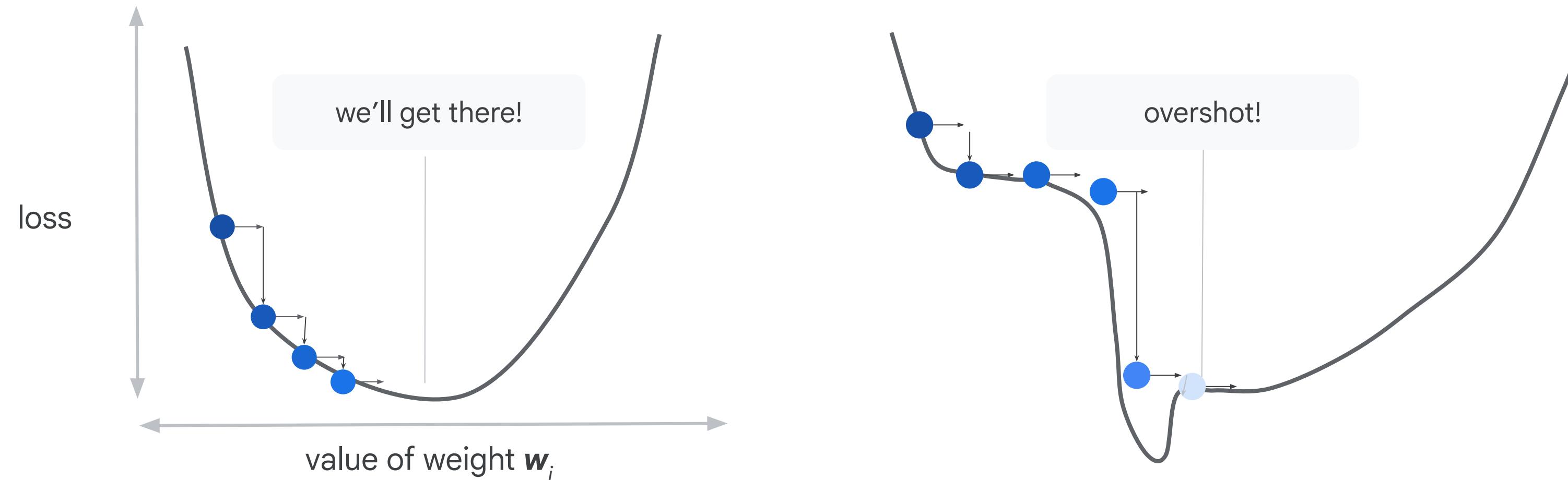
+1.10

+3.09

-2.10



A correct and constant step size (Learning Rate) can be difficult to find



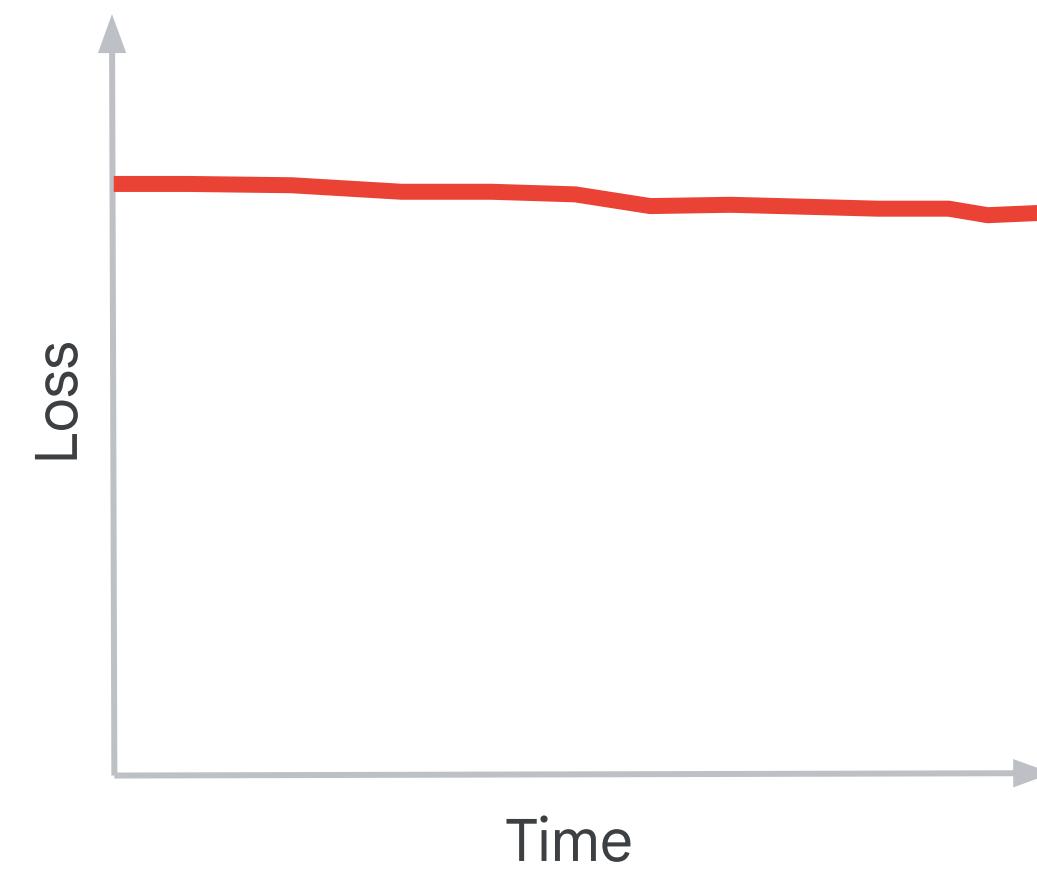
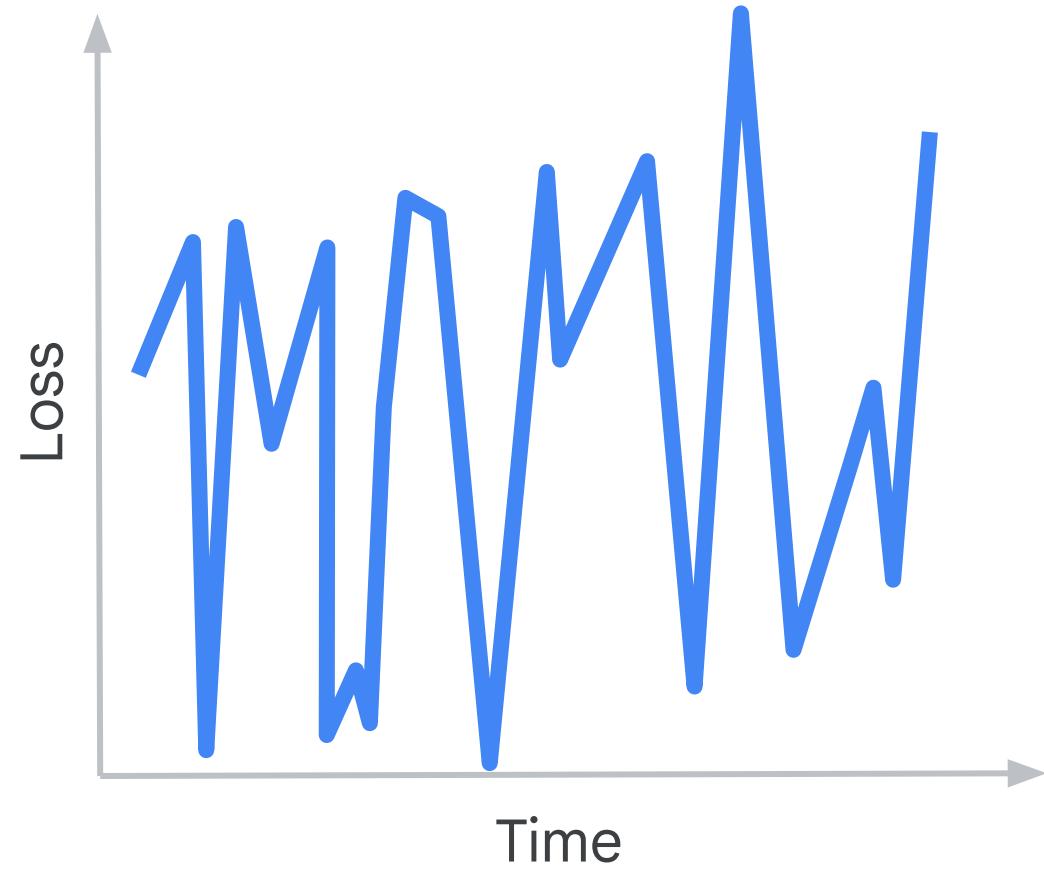
Step size or “learning rate” is a hyperparameter which is set before training.

One size does not fit all models.

A typical loss curve



Troubleshooting a loss curve



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()
```

Checking loss with reduced frequency

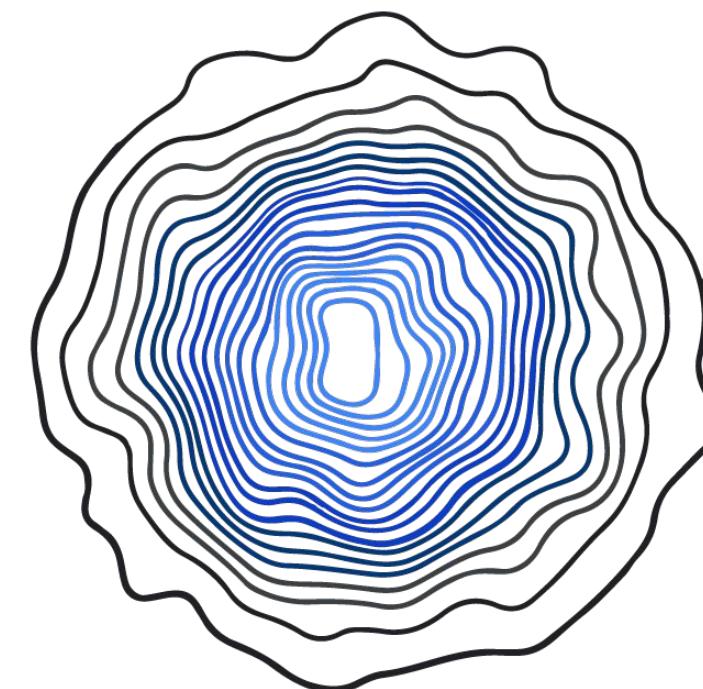
```
while loss is > Epsilon:  
    derivative = computeDerivative()  
    for i in range(self.params):  
        self.params[i] = //  
            self.params[i] //  
                - learning_rate //  
                    * derivative[i]  
    loss = computeLoss()
```

Popular implementations for readyToUpdateLoss():

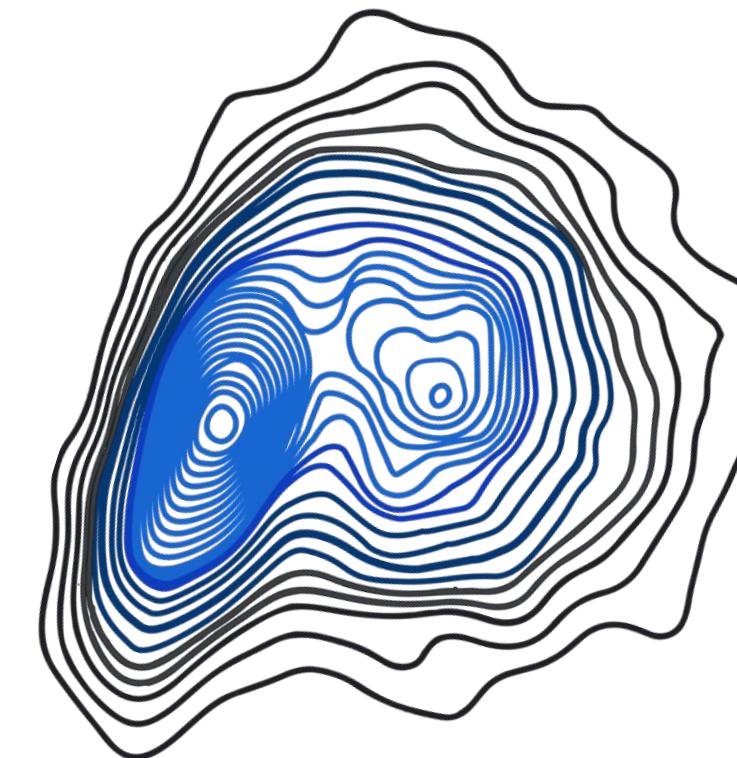
- Time-based (e.g., every hour)
- Step-based (e.g., every 1000 steps)

Problem: My model changes every time I retrain it

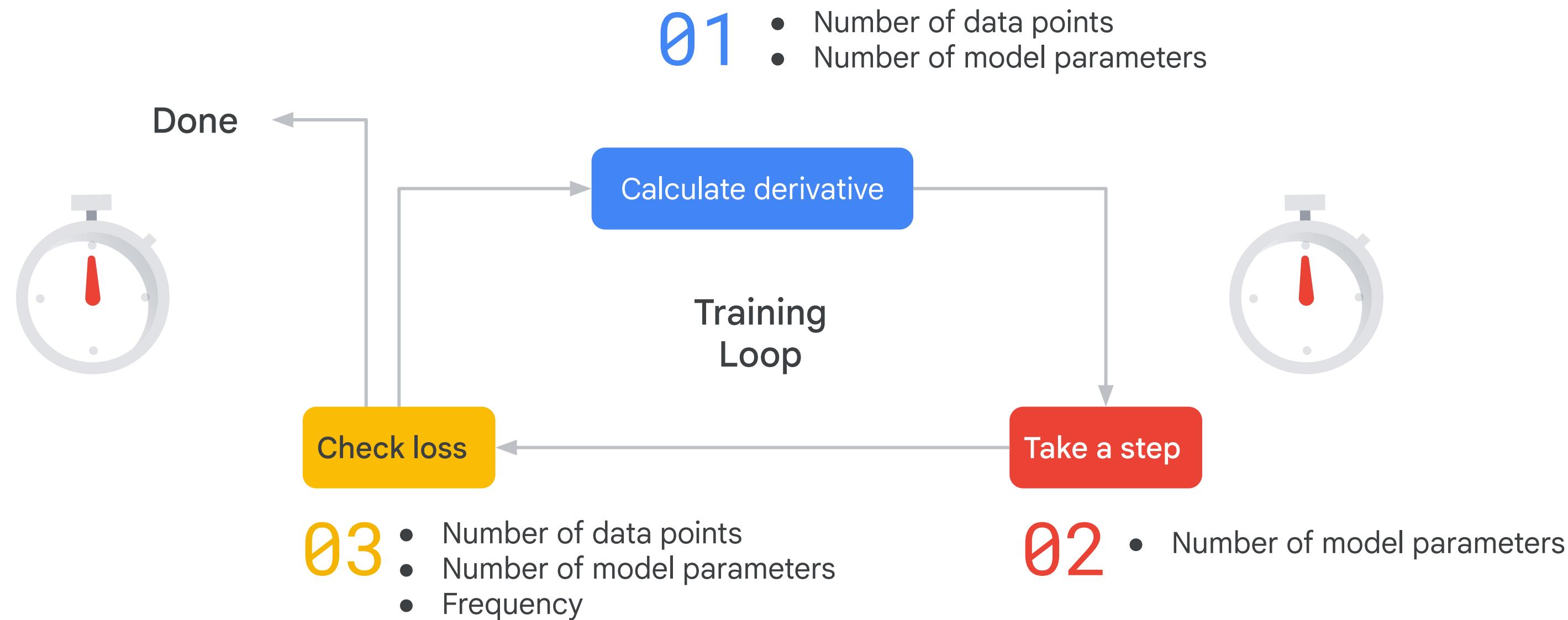
Loss surface with a
global minimum



Loss surface with more
than one minima



Problem: Model training is still too slow

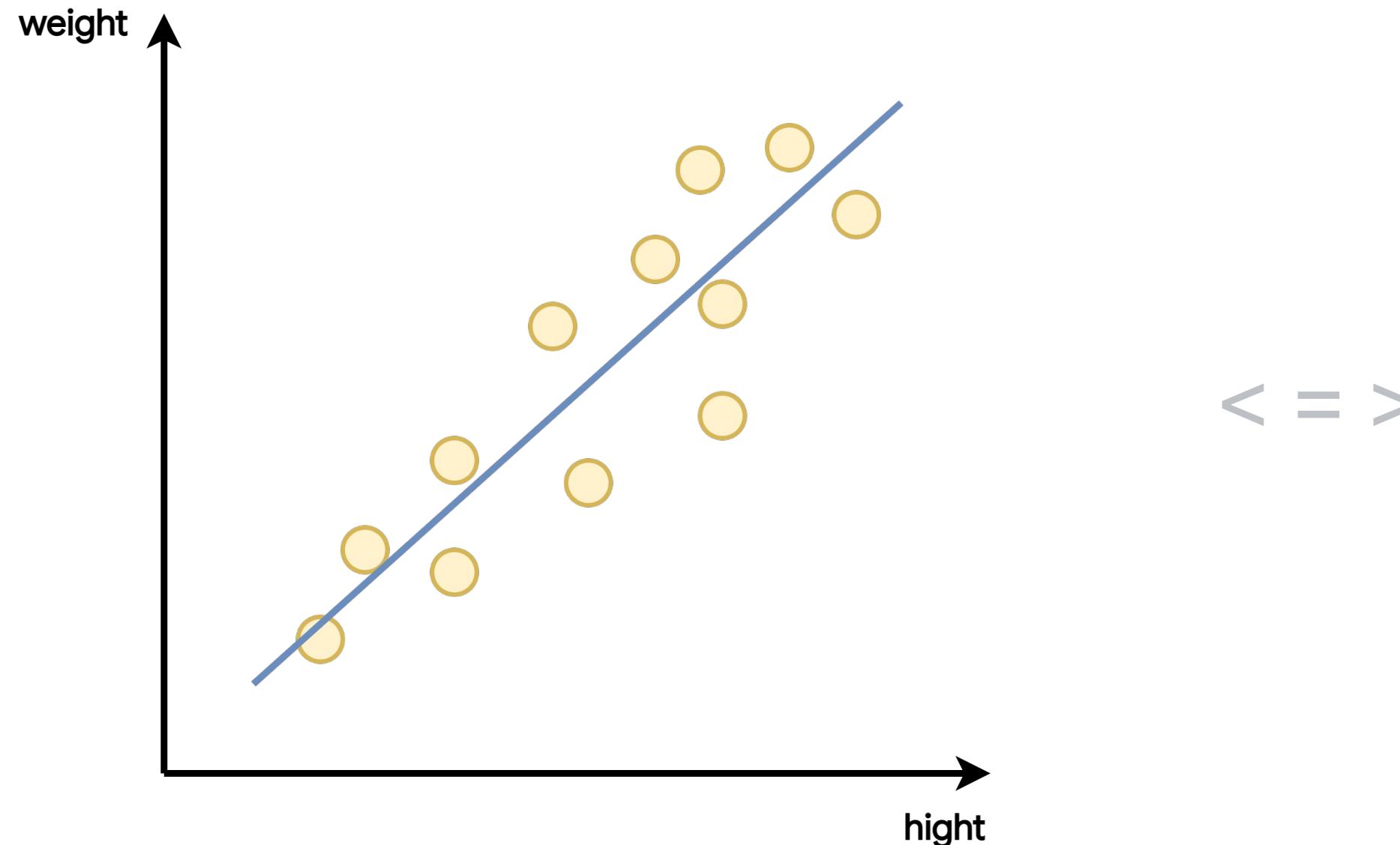




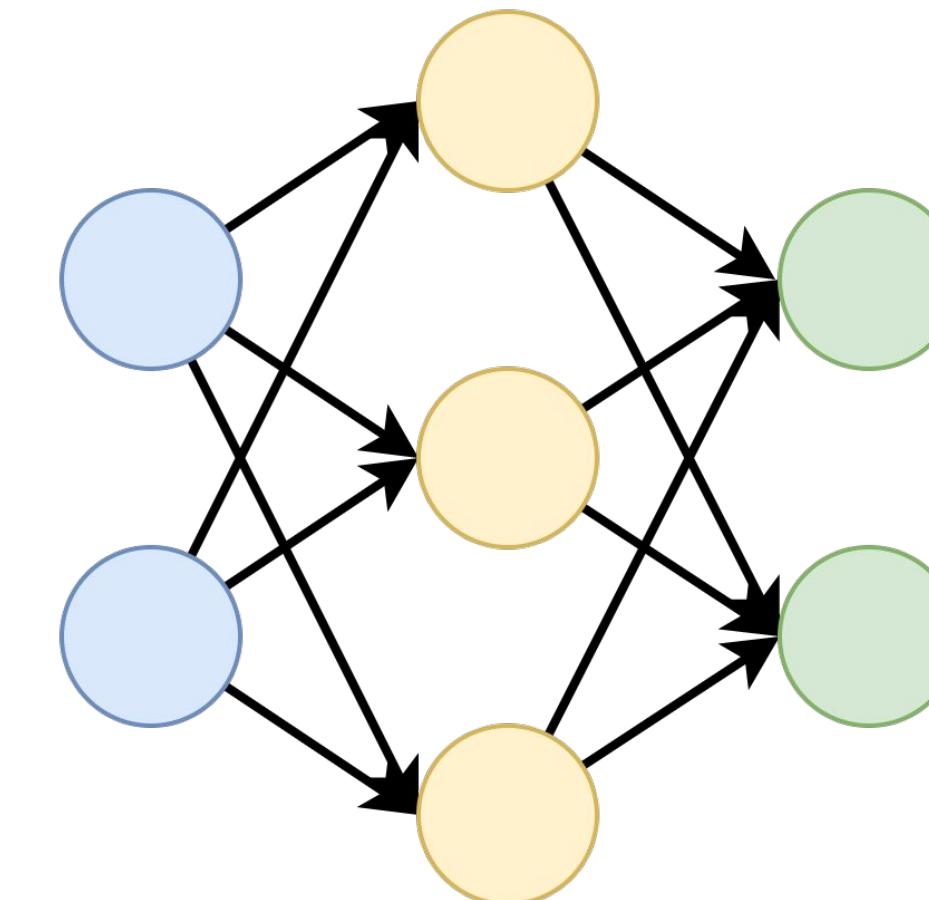
Model Metrics

RMSE: Regression problem is easy

CNN:



\leq



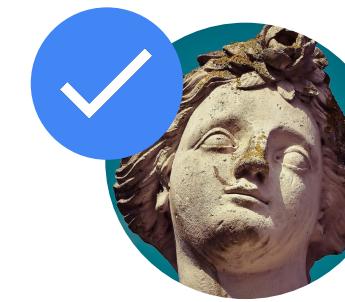
The lower RMSE wins (pss... careful not overfitting)



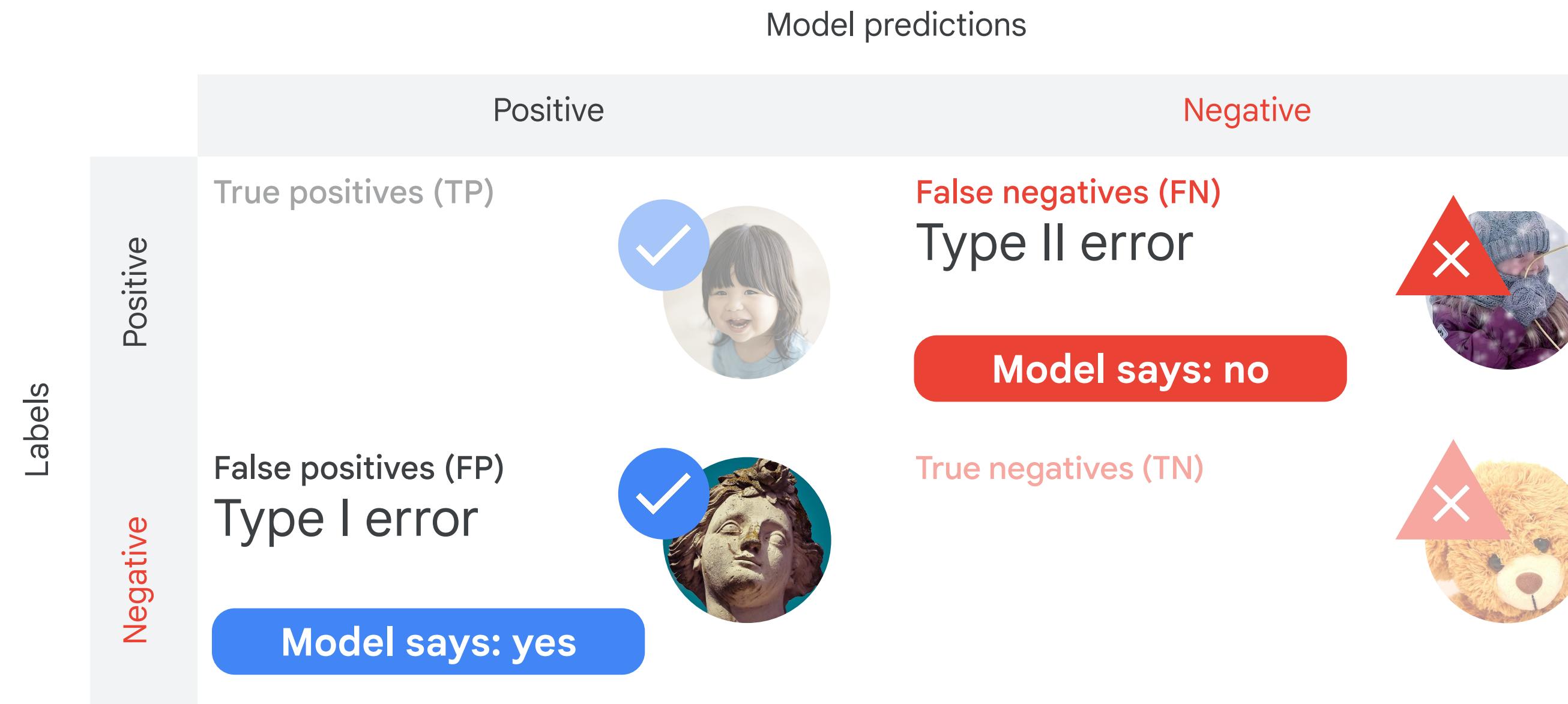
A confusion matrix leads to evaluation metric insights (for a classification model)

		Model predictions	
		Positive	Negative
Labels	Positive	True positives (TP) Label says something exists. Model predicts it.	
	Negative		

The confusion matrix leads to evaluation metric insights

		Model predictions	
		Positive	Negative
Labels	Positive	True positives (TP) Label = something exists. Model predicts it.	
	Negative	False positives (FP) Type I error Something doesn't exist. Model predicts it.	
		False negatives (FN) Type II error Label = something exists. Model doesn't predict it.	
		True negatives (TN) Something doesn't exist. Model doesn't predict it.	

False positives and false negatives errors occur when predictions and labels disagree



Evaluation metrics can help highlight areas where machine learning could be more inclusive

		Model predictions	
		Positive	Negative
Labels	Positive	True positives (TP) Label says something exists. The model predicts it.	False negatives (FN) Type II error Label says something exists. Model doesn't predict it.
	Negative	False positives (FP) Type I error Label says something doesn't exist. Model predicts it.	True negatives (TN) Label says something doesn't exist. Model doesn't predict it.

False negative rate is the fraction of true faces that are not detected by the ML system

		Model predictions	
		Positive	Negative
Labels	Positive	True positives (TP) Label says something exists. The model predicts it.	Type II error Label says something exists. Model doesn't predict it.
	Negative	False negative rate	$\frac{\text{False negatives}}{\text{False negatives} + \text{True positives}}$

False positive rate is the fraction of the faces that the ML model detects that are not really faces

		Model predictions	
		Positive	Negative
Labels	Positive	True positives (TP) Label says something exists. The model predicts it.	False positive rate = $\frac{\text{False positives}}{\text{False positives} + \text{True negatives}}$
	Negative	Type I error Label says something doesn't exist. Model predicts it.	

Classification evaluation metrics: Precision

Precision attempts to answer the following question:

What proportion of positive identifications was actually correct?

Precision is defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Confidence threshold  0.5

All labels

PR AUC 0.696

ROC AUC 0.81

Log loss 0.815

F1 score 0.5686126

Precision 60.4%

Recall 53.7%

Created Sep 29, 2021 12:52:20 AM

Classification evaluation metrics: Recall

Recall attempts to answer the following question:

What proportion of actual positives was identified correctly?

Mathematically, recall is defined as follows:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Note: A model that produces no false negative has a recall of 1.0

Confidence threshold  0.5

All labels

PR AUC 0.696

ROC AUC 0.81

Log loss 0.815

F1 score 0.5686126

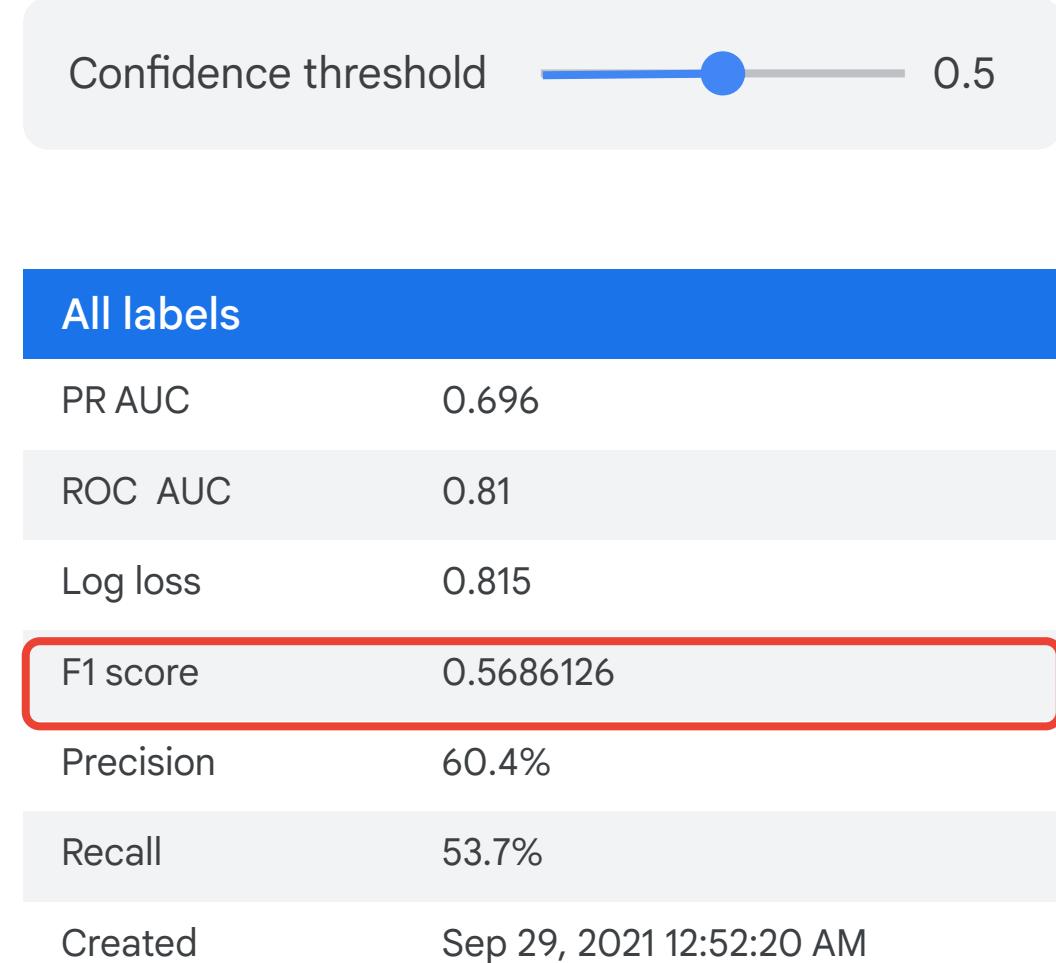
Precision 60.4%

Recall 53.7%

Created Sep 29, 2021 12:52:20 AM

F1 score

- Precision answers the question
 - Out of the equipment classified “will fail,” what fraction was correct?
- Recall answers the question
 - Out of the equipment that actually failed, what fraction did the classifier pick up?
- F1 score is the harmonic mean of precision and recall.



Sometimes, false positives are better than false negatives

Privacy in images



False positive



False negative

Sometimes, false negatives are better than false positives

False negative:

E-mail that is SPAM is not caught, so you see it in your inbox.



Jan Smith

Win the lottery with these numbers!

False positive:

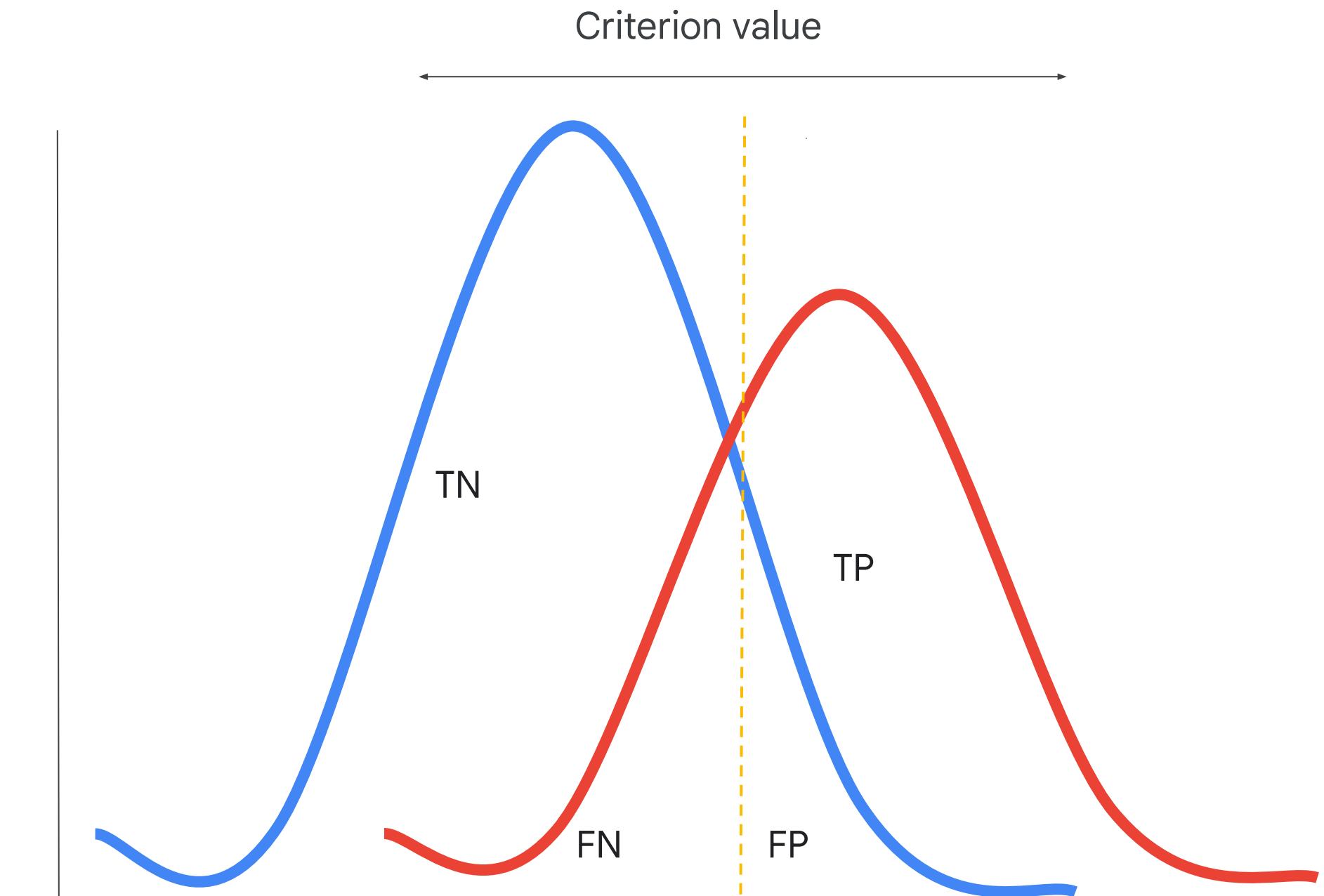
E-mail flagged as SPAM is removed from your inbox.



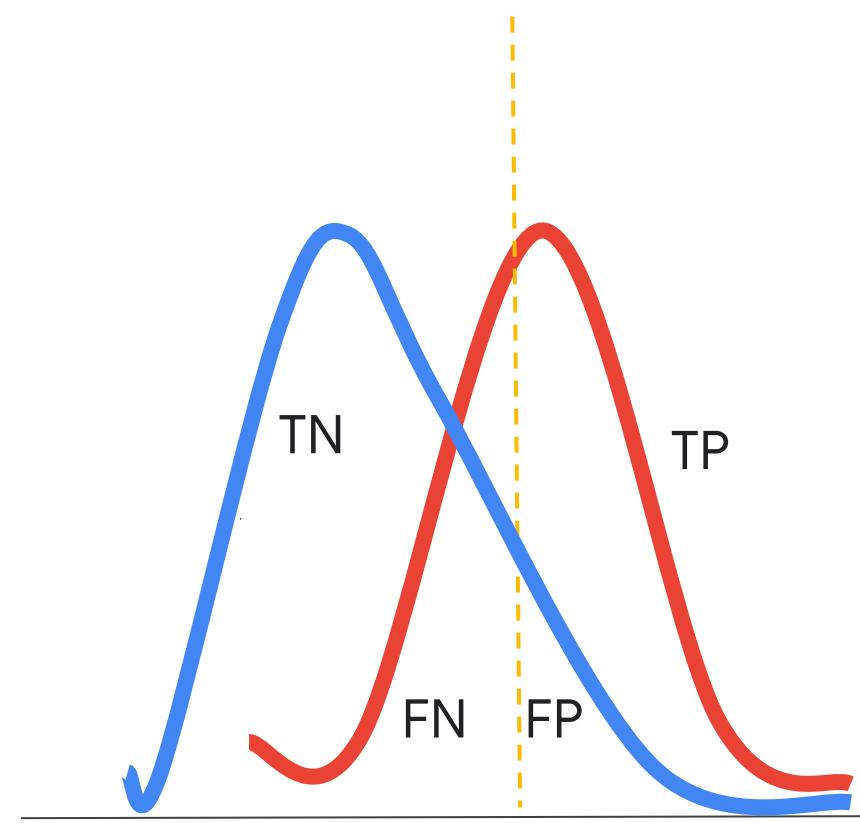
Karla Brown

Lunch today?

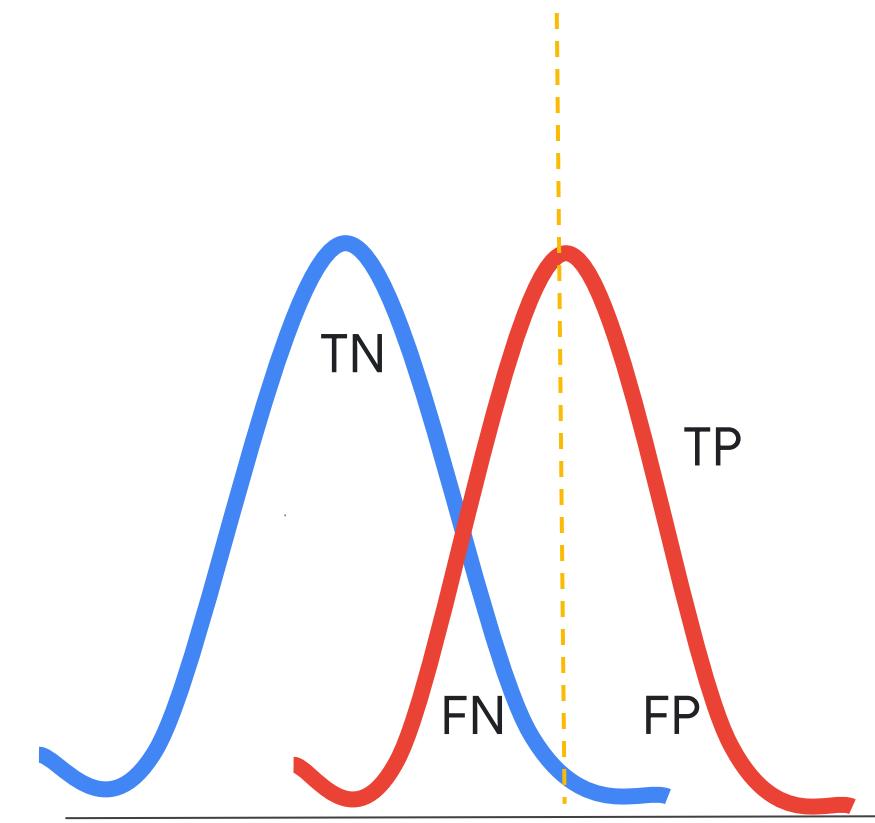
Find the threshold
that brings the
precision or recall to
acceptable values



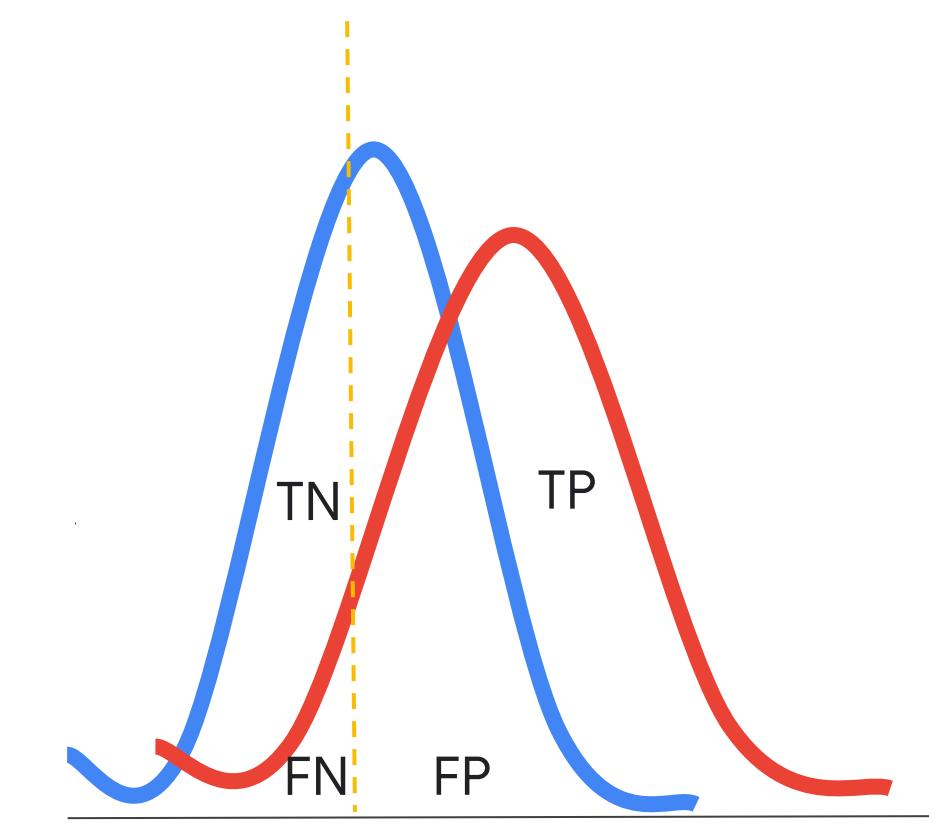
Check the precision/recall you obtain with that threshold in each of your subgroups



Sub-group 1

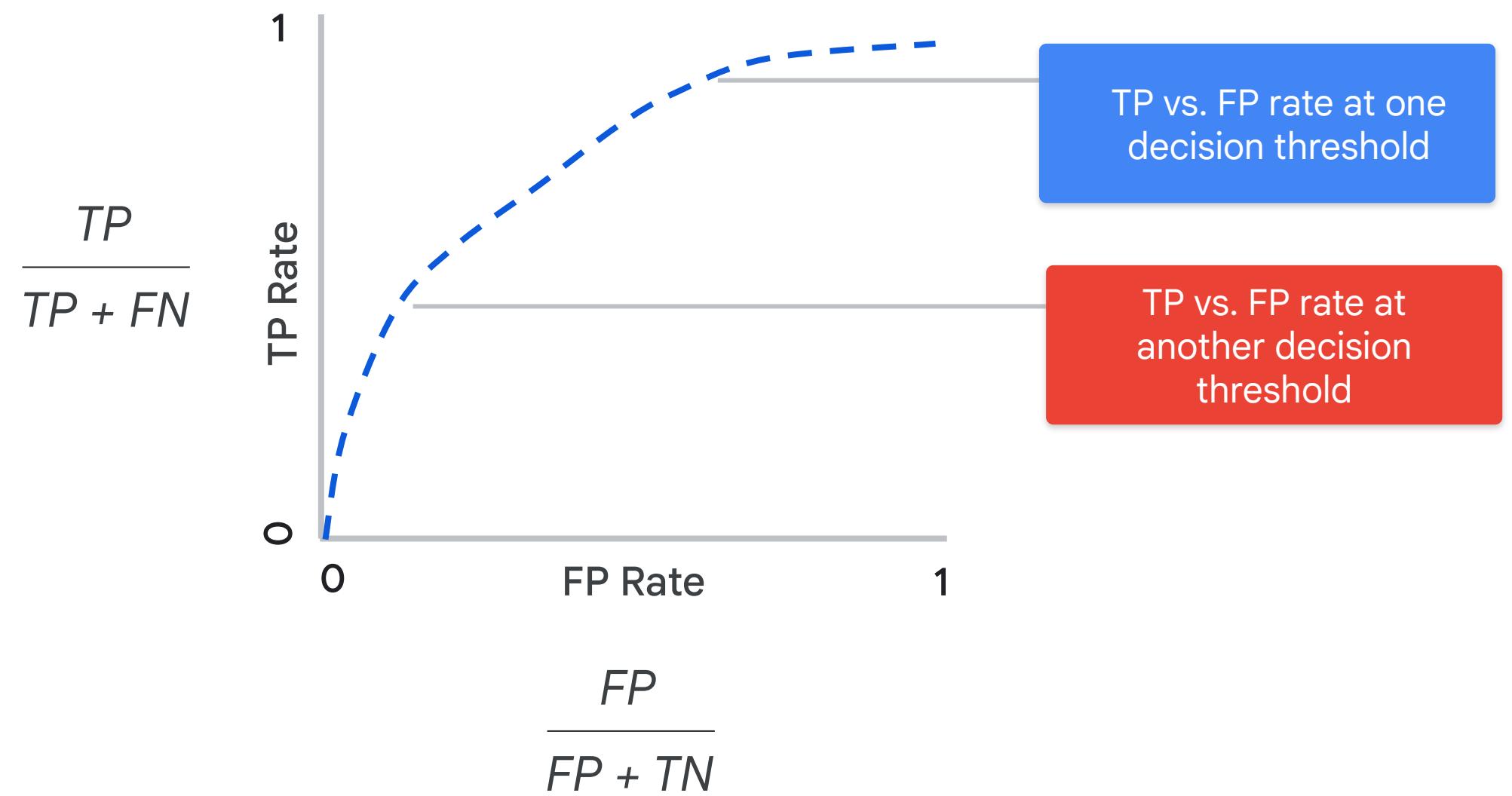


Sub-group 2

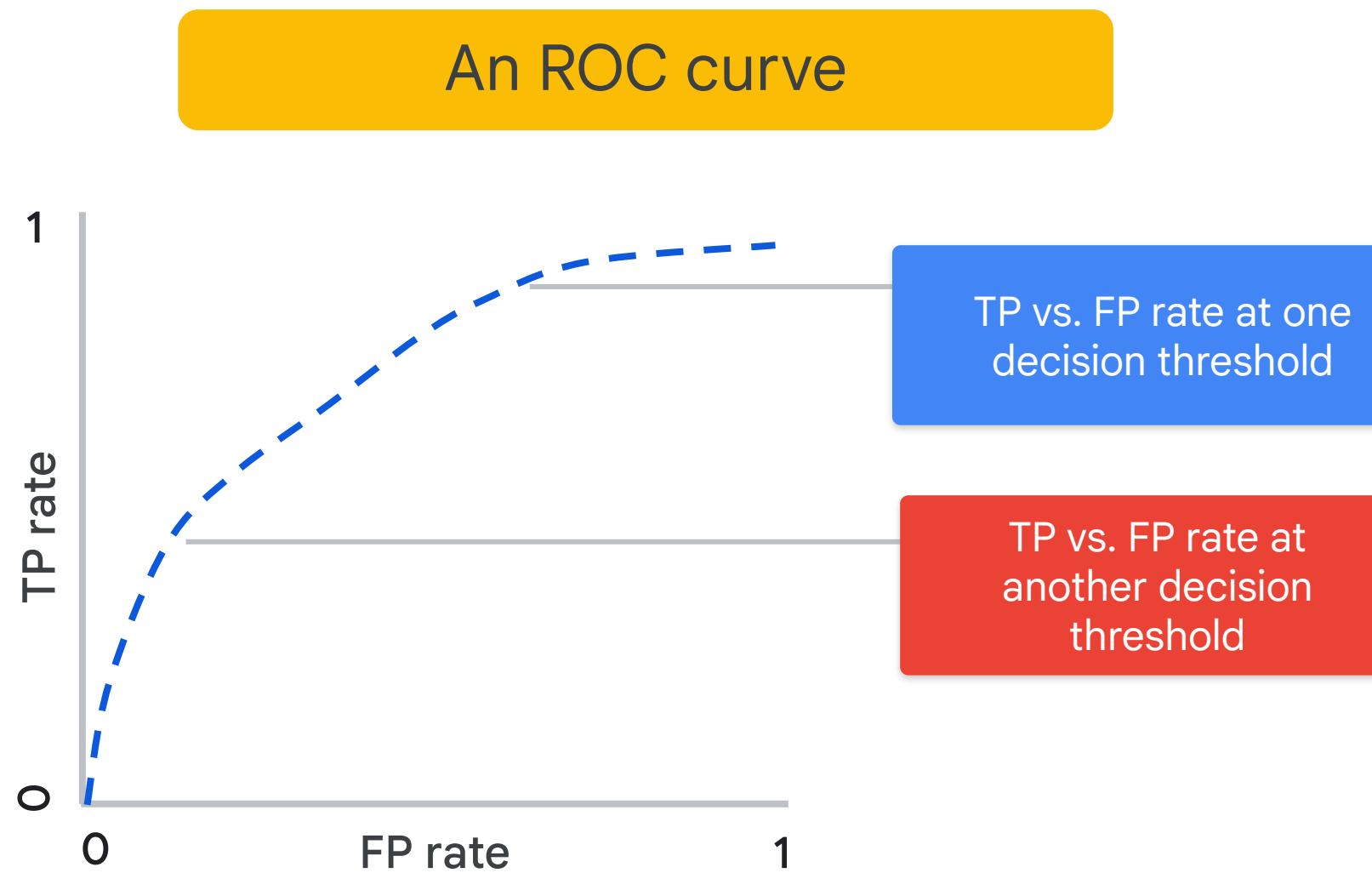


Sub-group 3

Use the ROC curve
to choose the
decision threshold
based on decision
criteria



ROC curve



Confidence threshold 0.5

All labels

PR AUC 0.696

ROC AUC 0.81

Log loss 0.815

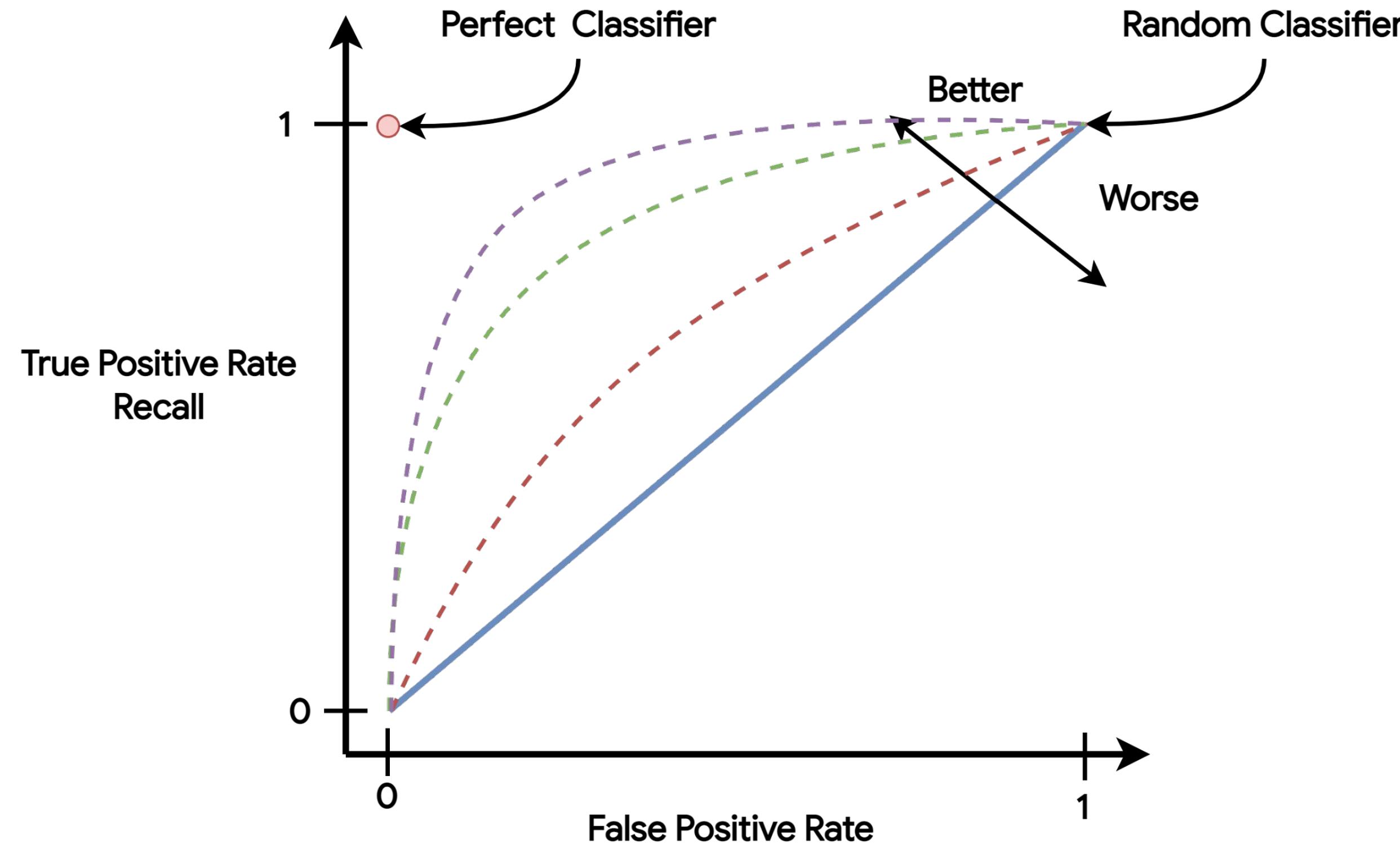
F1 score 0.5686126

Precision 60.4%

Recall 53.7%

Created Sep 29, 2021 12:52:20 AM

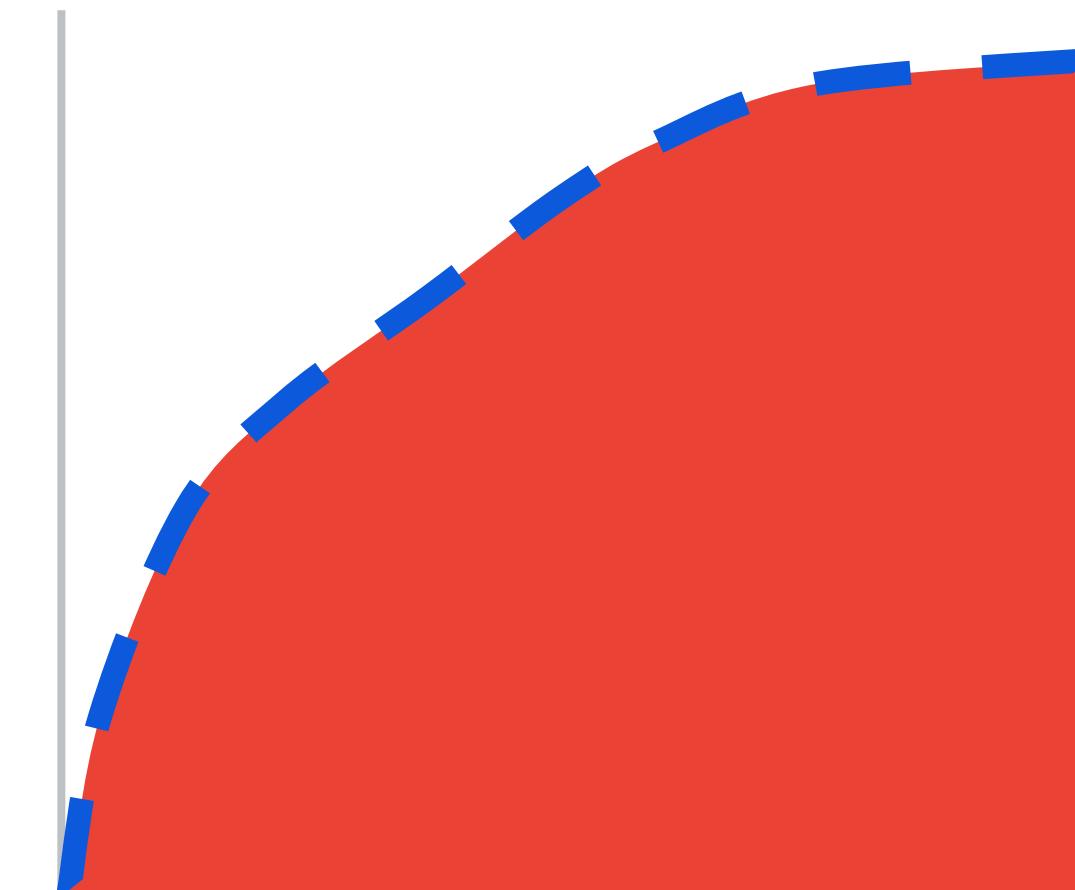
AUC between Classifiers



The Area-Under-Curve (AUC) provides an aggregate measure of performance across all possible classification thresholds

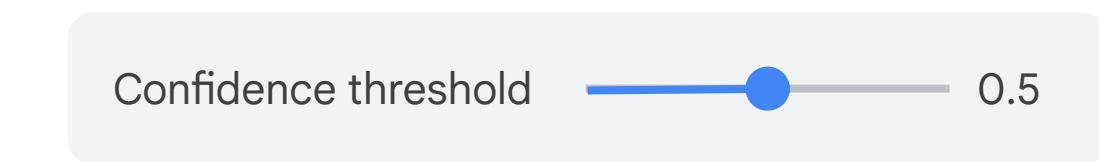
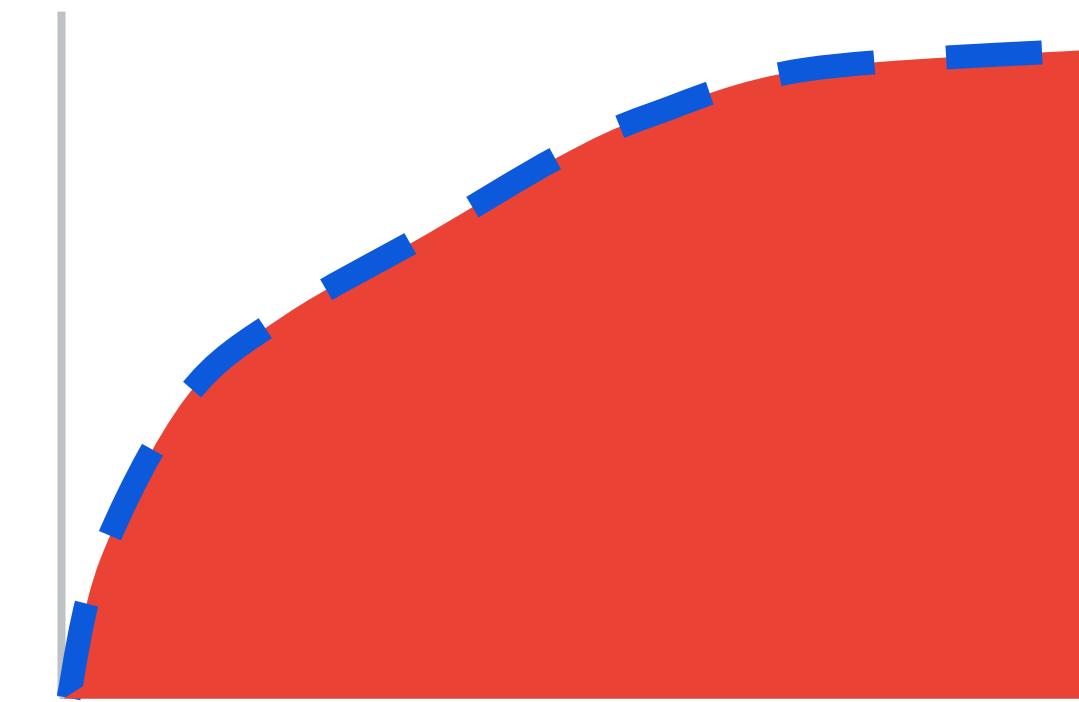
AUC helps you choose between models when you don't know what decision threshold is going to be ultimately used.

"If we pick a random positive and a random negative, what's the probability my model scores them in the correct relative order?"



AUC

- AUC = area under the ROC curve.
- Interpretation: If we pick a random positive and a random negative, what's the probability that my model scores them in the correct relative order?
- Intuition: Gives an aggregate measure of performance aggregated across all possible classification thresholds.



All labels	
PR AUC	0.696
ROC AUC	0.81
Log loss	0.815
F1 score	0.5686126
Precision	60.4%
Recall	53.7%
Created	Sep 29, 2021 12:52:20 AM

Recommended Additional Activities

[Recommendation Systems with TensorFlow on Google Cloud](#)

Sometimes, false positives are better than false negatives

Privacy in images



False positive



False negative

Sometimes, false negatives are better than false positives

False negative:

E-mail that is SPAM is not caught, so you see it in your inbox.



Jan Smith

Win the lottery with these numbers!

False positive:

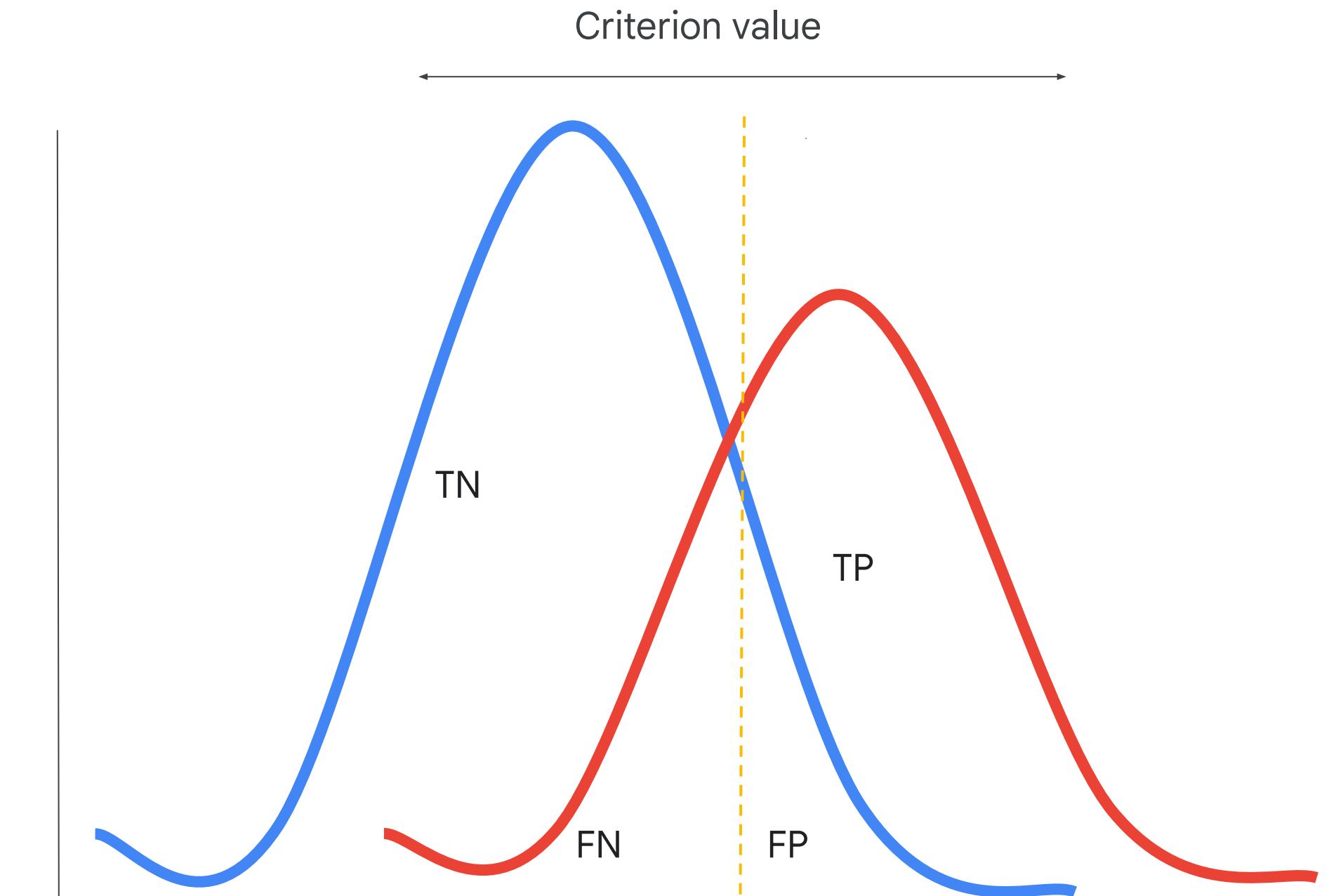
E-mail flagged as SPAM is removed from your inbox.



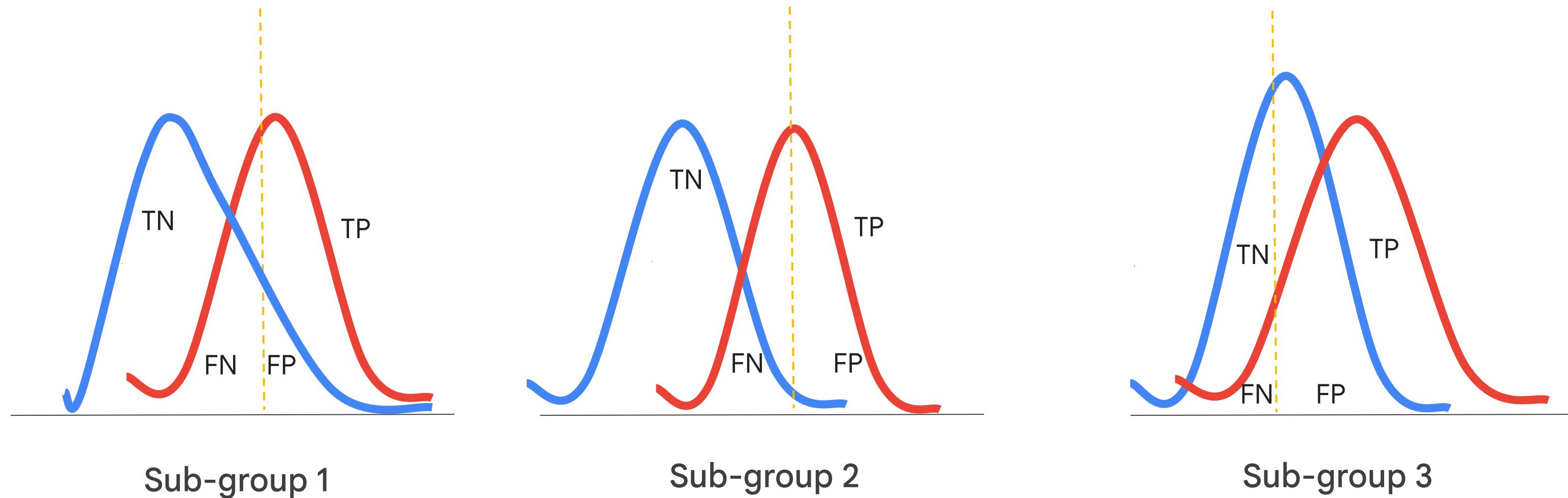
Karla Brown

Lunch today?

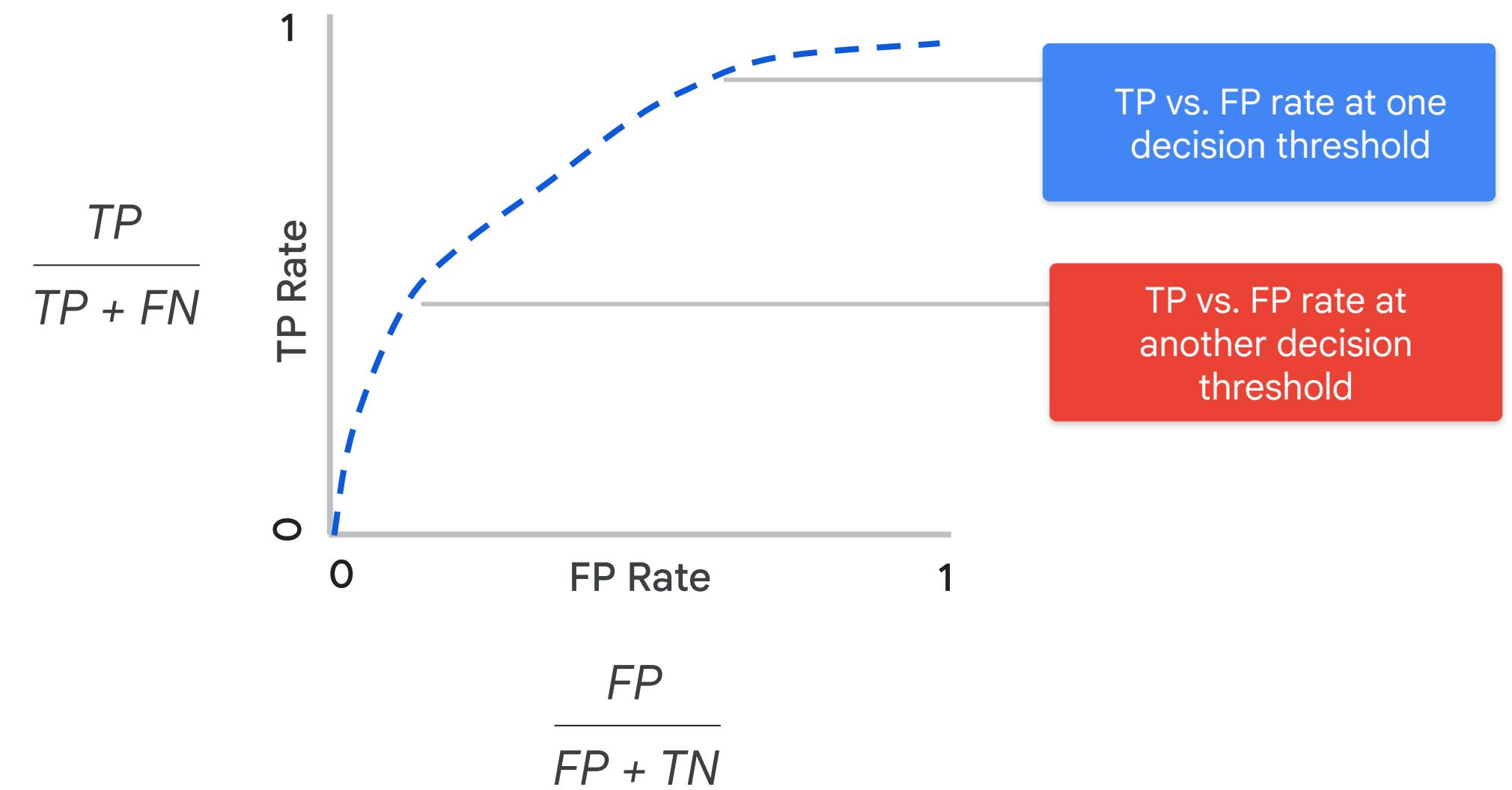
Find the threshold
that brings the
precision or recall to
acceptable values



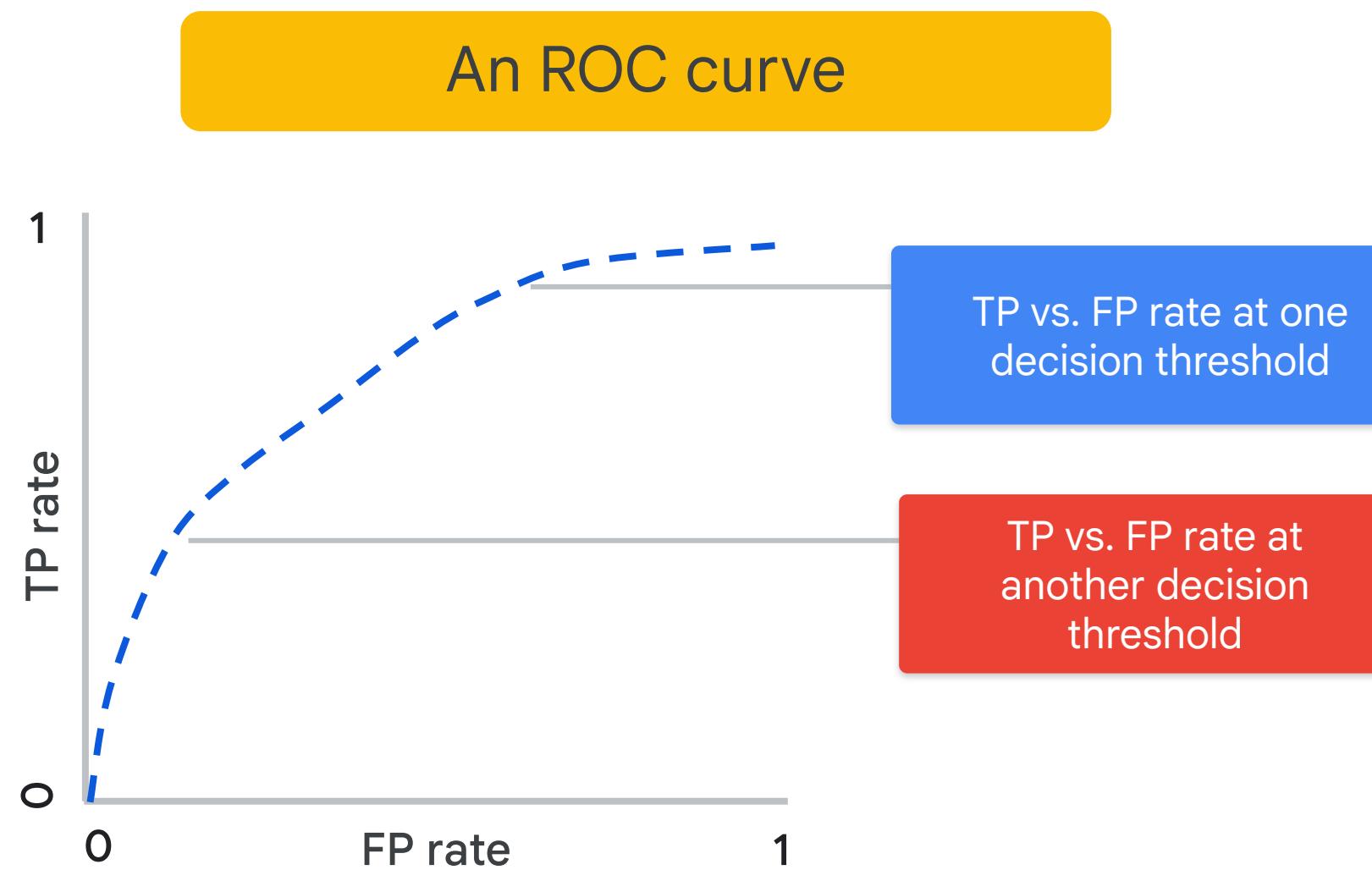
Check the precision/recall you obtain with that threshold in each of your subgroups



Use the ROC curve
to choose the
decision threshold
based on decision
criteria



ROC curve



Confidence threshold 0.5

All labels

PR AUC 0.696

ROC AUC 0.81

Log loss 0.815

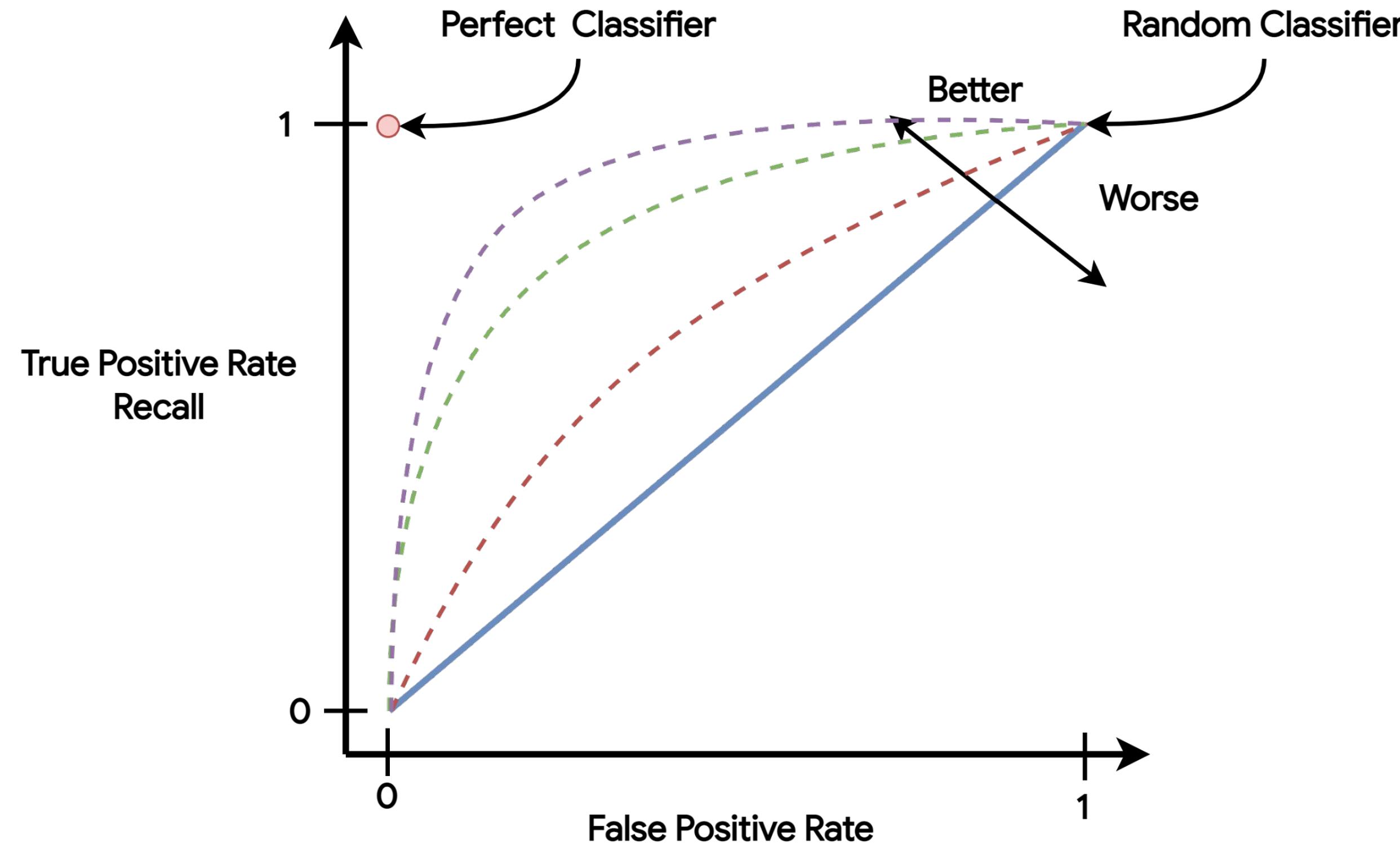
F1 score 0.5686126

Precision 60.4%

Recall 53.7%

Created Sep 29, 2021 12:52:20 AM

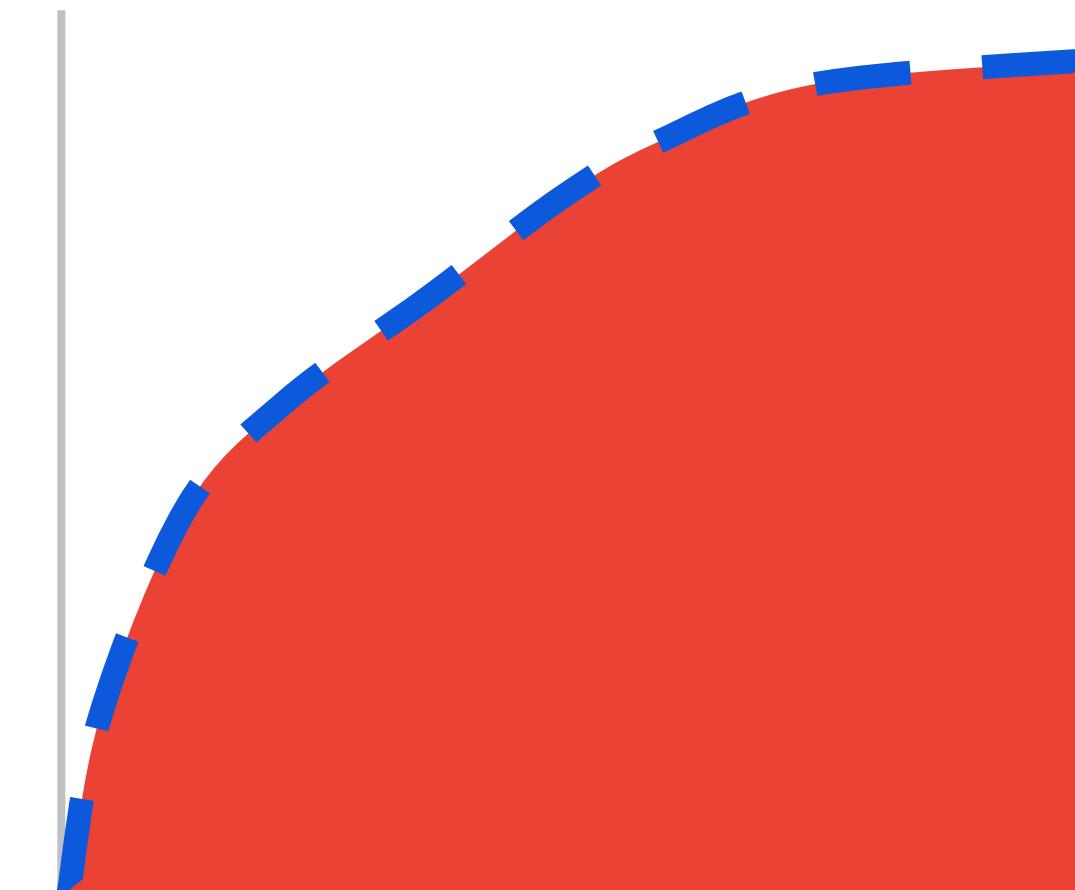
AUC between Classifiers



The Area-Under-Curve (AUC) provides an aggregate measure of performance across all possible classification thresholds

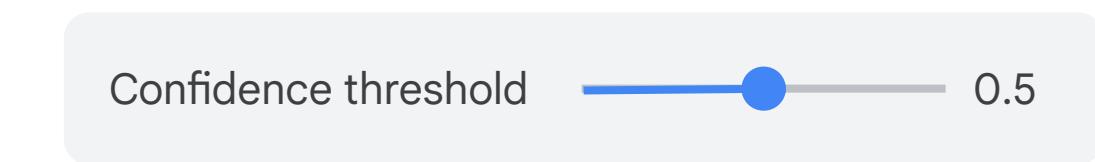
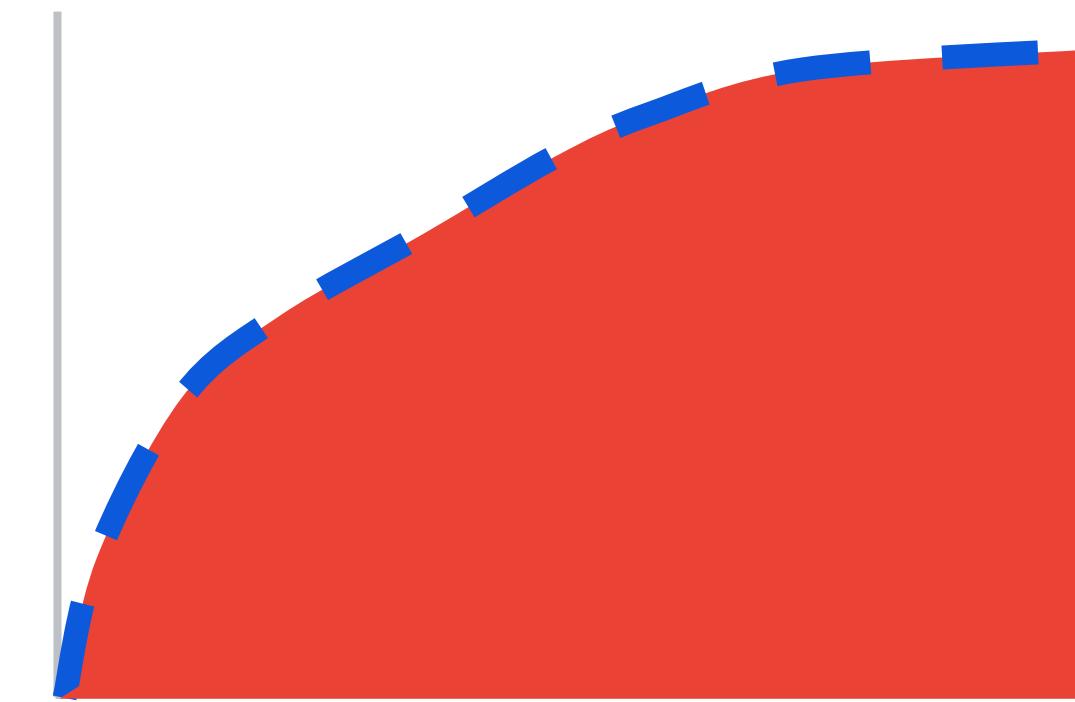
AUC helps you choose between models when you don't know what decision threshold is going to be ultimately used.

"If we pick a random positive and a random negative, what's the probability my model scores them in the correct relative order?"



AUC

- AUC = area under the ROC curve.
- Interpretation: If we pick a random positive and a random negative, what's the probability that my model scores them in the correct relative order?
- Intuition: Gives an aggregate measure of performance aggregated across all possible classification thresholds.



All labels	
PR AUC	0.696
ROC AUC	0.81
Log loss	0.815
F1 score	0.5686126
Precision	60.4%
Recall	53.7%
Created	Sep 29, 2021 12:52:20 AM

Recommended Additional Activities

[Recommendation Systems with TensorFlow on Google Cloud](#)

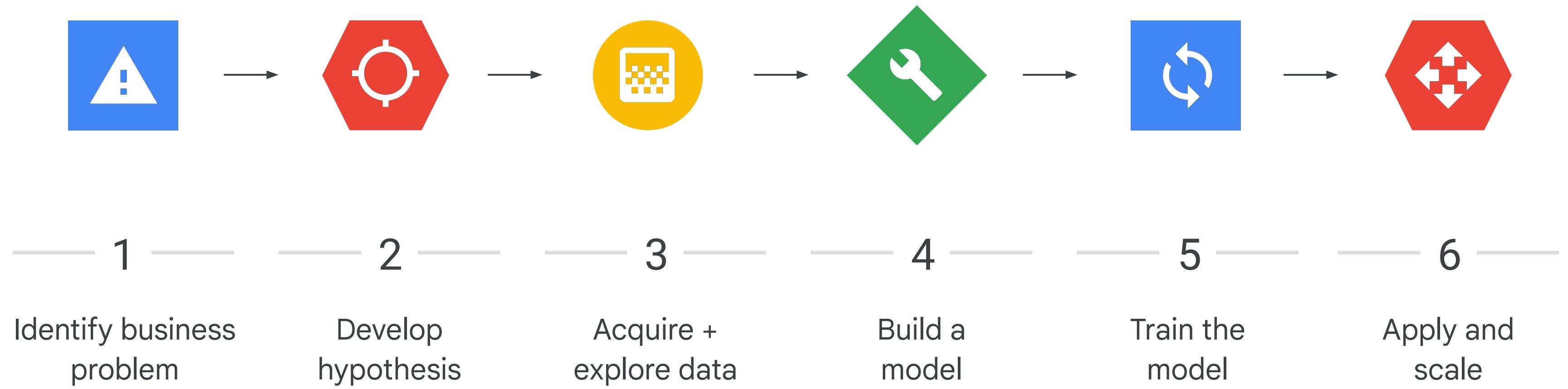


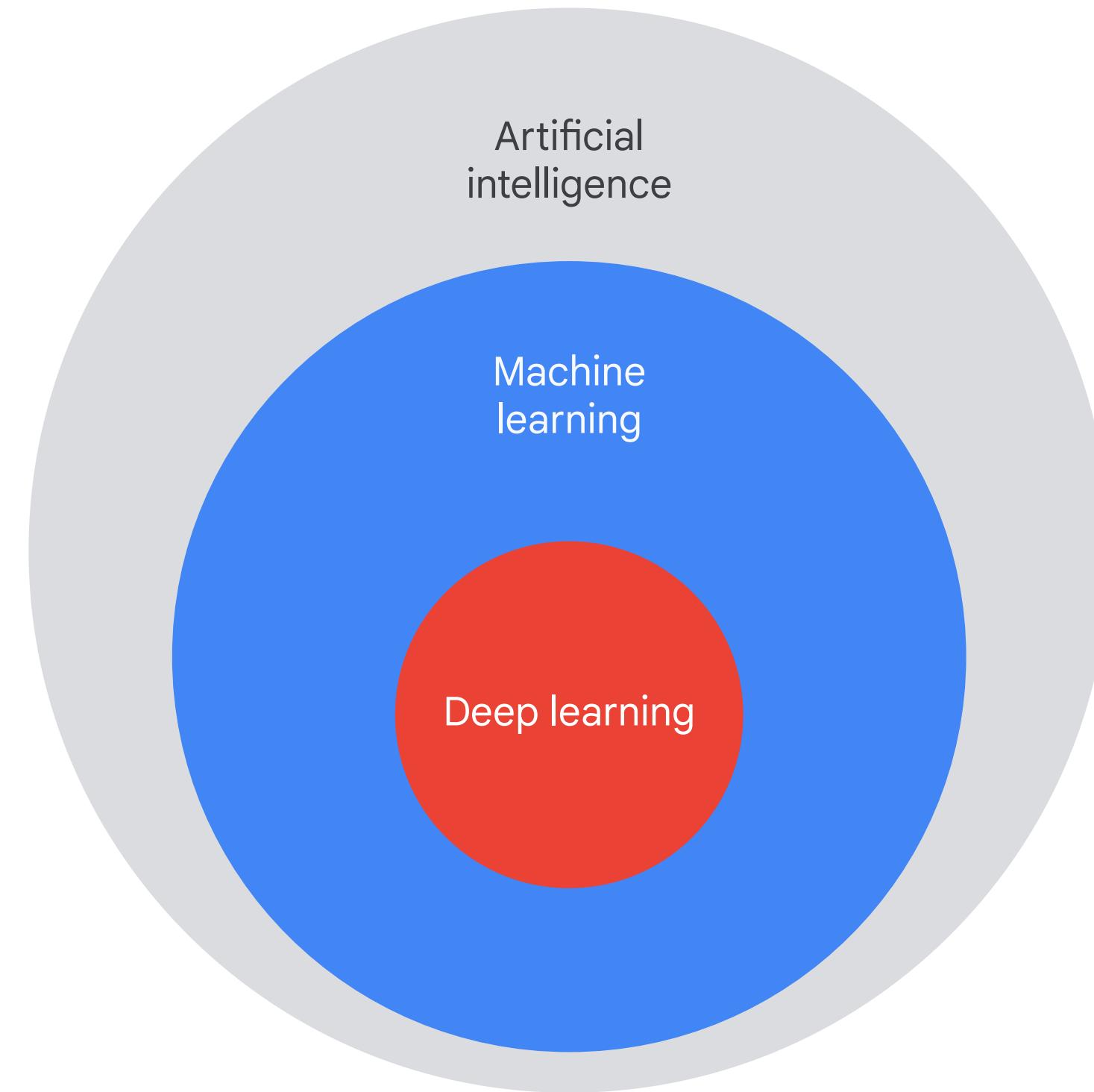
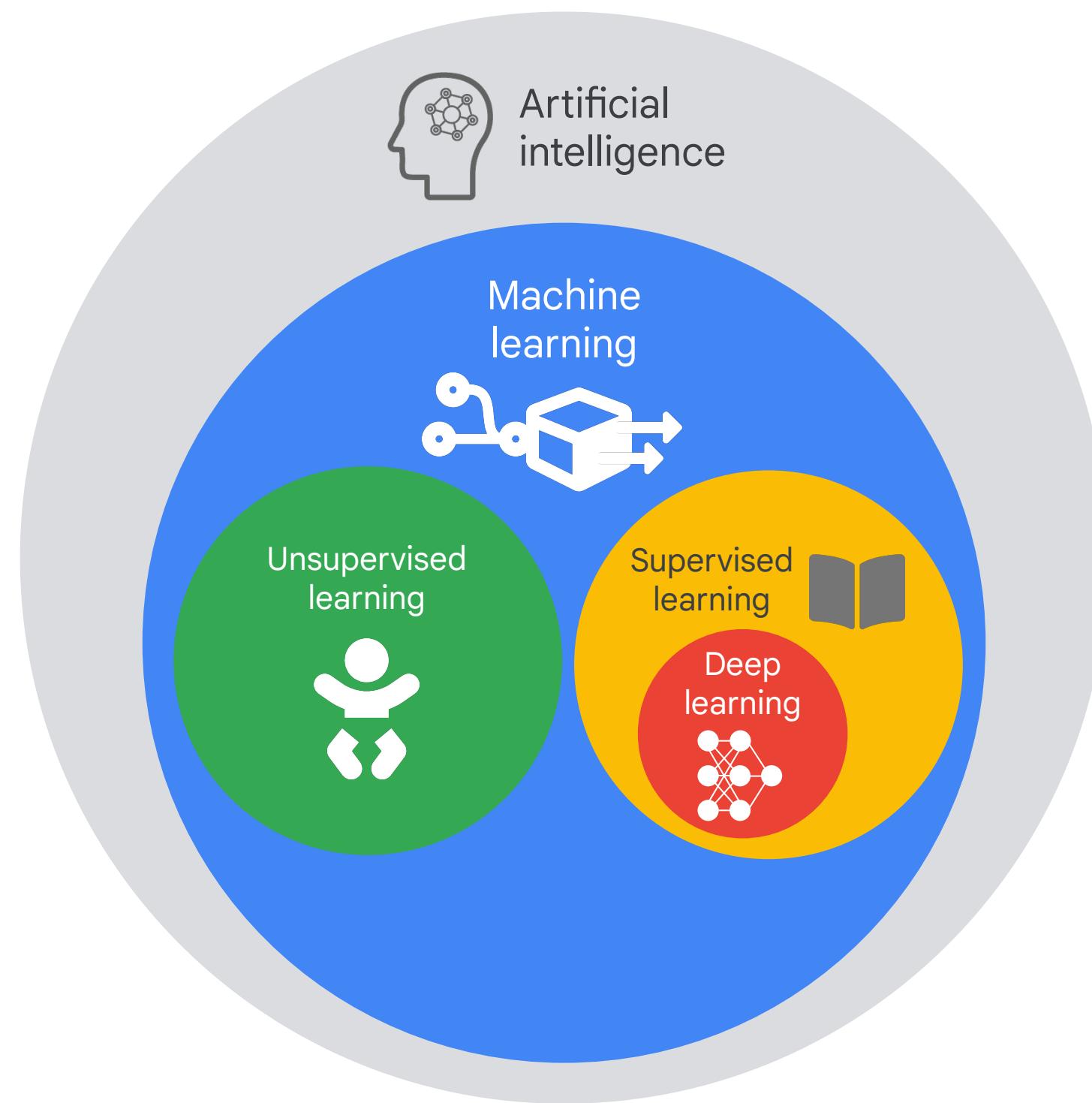
Machine Learning vs Deep Learning

All machine learning starts with a
business requirement or problem
you are **trying to solve.**



To build a machine learning model







Factors

- Data requirement
- Accuracy
- Training time
- Hardware dependency
- Hyperparameter tuning



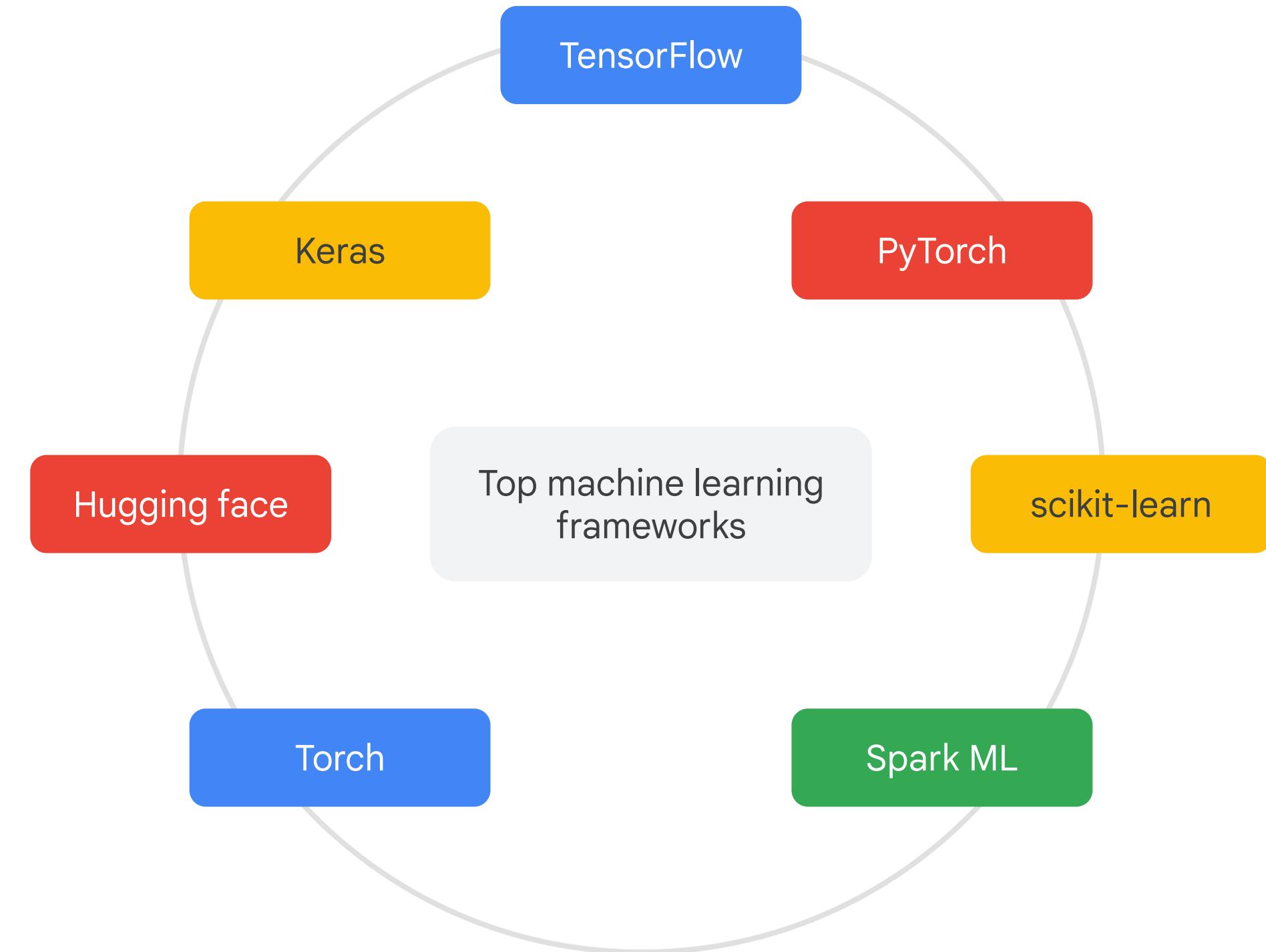
Deep learning

- Requires large data
- Provides high accuracy
- Takes longer to train
- Requires GPU to train properly
- Can be tuned in various different ways



Machine learning

- Can train on lesser data
- Gives lesser accuracy
- Takes less time to train
- Trains on CPU
- Limited tuning capabilities





**Thank you
End of Session 3**

Google Cloud



Google Cloud