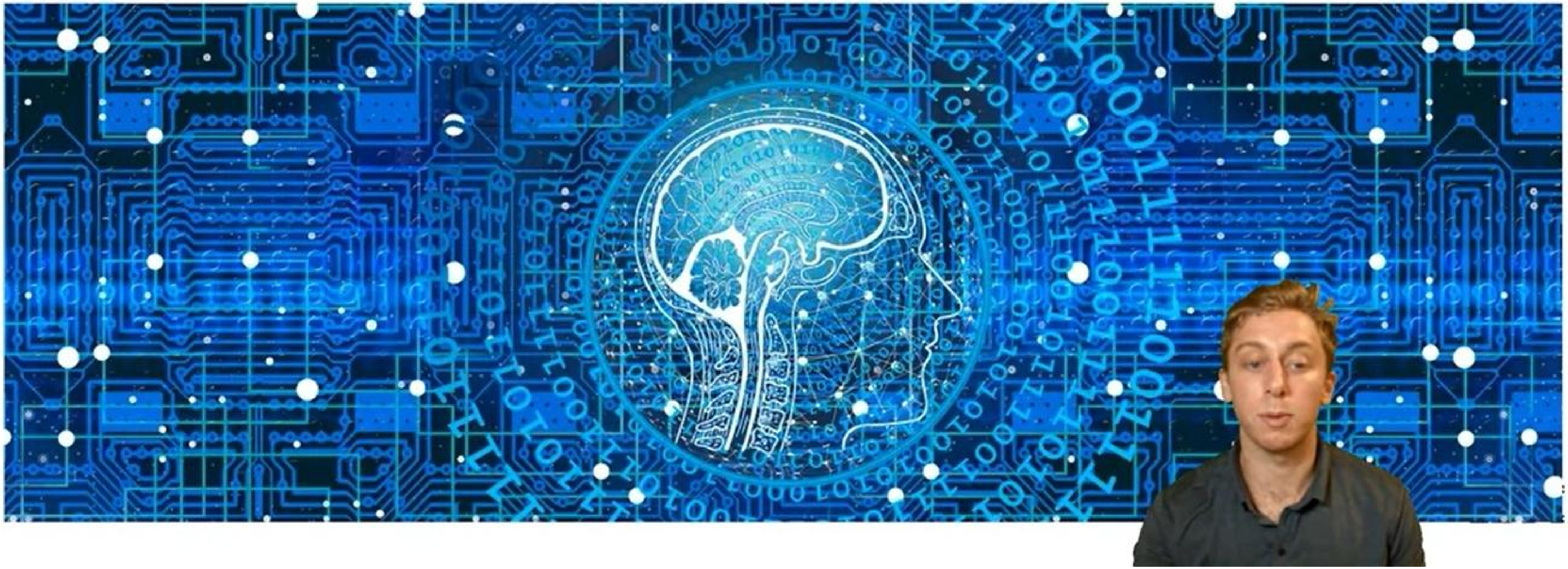
Introduction to Supervised Machine Learning

Course 2:
Supervised Learning: Regression

What Is Machine Learning?

Machine learning: is the process through which computers are said to "learn" and infer predictions from data.



Machine Learning in Context: AI

Thinking Humanly	**Thinking Rationally**
"The exciting new effort to make computers think . . . *machines with minds*, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Acting Humanly	**Acting Rationally**
"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	"Computational Intelligence is the study of the design of intelligent agents." (Poole *et al.*, 1998) "AI . . . is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

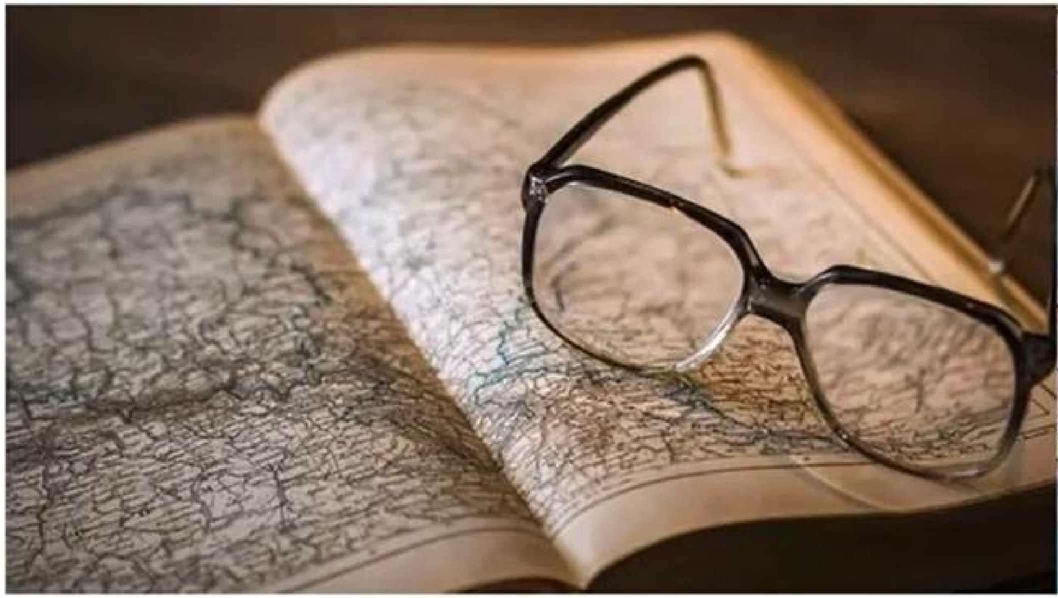
Some definitions of artificial intelligence, organized into four categories.

Source: Russel & Norvig: Artificial Intelligence, A Modern Approach

Model: A Learning Algorithm

A **model** is a small thing that captures a larger thing.

A *good* model omits unimportant details while retaining what's important.



Machine Learning in Our Daily Lives

Spam Filtering	Web Search	Postal Mail Routing
Fraud Detection	Movie Recommendations	Vehicle Driver Assistance
Web Advertisements	Social Networks	Speech Recognition

IBM

Types of Machine Learning

Supervised

data points have known outcome

Unsupervised

data points have unknown outcome

Semi-Supervised

uses both: data with outcomes, data without outcomes

How Does Machine Learning Work?

ML Framework (applies to Supervised Machine Learning Models):

$$y_p = f(\Omega, x)$$

$\boldsymbol{x}$: **Input**

$\boldsymbol{y_p}$: **Output** (values predicted by the model)

- **Observations**: rows or examples the model will see.
- **Features**: the different ways that we measure each observation.

Here, we distinguish between $\boldsymbol{y_p}$ (the prediction of our model)
and $\boldsymbol{y}$ (the observed values of the target variable).

Machine Learning Framework

ML Framework (applies to Supervised Machine Learning Models):

$$y_p = f(\Omega, x)$$

$\boldsymbol{x}$ and $\boldsymbol{y}$ are **arrays** with 1 or more rows and columns.

- A single observation can be represented by a **row**.
- A single feature can be represented by a **column**.
- $\boldsymbol{\Omega}$ represents **parameters** of the model (1 or more variables).

It is important to specify exactly what y_p, f, Ω, and x are for any model we work with.

Machine Learning Framework

ML Framework (applies to Supervised Machine Learning Models):

$$y_p = f(\Omega, x)$$

Ω represents **parameters** of the model (1 or more variables).

- This is what changes as the model learns.
- Some models have many different parameters, some have very few.

As we implement our modeling approach, we will also select **hyperparameters:**

- A **hyperparameter**: is a parameter that is not learned directly from the data, but relates to implementation: training our ML model.
- We will apply techniques for using model performance to inform hyperparameter selection.

IBM

Fit Parameters vs. Hyperparameters

ML Framework (applies to Supervised Machine Learning Models):

$$y_p = f(\Omega, x)$$

Our framework estimates a relationship between the features and target:

- Here, **Ω** (the **Fit Parameters**) involve aspects of the model we estimate (fit) using the data.
- To implement our approach, we make decisions regarding how to produce these estimates.
- These decisions lead to **hyperparameters**, that are an important part of the machine learning workflow (though not explicit components of the model).

Machine Learning Framework

ML Framework (applies to every Supervised Machine Learning Models):

$$y_p = f(\Omega, x)$$

Two main modeling approaches:

– **Regression**: y is numeric.

- E.g.: stock price, box office revenue, location (x,y coordinates).

– **Classification**: y is categorical.

- E.g.: face recognition customer churn, which word comes next.

How Does Machine Learning Work?

ML Framework (applies to Supervised Machine Learning Models):

$$y_p = f(\Omega, x)$$

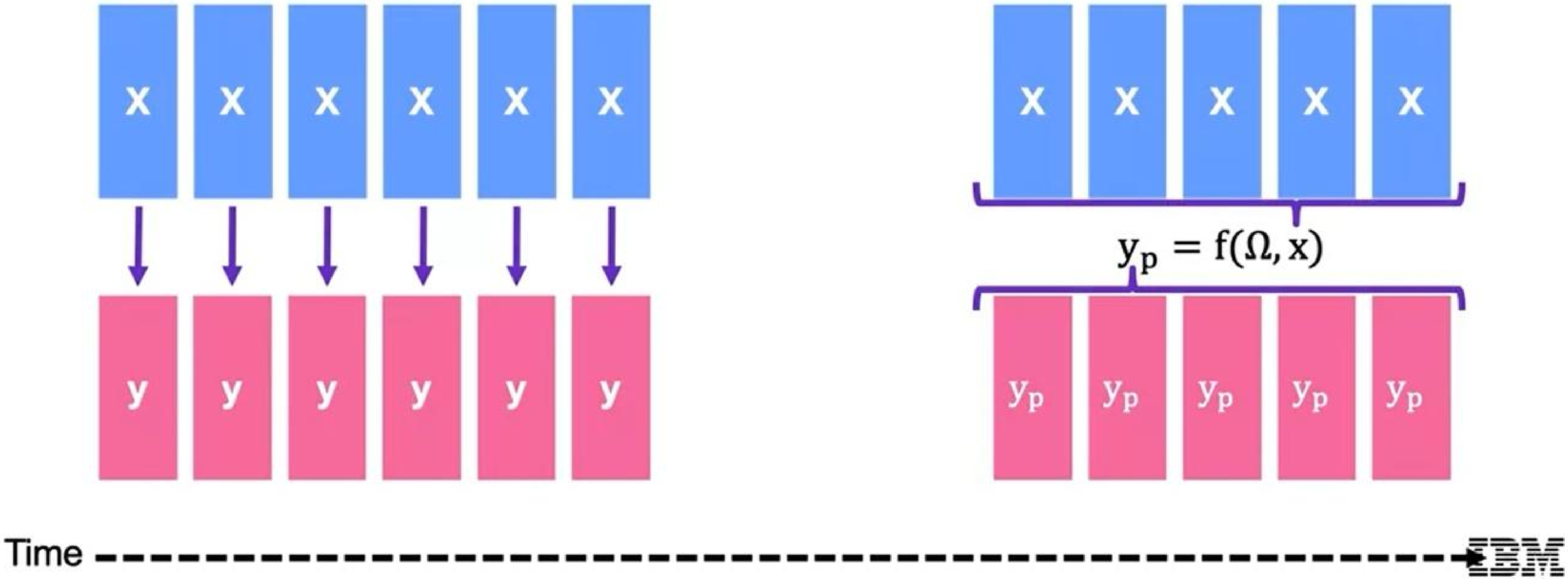
$\boldsymbol{x}$: Input.

$\boldsymbol{y_p}$: Output (values predicted by the model).

$\boldsymbol{f}(\cdot)$: Prediction function that generates predictions from $\boldsymbol{x}$ and $\boldsymbol{\Omega}$.

Machine Learning Framework

Data Scientists train the model to find the best Ω given past experience.



How Does Machine Learning Work?

ML Framework (applies to Supervised Machine Learning Models):

$$y_p = f(\Omega, x)$$

$\boldsymbol{x}$: Input.

$\boldsymbol{y_p}$: Output (values predicted by the model).

$\boldsymbol{f}(\cdot)$: Prediction function that generates predictions from $\boldsymbol{x}$ and $\boldsymbol{\Omega}$.

$\boldsymbol{J}(\boldsymbol{y}, \boldsymbol{y_p})$: Loss

– Most ML models define a quantitative score for how “good” our predictions are.

– Typically measures how close our predictions are to the true values.

How Does Machine Learning Work?

ML Framework (applies to Supervised Machine Learning Models):

$$y_p = f(\Omega, x)$$

$\boldsymbol{x}$: Input (features).

$\boldsymbol{y_p}$: Output (values predicted by the model).

$\boldsymbol{f}(\cdot)$: Prediction function that generates predictions from $\boldsymbol{x}$ and $\boldsymbol{\Omega}$.

$\boldsymbol{J}(\boldsymbol{y}, \boldsymbol{y_p})$: Loss

Update rule: using features x and outcome y, choose parameters $\boldsymbol{\Omega}$ to minimize loss $\boldsymbol{J}$

Supervised Machine Learning: Approaches

Course 2:
Supervised Learning: Regression

Learning Goals

In this section, we will cover:

- Machine Learning Objectives: interpretation and prediction
- Examples of Machine Learning applications
- The Machine Learning Framework

IBM

Interpretation and Prediction

Interpretation:

– In some cases, the primary objective is to train a model to find insights from the data.

– In $y_p = f(\Omega, x)$, the **interpretation approach** uses Ω to give us insight into a system.

– Common workflow:

- **Gather** x, y; **Train** model by finding the Ω that gives the best prediction $y_p = f(\Omega, x)$.
- **Focus on** Ω (rather than y_p) to generate insights.

– Example interpretation exercises:

- x = customer demographics, y = sales data; examine Ω to understand loyalty by segment
- x = car safety features, y = traffic accidents; examine Ω to understand what makes cars safer
- x = marketing budget, y = movie revenue; examine Ω to understand marketing effectiveness

IBM

Interpretation and Prediction

Prediction:

– In some cases, the primary objective is to make the best prediction.

– In $y_p = f(\Omega, x)$, the **prediction approach** compares y_p with y.

– The focus is on **performance metrics**, which measure the quality of the model's predictions.

- **Performance metrics** usually involve some measure of closeness between y_p with y.
- Without focusing on interpretability, we risk having a Black-box model.

– Example prediction exercises:

- x = customer purchase history , y = customer churn; focus on predicting customer churn
- x = financial information, y = flagged default/non-default; focus on predicting loan default
- x = purchase history, y = next purchase; focus on predicting the next purchase

IBM

Example: Regression with Housing Data

In this course, we will examine housing datasets.

Here, our target is the **price** of housing,
and our features include characteristics
about the house and area.

Suppose we fit our model $y_p = f(\Omega, x)$,
based on data on housing sales in Ames, Iowa,
and obtain estimates of parameters Ω.

These parameters represent coefficients relating
the features x with expected target values.

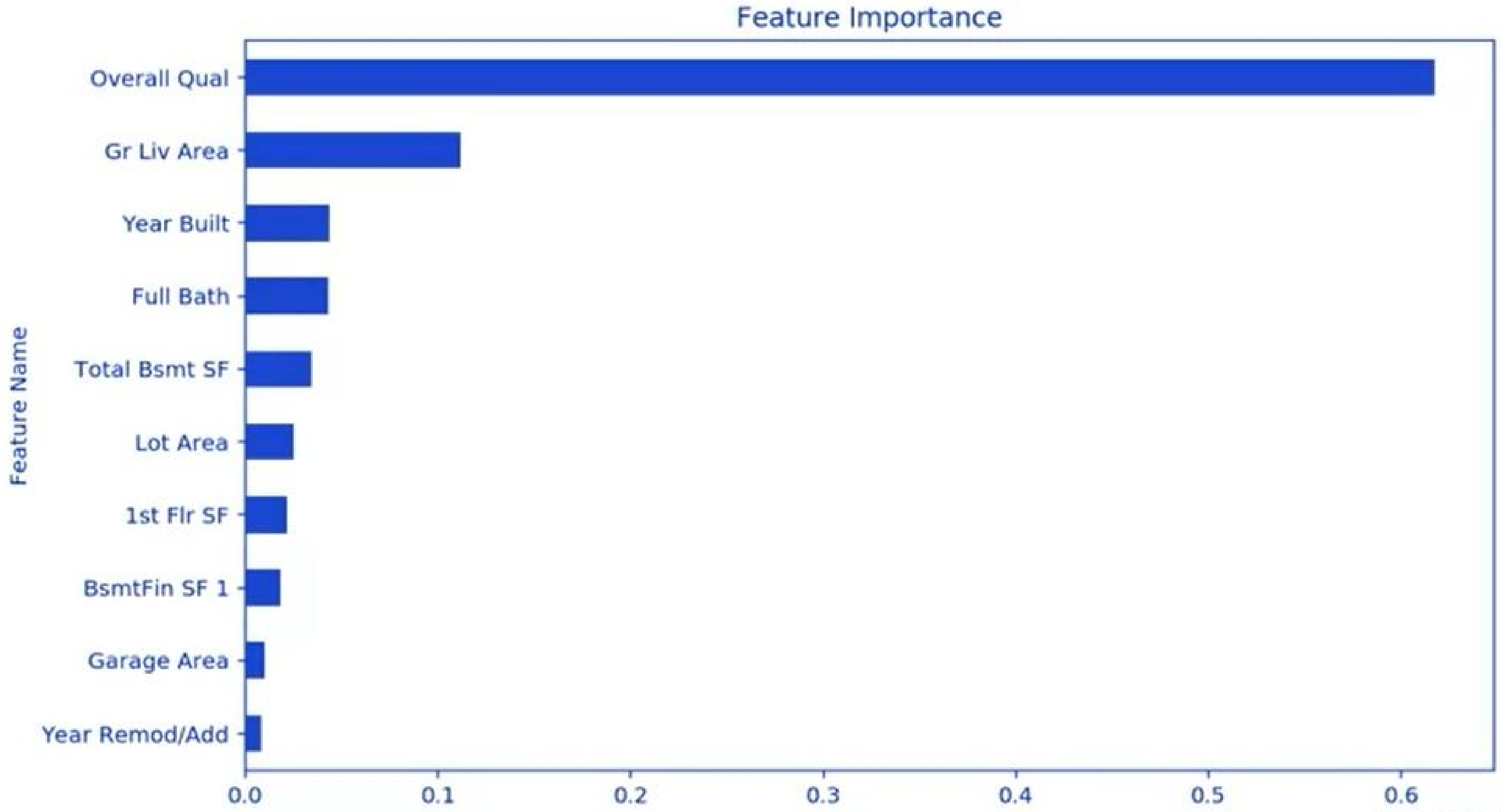
We can **interpret** our results to learn about
feature importance.

Housing Data: Regression Interpretation

Regression **Interpretation:**

Examining results from regression of housing sale prices in Ames, Iowa.

Which features are **most** important?



Example: Regression with Housing Data

Here, our target is the **price** of housing, and our features include characteristics about the house and area.

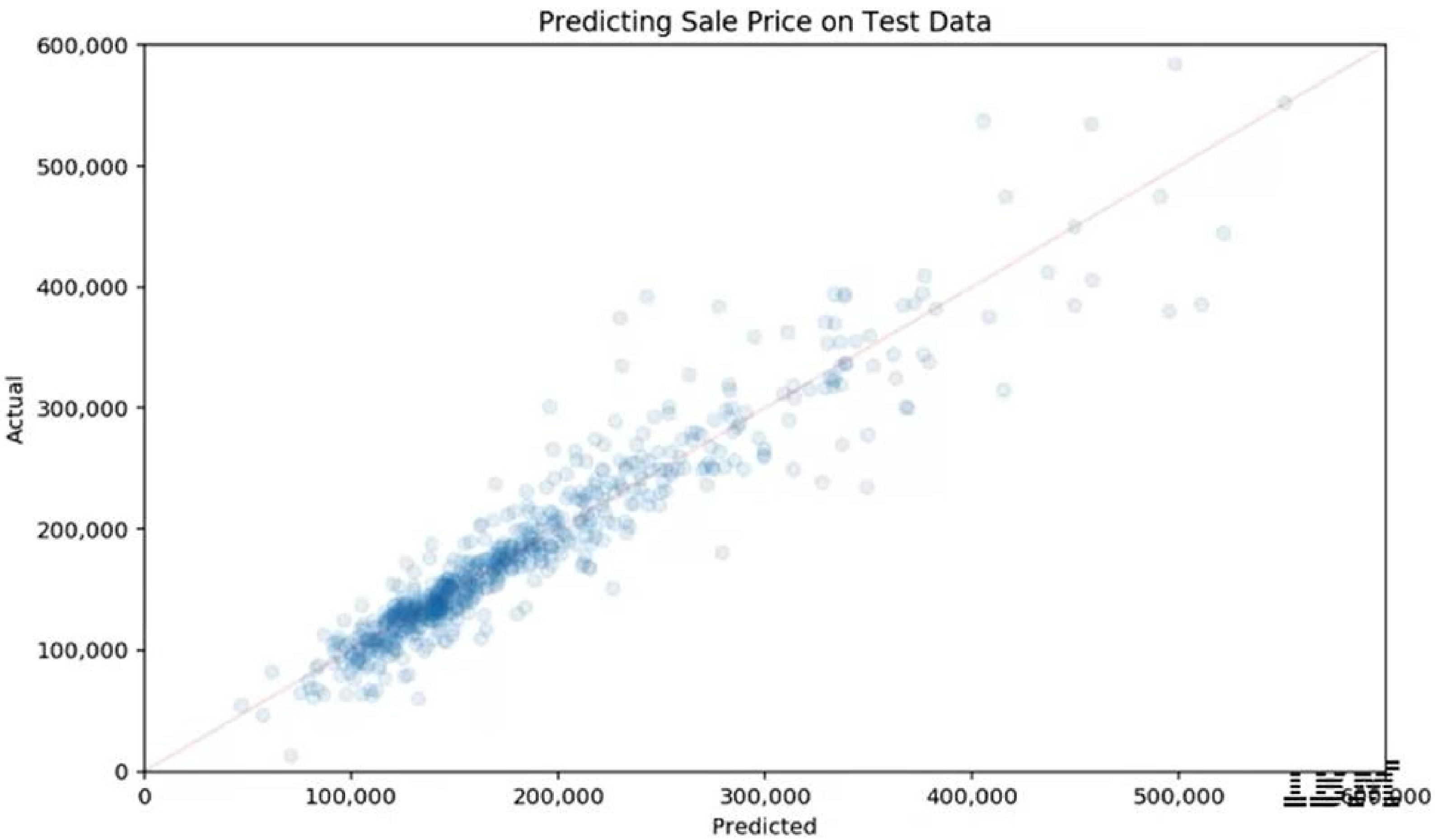
Suppose we fit our model $y_p = f(\Omega, x)$, based on data on housing sales in Ames, Iowa, and obtain estimates of parameters Ω.

Our primary aim may be **prediction**, in which case we are more focused on **generating** values for y_p than in interpreting parameters.

Housing Data: Regression Prediction

Prediction using Regression Example:

Use characteristics to predict unknown sale prices, focusing on how accurately we are able to predict.



Classification Example: Customer Churn

Customer **churn** occurs when a customer leaves a company.

Data related to churn may include a target variable for whether or not the customer left, as well as information on customer characteristics.

Here, we may be interested in both:
- **Interpretation:** understanding factors that may lead to customers leaving.
- **Prediction:** estimating how long customers are likely to stay can help us understand how many we still need to support, and how valuable they are to the company.

Two Common Approaches

Interpretation and prediction in Supervised Machine Learning

- Majority of projects will call for a balance.
- Interpretation can provide insight into improvements in prediction, and vice-versa.
- Not all models will allow both: Supervised Machine Learning models provide varying levels of support for interpretation vs. prediction

IBM

ML Framework: Takeaways

Machine Learning is the subset of AI that focuses on model building to support a goal of interpretation and/or prediction.

ML algorithms;

- Use past experience to build a model that is useful for future experience.
- Follow a general form: $y_p = f(\Omega, x)$.

Learning Recap

In this section, we discussed:

- Machine Learning Objectives: interpretation and prediction
- Examples of Machine Learning applications
- The Machine Learning Framework

IBM

Regression vs. Classification Problems

Course 2:
Supervised Learning: Regression

IBM

Learning Goals

In this section, we will cover:

- Types of Supervised Learning
- Examples of Regression and Classification Problems

Types of Supervised Learning

Regression

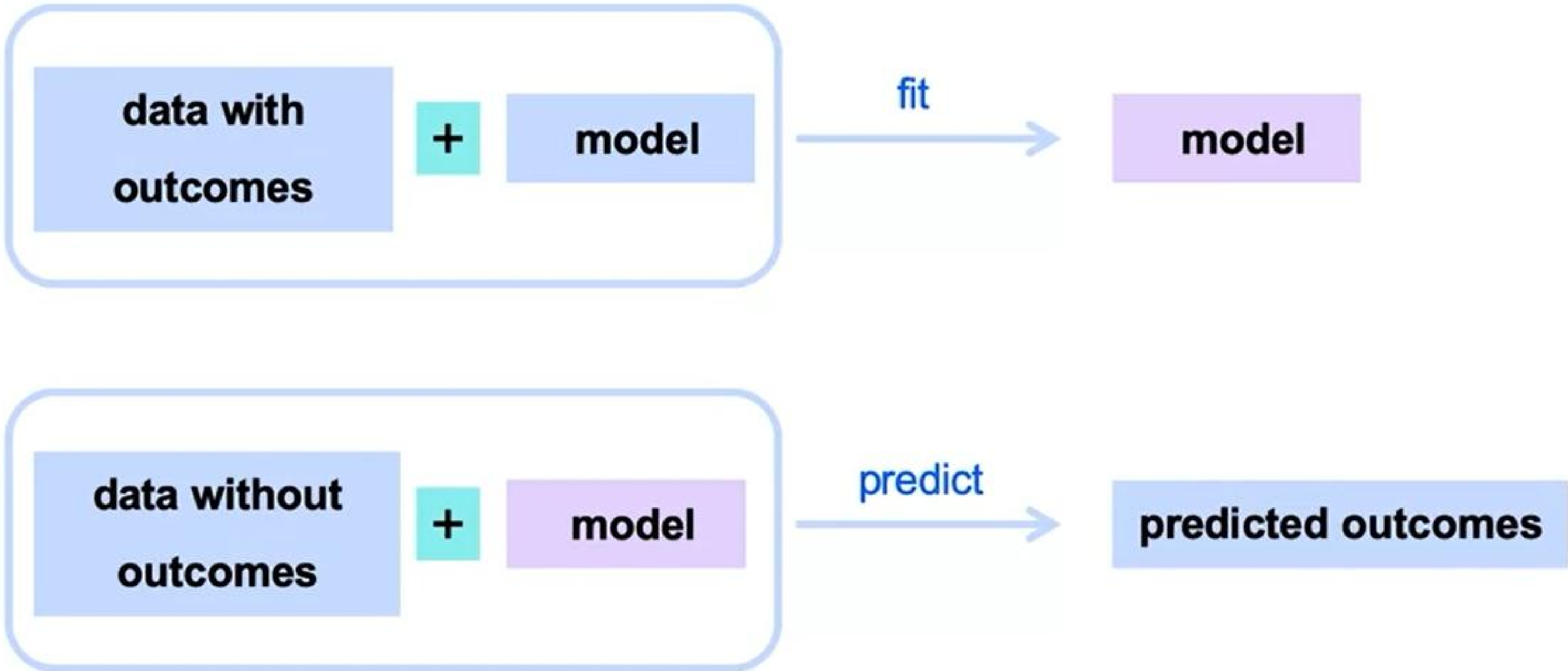
outcome is continuous (numerical)

Classification

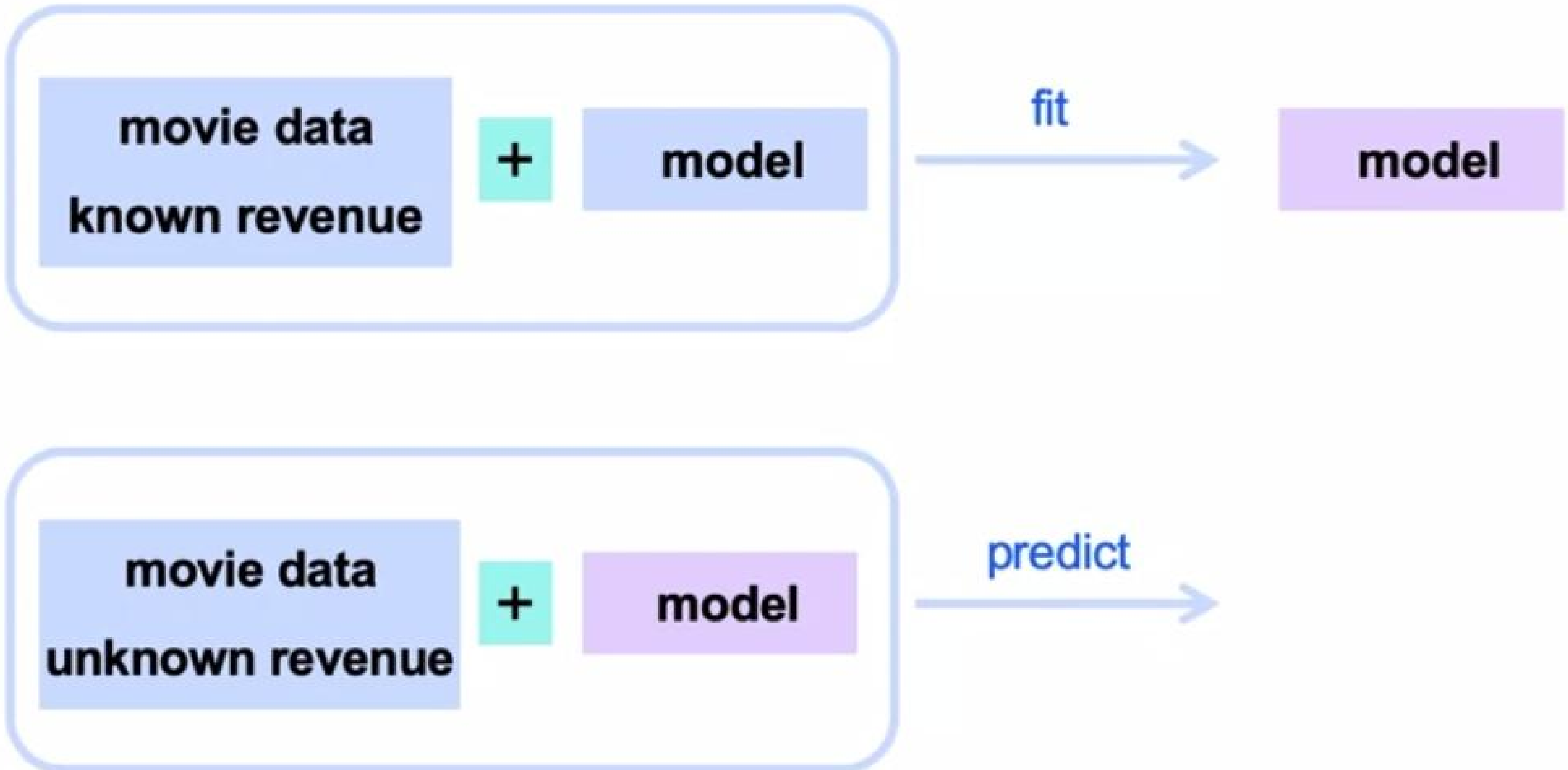
outcome is a category

IBM

Supervised Learning Overview

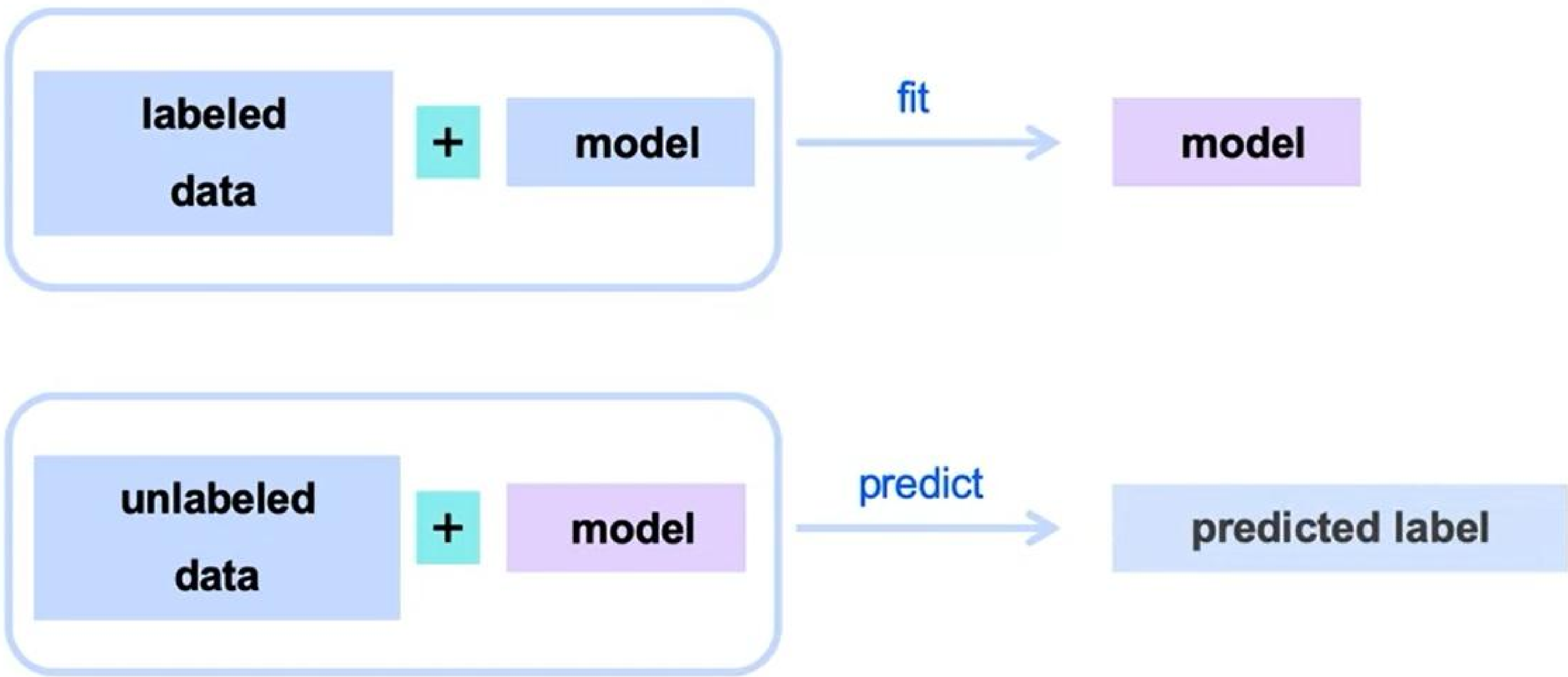


Numeric Predicting: Movie Revenue

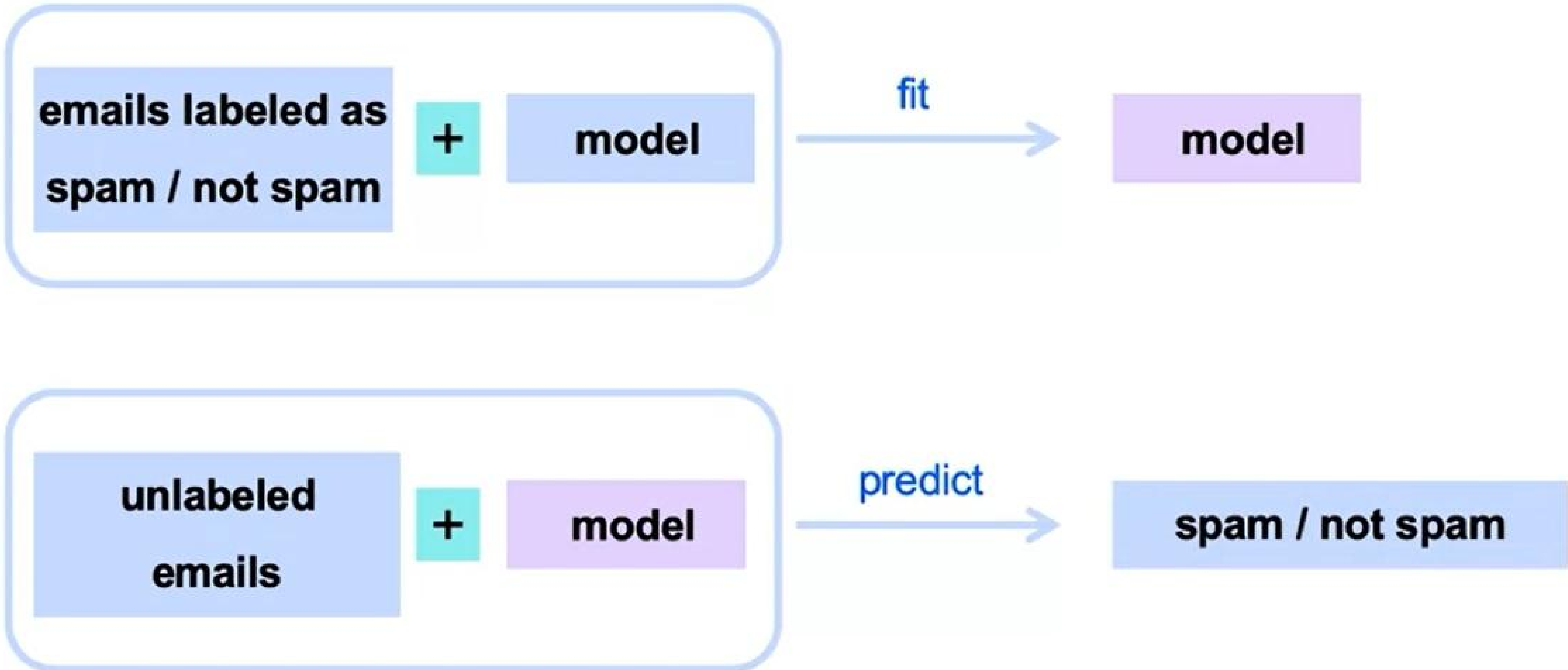


IBM

Classification: Categorical Answers



Classification: Predicting Spam Emails



What is Needed for Classification?

Model data with:

- Features that can be quantified
- Labels that are known
- Method to measure similarity

Learning Recap

In this section, we discussed:

- Types of Supervised Learning
- Examples of Regression and Classification Problems