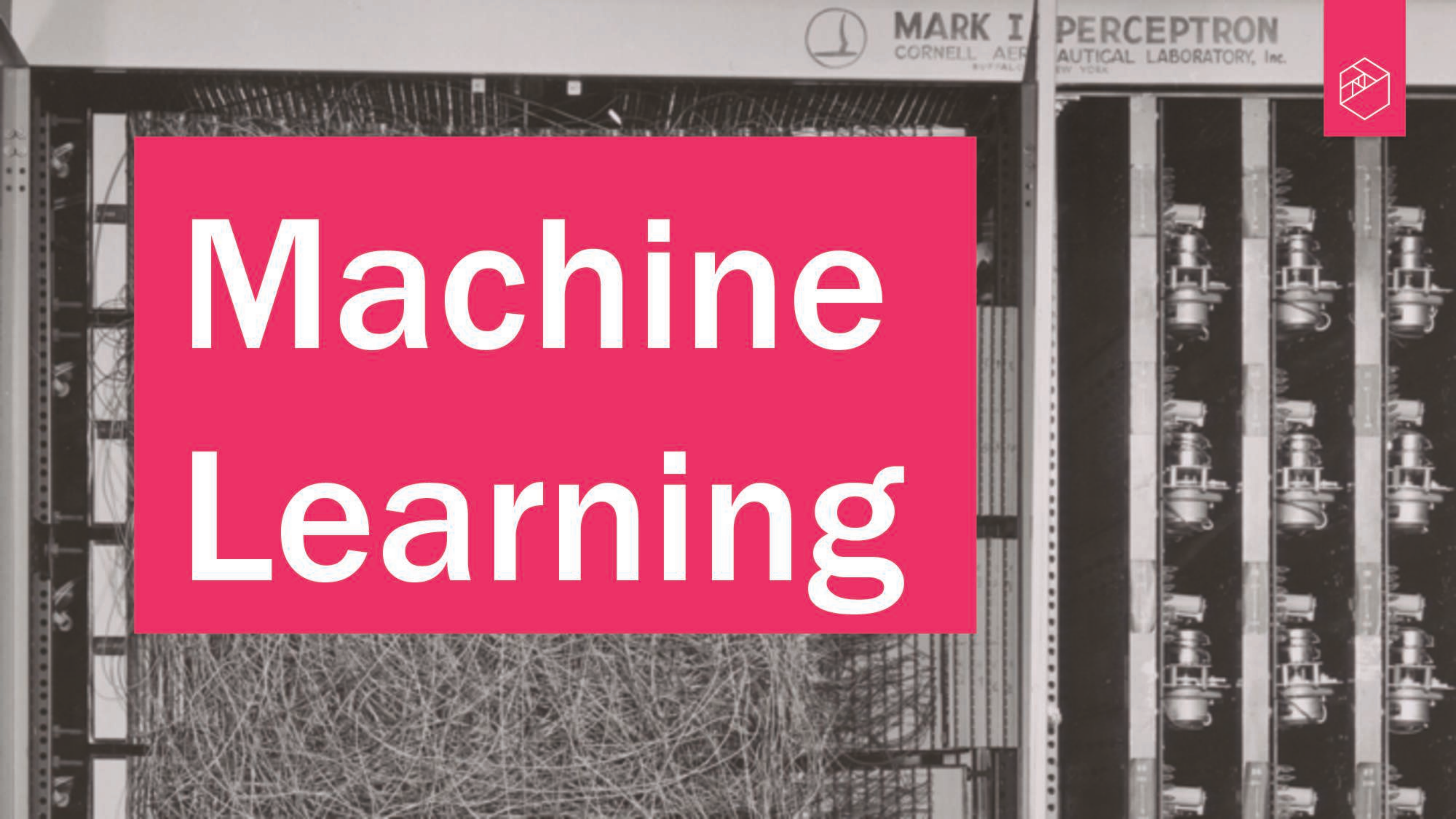
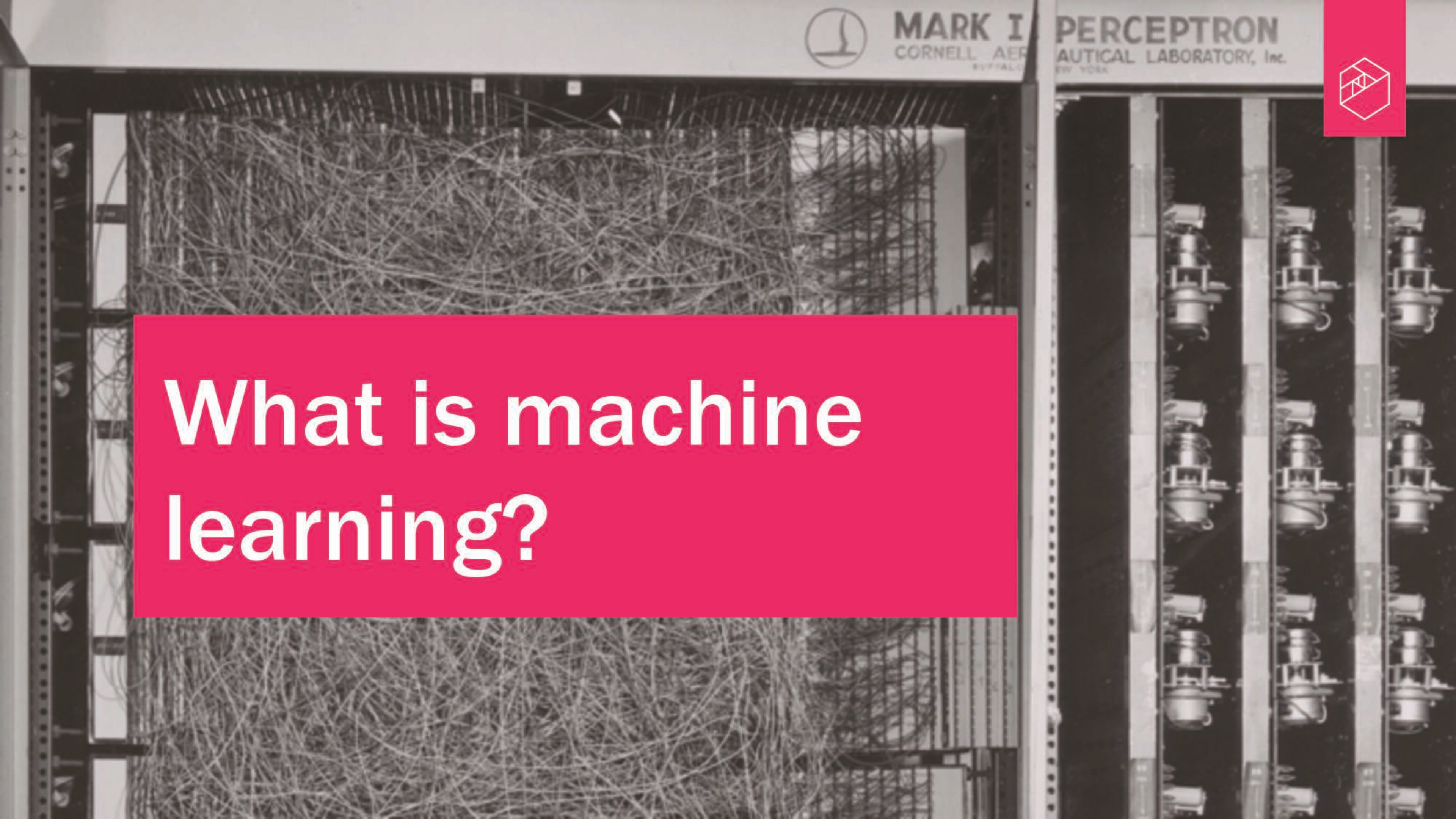
Machine Learning



What is machine learning?



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)

Figure 1.1 Some definitions of artificial intelligence, organized into four categories.

Russel & Norvig: Artificial Intelligence, A Modern Approach

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)

Figure 1.1 Some definitions of artificial intelligence, organized into four categories.

Russel & Norvig: Artificial Intelligence, A Modern Approach

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)

Russel & Norvig: Artificial Intelligence, A Modern Approach

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 study of **mental faculties** through the use of computational models." (Charniak and McDermott, 1985)
"[The automation of] activities that we associate with **human thinking**, activities such as **decision-making, problem solving, learning** . . ." (Bellman, 1978)	"The study of the computations that make it possible to **perceive, reason, and act**." (Winston, 1992)

Russel & Norvig: Artificial Intelligence, A Modern Approach

What is learning?

On Exactitude in Science
Jorge Luis Borges, *Collected Fictions*, translated by Andrew Hurley.

…In that Empire, the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations, who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast Map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars; in all the Land there is no other Relic of the Disciplines of Geography.

—Suarez Miranda,*Viajes de varones prudentes*, Libro IV,Cap. XLV, Lerida, 1658

Learn verb

\'lern\

to process past experience and update a model such that the the model is more useful for future experience

Learn verb

\’lern\

to process past experience and update a model such that the the model is useful for future experience

Learn verb

\’lern\

to process past experience and **update a model** such that the the model is useful for future experience

Learn verb

\'lern\

to process past experience and update a model

such that the the model is more useful for

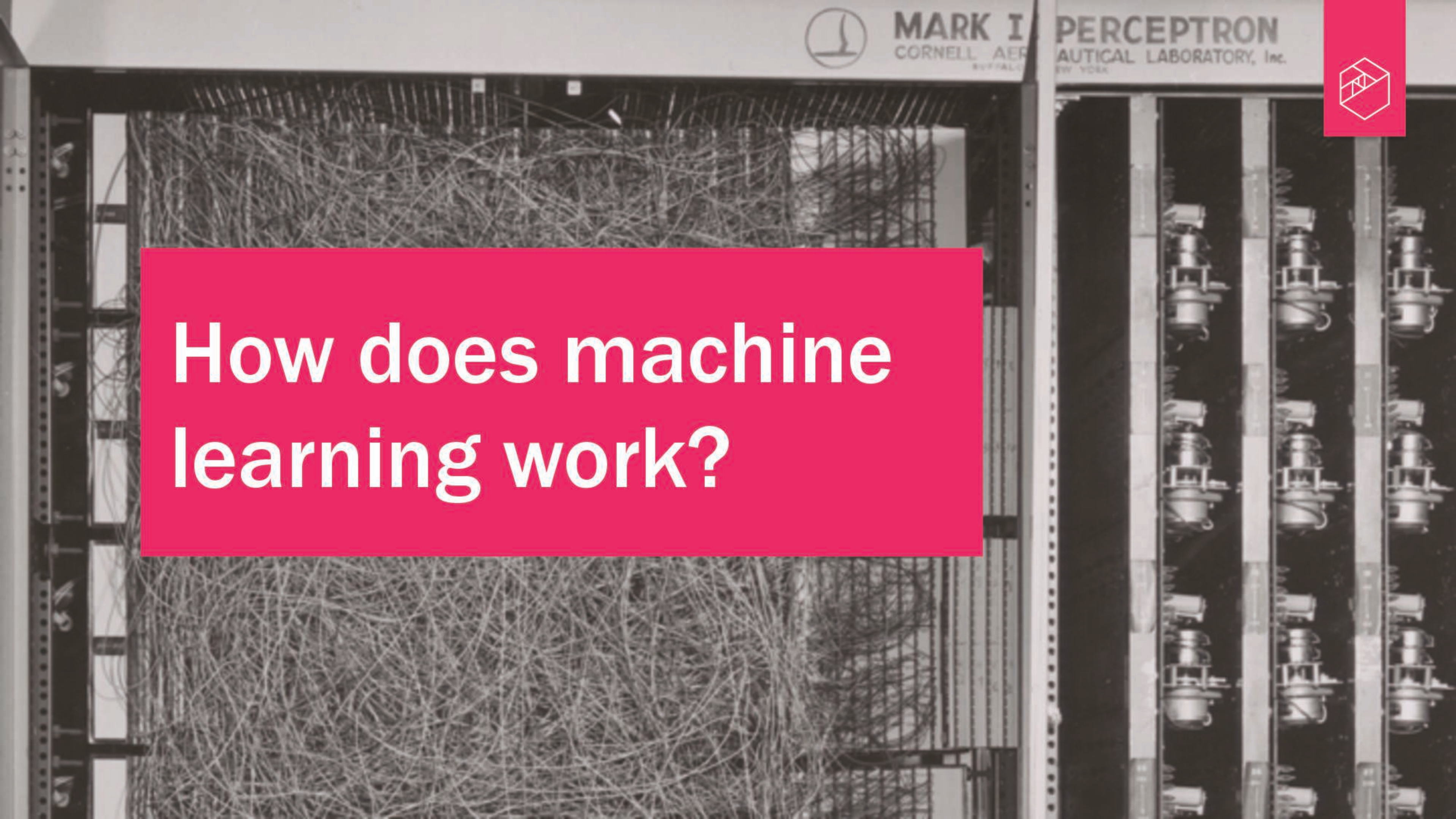
future experience

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.

Model: a learning algorithm

- Industry sometimes uses "algorithm" and "model" interchangeably.
  - Words are complicated (more on this in project 4)



How does machine learning work?

ML framework

Applies to every machine learning model

- You should eventually be able to explain what $y_p, f, \Omega,$ and x are for every model we cover in the bootcamp.
- Notation:
  - x and y are arrays with 1 or more rows and columns. A single example is just an array with 1 row. A single feature is just an arrow with 1 column.
  - Ω is standing in for 1 or more variables, rather than a single variable.

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

ML framework

Applies to every machine learning model

- x : Input
  - Observations: rows or examples the model will see.
  - Features: the different ways that we measure each observation
  - What were the observations and features in the MTA challenge?

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

ML framework

Applies to every machine learning model

- x : Input
  - Observations: rows or examples the model will see.
  - Features: the different ways that we measure each observation
  - What were the observations and features in the MTA challenge?

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

	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
0	A002	R051	02-00-00	59 ST	NQR456	BMT	08/27/2016	00:00:00	REGULAR	5799442	1966041
1	A002	R051	02-00-00	59 ST	NQR456	BMT	08/27/2016	04:00:00	REGULAR	5799463	1966044
2	A002	R051	02-00-00	59 ST	NQR456	BMT	08/27/2016	08:00:00	REGULAR	5799492	1966079
3	A002	R051	02-00-00	59 ST	NQR456	BMT	08/27/2016	12:00:00	REGULAR	5799610	1966155
4	A002	R051	02-00-00	59 ST	NQR456	BMT	08/27/2016	16:00:00	REGULAR	5799833	1966214

ML framework

Applies to every machine learning model

- x : Input
- y_p : Output
  - Be careful to keep separate y_p (the prediction of our model) and y (the observed values of y).
  - <u>Regression</u>: y is numeric

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

ML framework

Applies to every machine learning model

- x : Input
- y_p : Output
  - Be careful to keep separate y_p (the prediction of our model) and y (the observed values of y)
  - Regression: y is numeric
    - Stock price, customer churn, location (x,y coordinates)

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

ML framework

Applies to every machine learning model

- x : Input
- y_p : Output
  - Be careful to keep separate y_p (the prediction of our model) and y (the observed values of y)
  - Regression: y is numeric
    - Stock price, customer churn, location (longitude, latitude)
  - Classification: y is categorical

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

ML framework

Applies to every machine learning model

- x : Input
- y_p : Output
  - Be careful to keep separate y_p (the prediction of our model) and y (the observed values of y)
  - Regression: y is numeric
    - Stock price, customer churn, location (x,y coordinates)
  - Classification: y is categorical
    - Face recognition, winner of award, which word comes next

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

ML framework

Applies to every machine learning model

- x: Input
- y_p: Output
- Ω: Parameters
  - This is what changes as the model learns.
  - Some models have many different parameters, some have very few.
  - Hyper-parameter: A parameter that is not learned directly from the data.
    - Later we will discuss techniques for using model performance to inform hyperparameter selection.

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

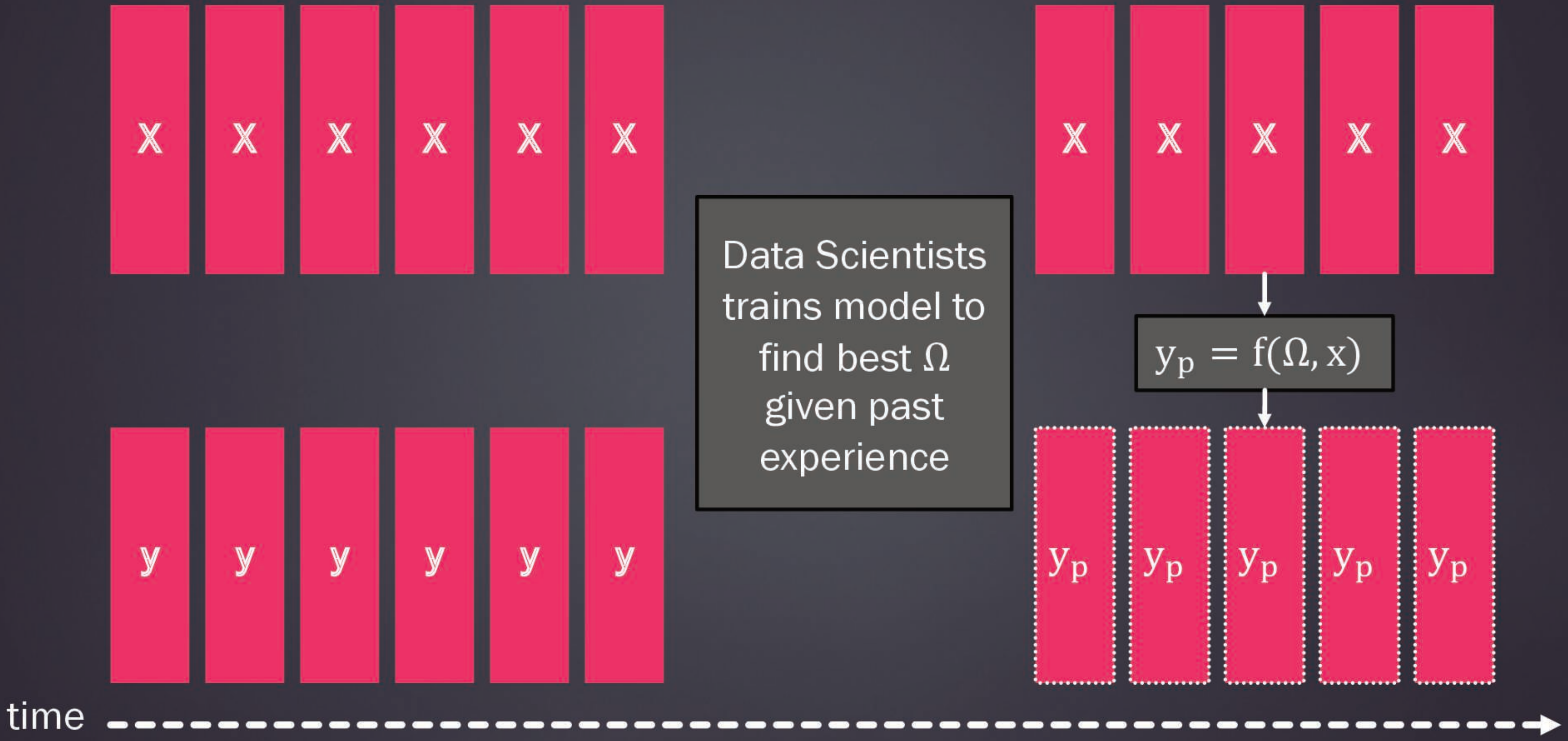
ML framework

Applies to every machine learning model

- x: Input
- y_p: Output
- Ω: Parameters
- $f(\cdot)$: Prediction function
  - Generates predictions from x and Ω

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

ML framework



ML framework

Applies to every machine learning model

- x: Input
- y_p: Output
- Ω: Parameters
- $f(\cdot)$: Prediction function

Applies to most machine learning models

- $J(y, y_p)$: 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.

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

ML framework

Applies to every machine learning model

- x: Input
- y_p: Output
- Ω: Parameters
- $f(\cdot)$: Prediction function

Applies to most machine learning models

- $J(y, y_p)$: Loss
- Update rule:
  - Given observed y, x, the update rule determines how to update Ω
  - Typically, finds the Ω that minimizes $J(y, f(\Omega, x))$

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

ML framework

Applies to every machine learning model

- x: Input
- y_p: Output
- Ω: Parameters
- $f(\cdot)$: Prediction function

Applies to most machine learning models

- $J(y, y_p)$: Loss
- Update rule

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

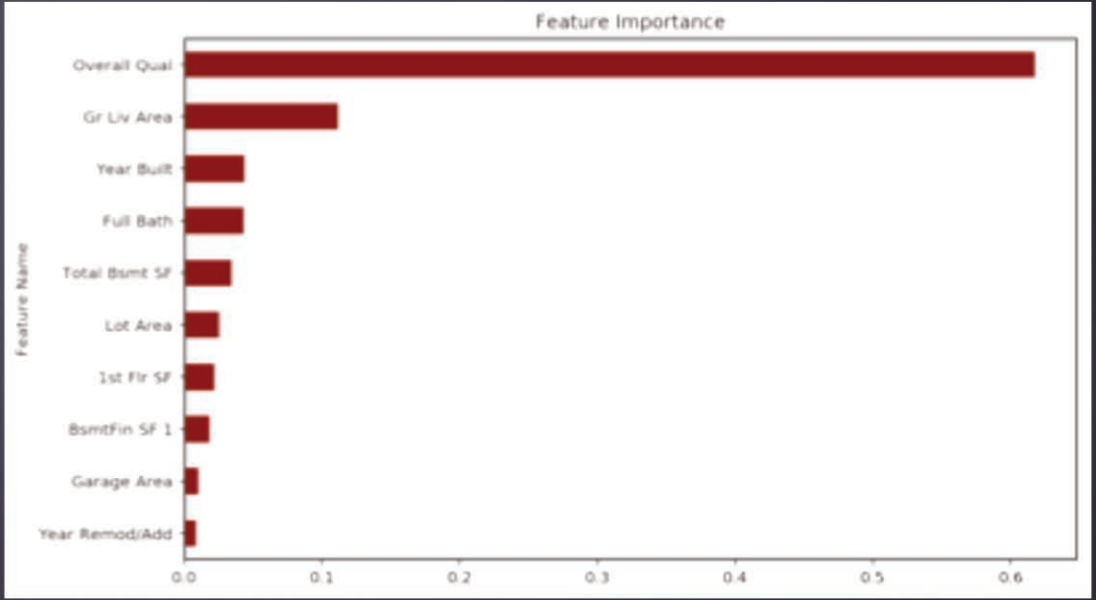
What can we do with Machine Learning?

Two common approaches

- Interpretation
  - In some cases, the primary concern what the model tells us.
  - In $y_p = f(\Omega, x)$, the interpretation approach uses Ω to give us insight into a system.
  - Example workflow:
    - Gather x, y
    - Train a model by finding the Ω that gives the best prediction $y_p = f(\Omega, x)$
    - Discard y_p. Peer into Ω and see what it tells us.

Two common approaches

- Interpretation
  - Example: Looking for insight into housing sale prices in Ames, Iowa. Which features are most important?



Two common approaches

- Interpretation
- Prediction
  - AKA “bake off”
  - In some cases, the primary concern is making the best prediction.
  - The focus is on optimizing certain performance metrics, such as accuracy or r^2 (more on this later).
  - ex: Kaggle, KDD Cup
  - <u>Black box model</u>: A model with little-to-no interpretability.

Two common approaches

- Interpretation
- Prediction
  - Example: Build an app that predicts the eventual sale price of homes in Ames, Iowa.

Two common approaches

- Interpretation
- Prediction
- Combined
  - Majority of projects will call for a balance.
  - Interpretation can provide insight into improvements in prediction and vice-versa.

Takeaways

- Machine Learning is the subset of AI that focuses on model building
- ML algorithms
  - Use past experience to update a model that is useful for future experience.
  - Follow a general form: $y_p = f(\Omega, x)$
- Models can be used for interpretation and/or prediction