Week – 1I
Intro to ML

ML Bootcamp 2021

GDG Cloud EDMONTON

AlbertaAI

Careera Analytics Lab

Slido: #ml_bootcamp_week2

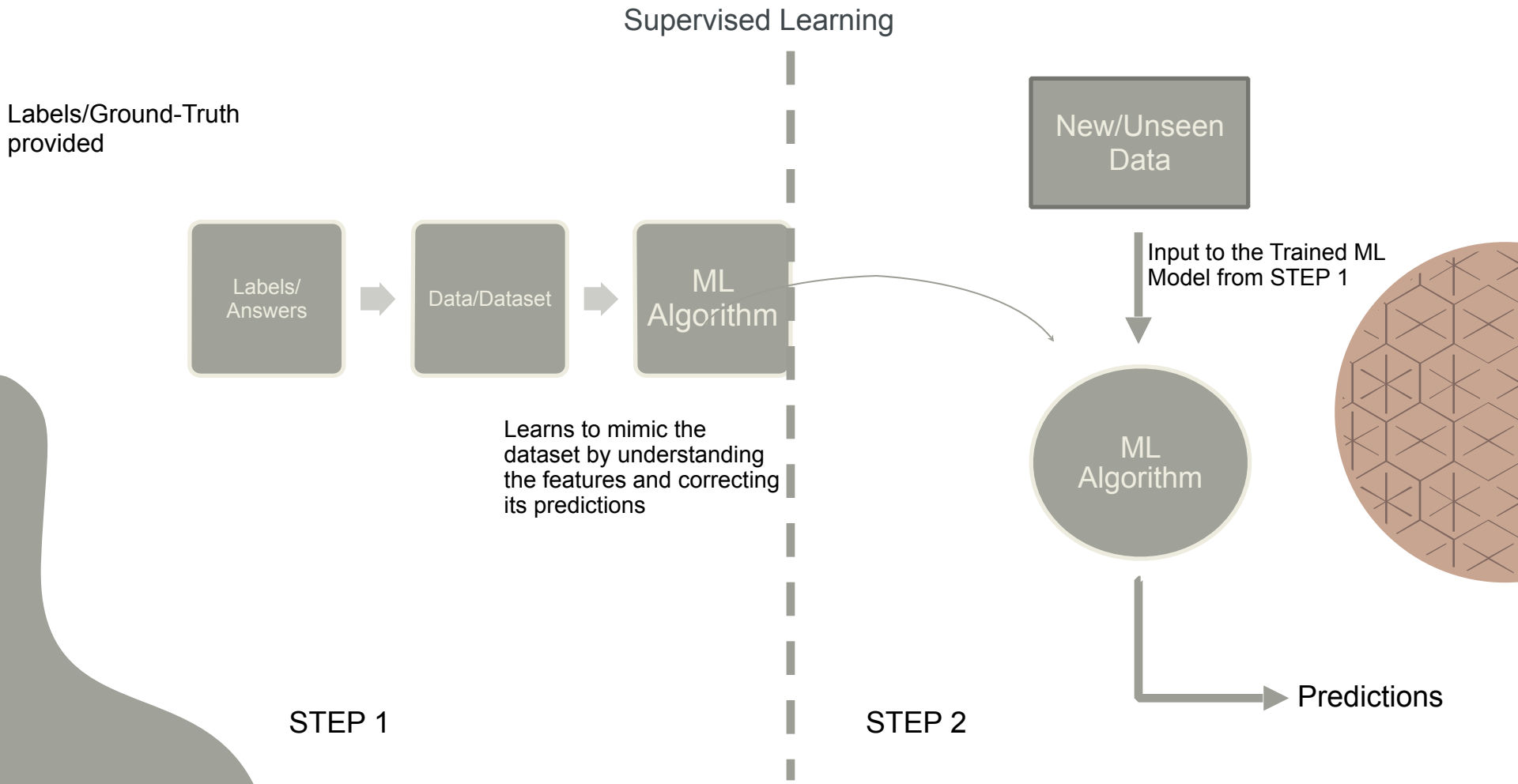
Table of contents

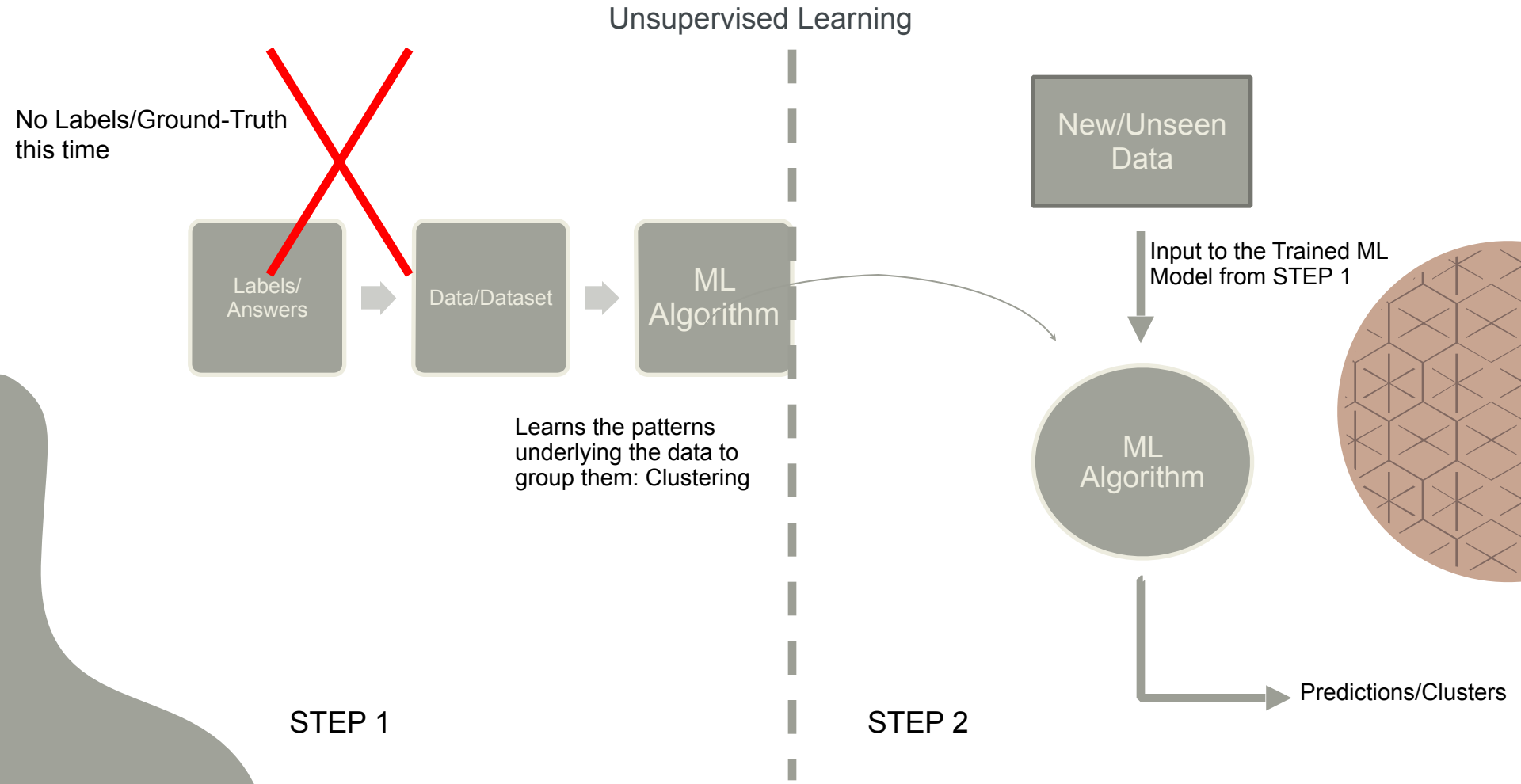
Slido: #ml_bootcamp_week2

Quick Refresher on ML Concepts

- **Datasets:** The data that you process, transform and use to train your ML models
- **Features:** The individual columns(if it is a table) in your data
- **Supervised/Unsupervised:** Either with or without answers
  - **Classification:** Predicting a limited set of outputs Eg: Pos/Neg sentiment, Cat/Dog, Bearish/Bullish
  - **Regression:** Output has an infinite possibilities Eg: Housing prices, stock prices, time etc.
- **Machine Learning:** A set of algorithms or a method to learn through experience using data
- **Training and Testing:** We usually split our data into two parts, one to train the model and another to test the model's performance
- **Decision Boundary:** The line separating the classes for a given ML problem
- **Hypothesis:** The model that you would like to determine
- **Weights/Parameters:** This is what defines your model as these are the values that are adjusted to train your model to make accurate predictions
- **Ground-Truth:** The labels/answers that are provided along with the dataset

Slido: #ml_bootcamp_week2


Supervised Learning
Labels/Ground-Truth provided
Labels/ Answers
Data/Dataset
ML Algorithm
Learns to mimic the dataset by understanding the features and correcting its predictions
New/Unseen Data
Input to the Trained ML Model from STEP 1
ML Algorithm
Predictions
STEP 1
STEP 2


Unsupervised Learning
No Labels/Ground-Truth this time
Labels/ Answers
Data/Dataset
ML Algorithm
Learns the patterns underlying the data to group them: Clustering
New/Unseen Data
Input to the Trained ML Model from STEP 1
ML Algorithm
Predictions/Clusters
STEP 1
STEP 2

Classification Vs Regression

Problem Definition	Classification or Regression?
Predicting the sentiment of a given Amazon Review. **Output/ Ground-Truth:** 1 → Positive 0 → Negative	
Weather Forecasting **Output/ Ground-Truth:** The temperature in Celsius or Farenheit	
Stock Market Price prediction **Output/ Ground-Truth:** The next day's stock price of a given company	
Next word prediction in E-mail: **Output/ Ground-Truth:** The next word based on the previous words	

Slido: #ml_bootcamp_week2

Linear Regression

1. $$y = \beta_o + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3 \ ...$$

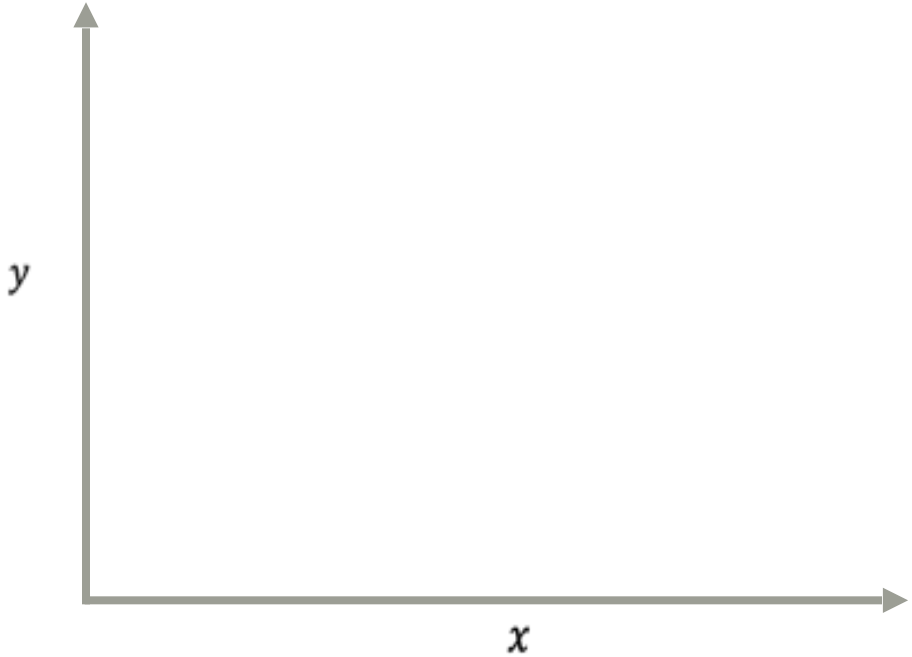
Dependent Variable/
Output/ Prediction

Coefficients / Weights

Independent Variable/
Input/ Predictors/
Features

This is an equation that is “Linear” in the input parameters β
Goal here is to find the best β that will get you close to the ground-truth y values

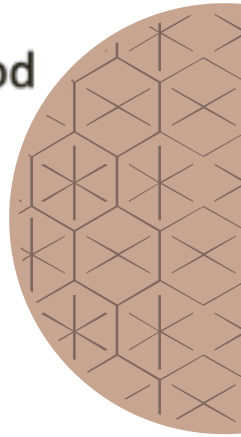
How do we model Linear Regression?



Slido: #ml_bootcamp_week2

How to find the best β ?

1. Linear Algebra, for example by using matrix inversion to find β directly
2. Ordinary Least Squares Method (which follows the Maximum Likelihood Estimation Framework)
3. Gradient Descent



Slido: #ml_bootcamp_week2

Gradient Descent

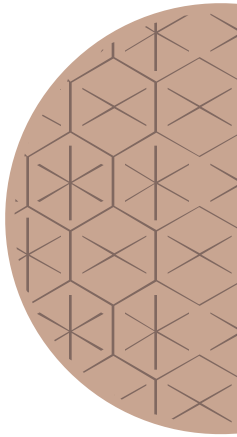
1.

$$\text{Cost Function } J(\theta) = \frac{1}{2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

These are the parameters like β earlier

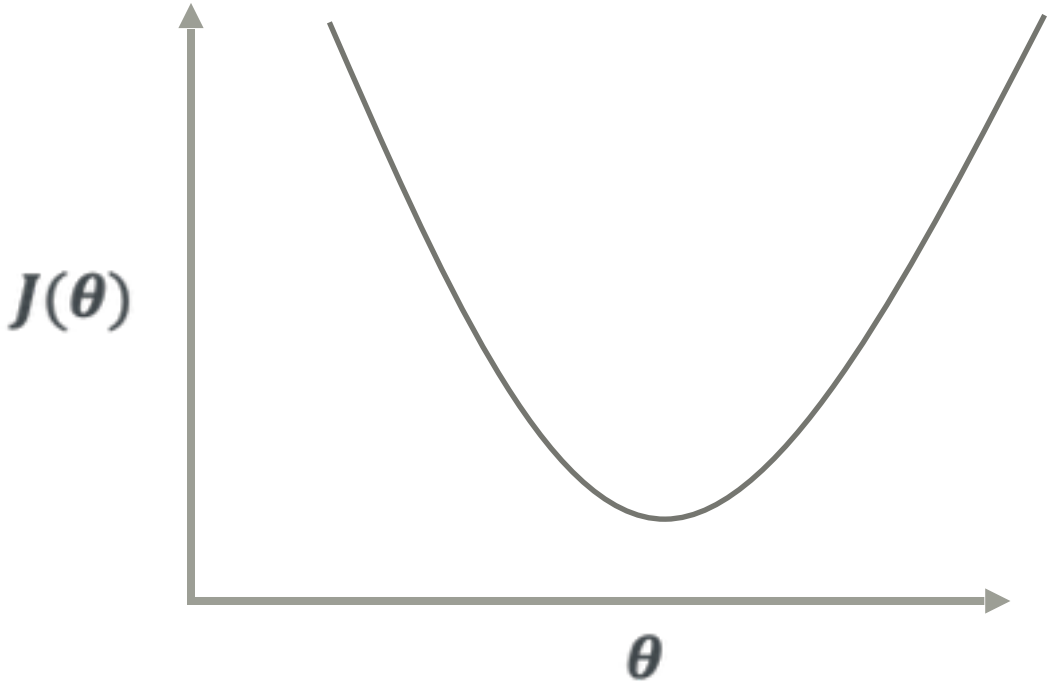
The ground-truth value

The model's predicted value



Slido: #ml_bootcamp_week2

Convex Optimization Problem



Slido: #ml_bootcamp_week2

Logistic Regression

1. The name is contradicting but this algorithm is meant to solve classification problems
- Let's look at our previous model for Linear Regression

$$y = \beta_o + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3 \ ...$$

- How can we transform the above model such that we can address the problem of Binary Classification?

$$y = f(\beta_o + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3 \ ...)$$

where,

$$f(x) = \frac{1}{1+e^{-x}}$$

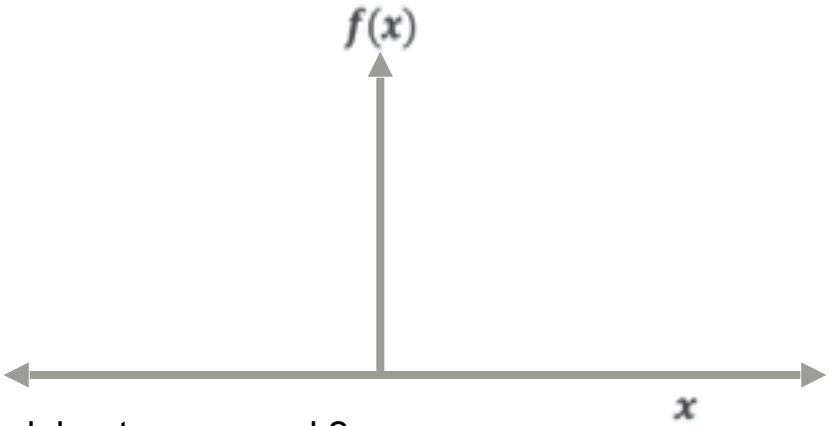
→ Logit Function

Expanding a bit further

1.

$$y = f(\beta_o + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3 \ldots)$$

$$= \frac{1}{1+e^{-(\beta_o + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3 \ldots)}} = \begin{cases} 1 \ if \ y \geq 0.5 \\ 0 \ if \ y < 0.5 \end{cases}$$

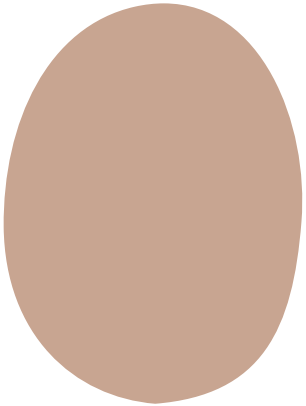


Slido: #ml_bootcamp_week2

Summary

- **Machine Learning** is a set of algorithms that learn from the data
- **Supervised learning** is a class of problems where the ground-truth/labels are provided
- **Unsupervised learning** problems do not have any explicit labels and the algorithms explore the underlying structure
- **Classification** problems have a finite set of outputs whereas **Regression** problems deal with the space of real numbers as outputs
- **Linear Regression**: An algorithm originating from Linear Algebra that is used for Regression problems. Eg: Weather Forecasting, House price prediction and Stock market trend analysis
- **Logistic Regression**: An algorithm that converts Linear Regression formulation into a classification problem. Eg: Sentiment Analysis, Titanic survival prediction etc.

Lets start Practicing!



Slido: #ml_bootcamp_week2

Thanks

Do you have any questions?


slidesgo