NDL GBM #3

Bootcamp Day 2

Modeling

9/21/25

Attendance: Meeting Name “Modeling”

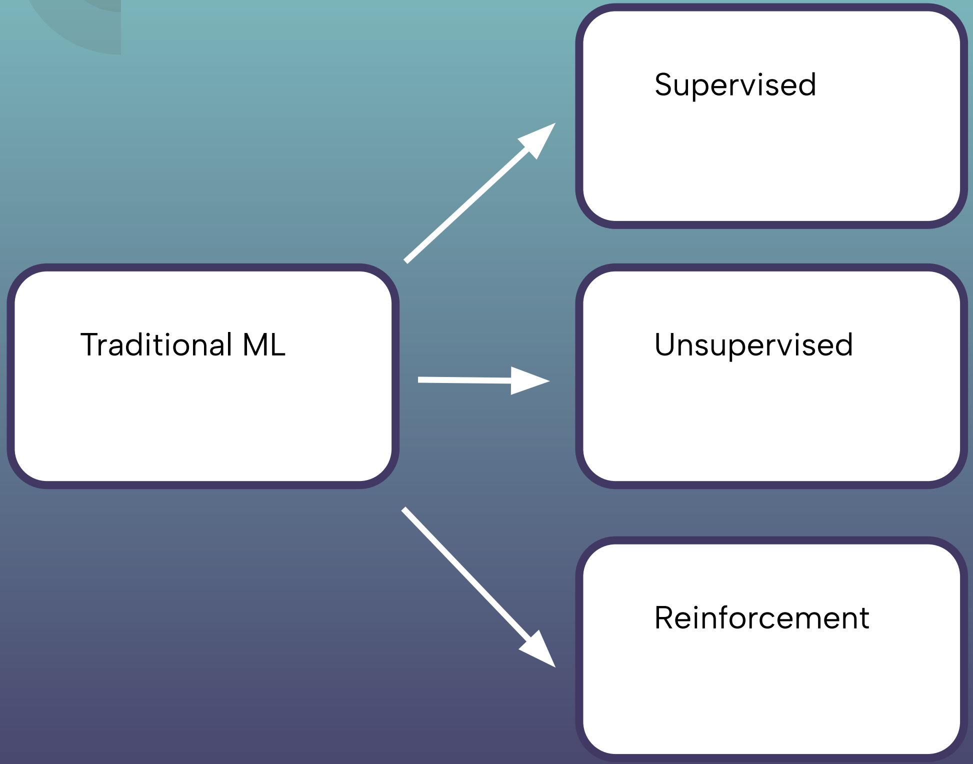
What is ML?

Traditional ML: Using the known to predict the unknown by training an algorithm

What kind can achieve our goal?

Regression - predicting quantitative variables

Where does regression fit?



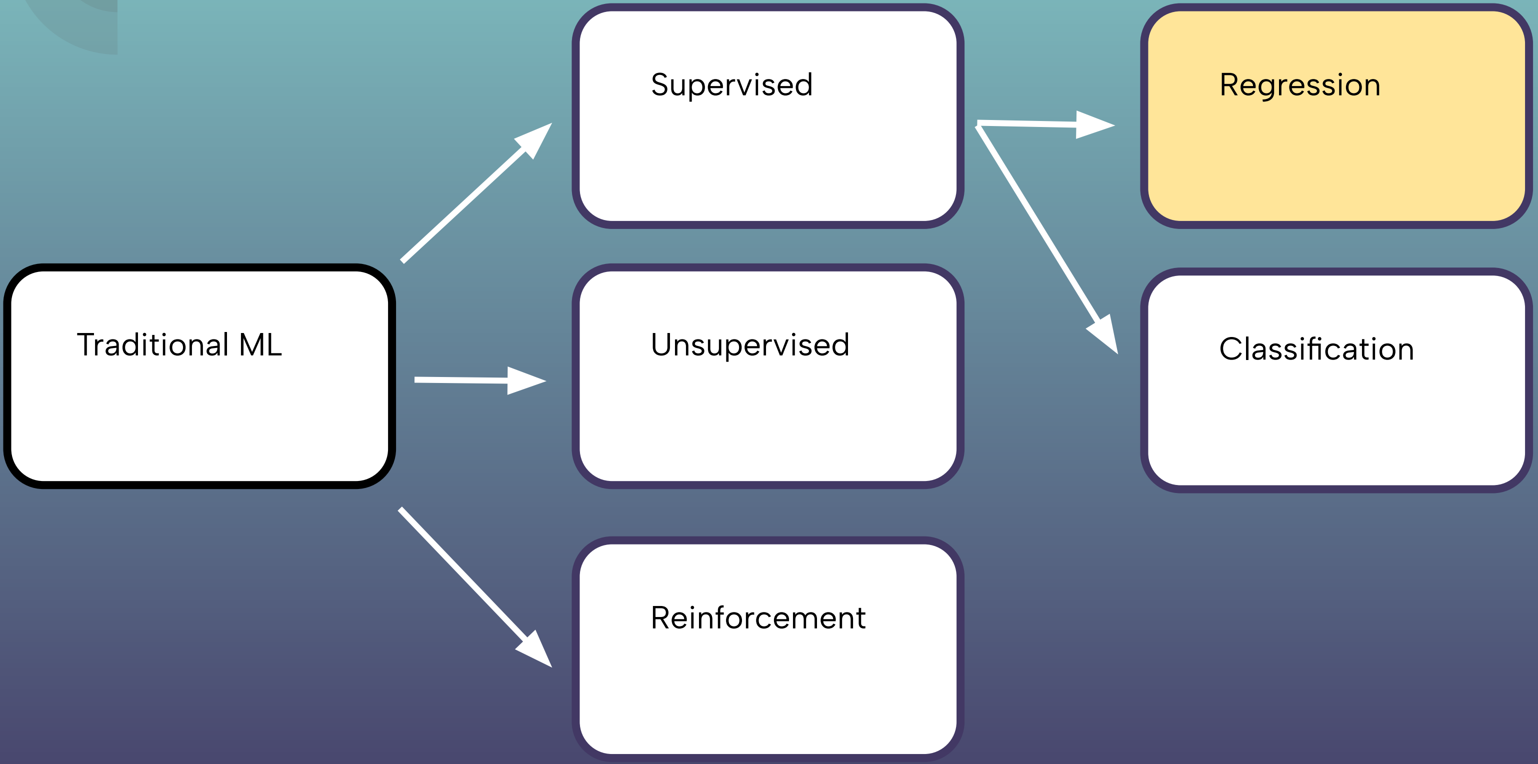
Traditional ML

Supervised

Unsupervised

Reinforcement

Where does regression fit?



Traditional ML

Supervised

Unsupervised

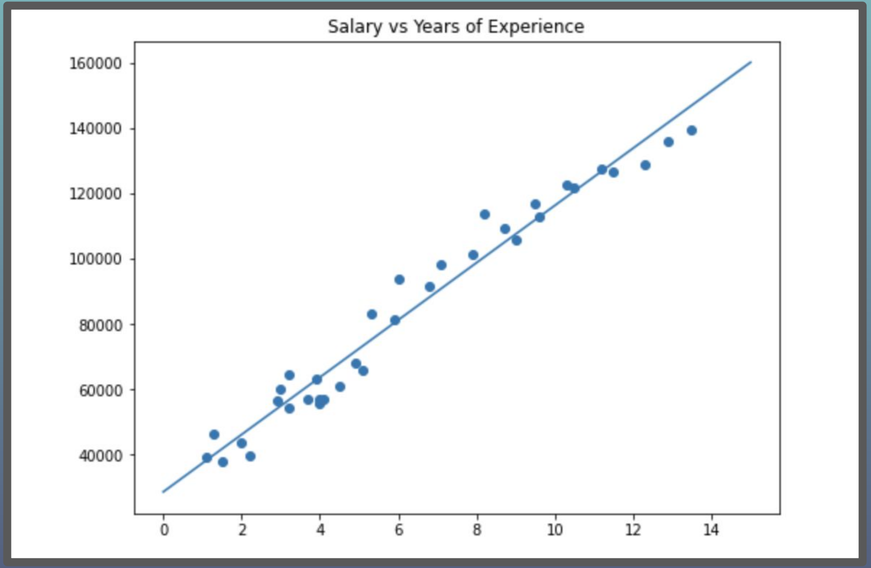
Reinforcement

Regression

Classification

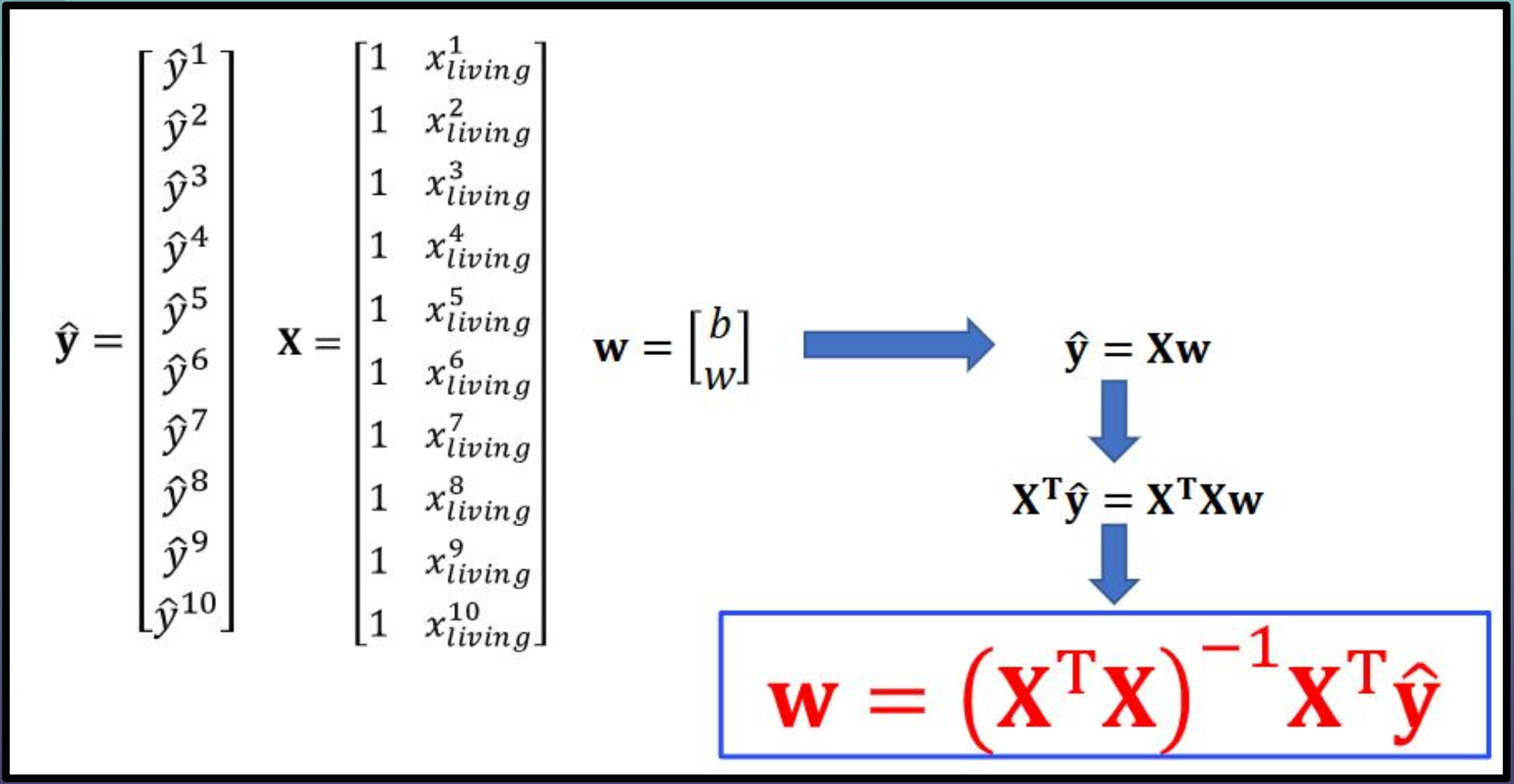
Simple linear regression

$$y = mx + b$$



$$\text{Salary} = 1000(\text{Y.o.E}) + 1000$$

Getting m and b

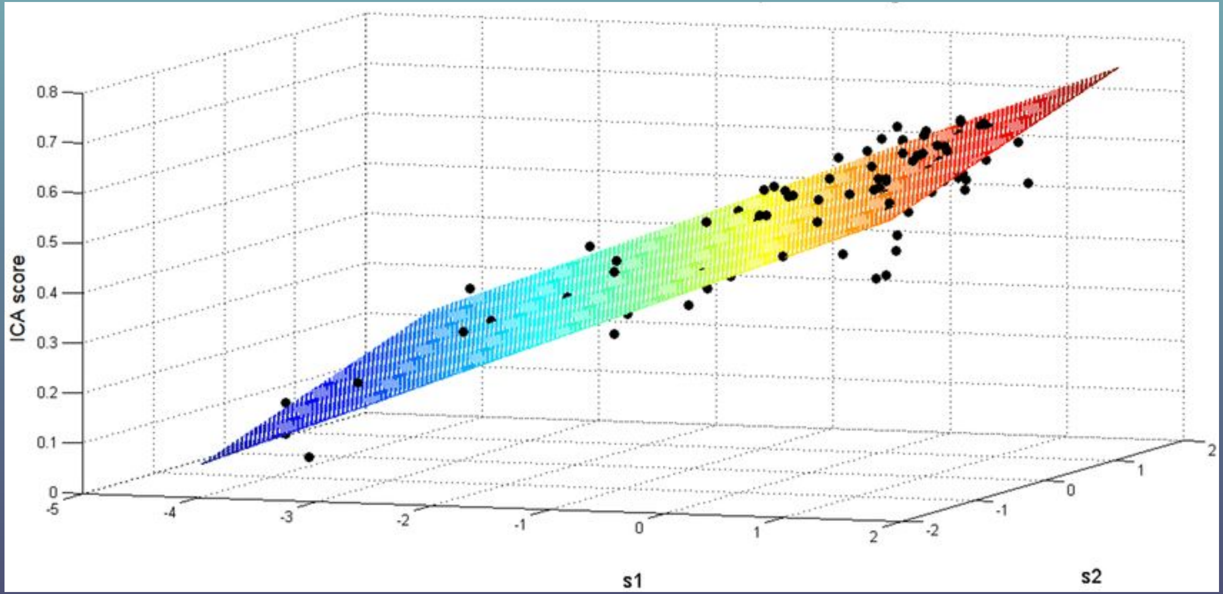


$$\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}^1 \\ \hat{y}^2 \\ \hat{y}^3 \\ \hat{y}^4 \\ \hat{y}^5 \\ \hat{y}^6 \\ \hat{y}^7 \\ \hat{y}^8 \\ \hat{y}^9 \\ \hat{y}^{10} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{living}^1 \\ 1 & x_{living}^2 \\ 1 & x_{living}^3 \\ 1 & x_{living}^4 \\ 1 & x_{living}^5 \\ 1 & x_{living}^6 \\ 1 & x_{living}^7 \\ 1 & x_{living}^8 \\ 1 & x_{living}^9 \\ 1 & x_{living}^{10} \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} b \\ w \end{bmatrix}$$

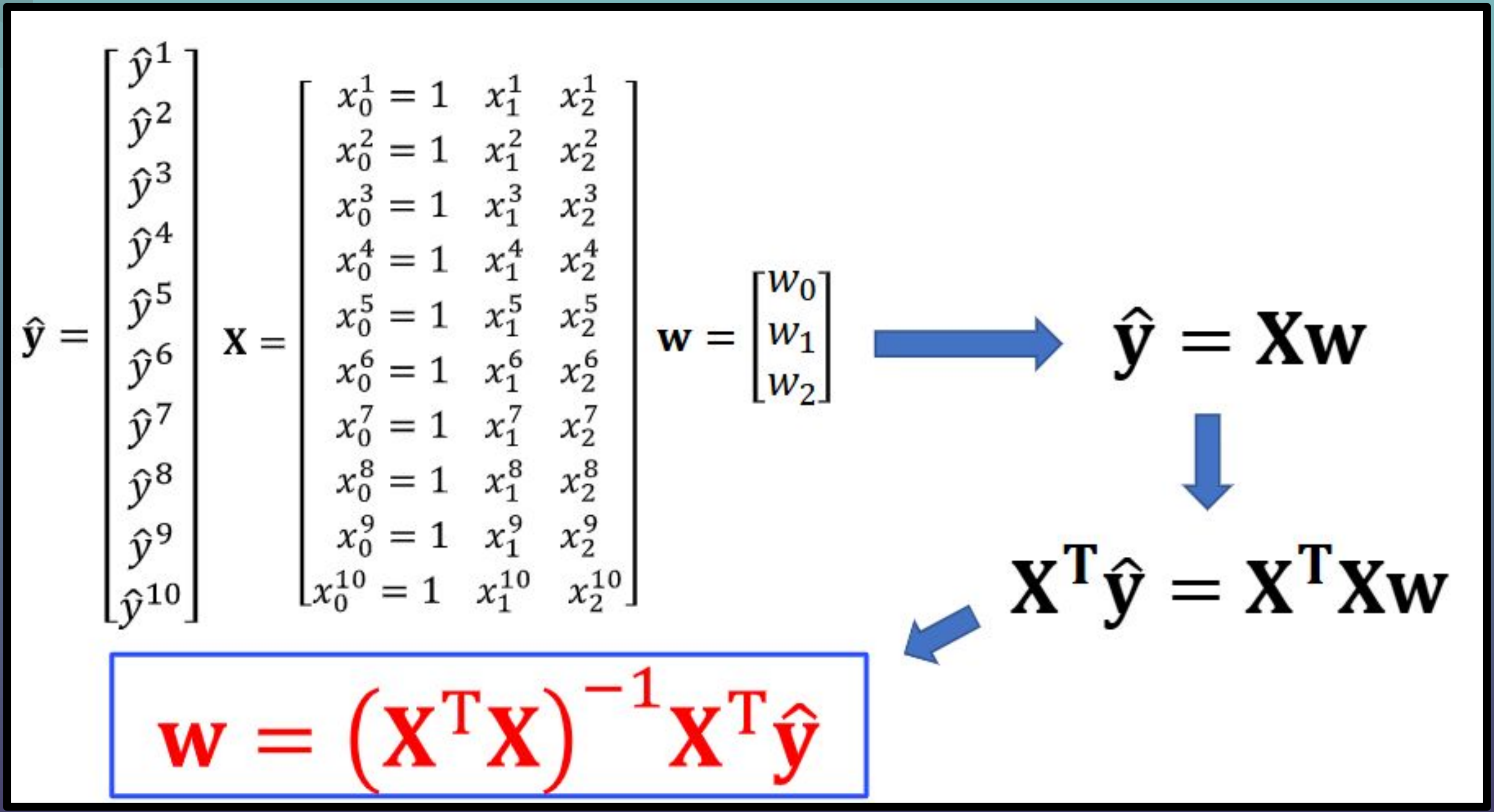
$$\begin{tikzcd}
 \arrow[r] & \hat{\mathbf{y}} = \mathbf{X}\mathbf{w} \arrow[d] \\
& \mathbf{X}^{\mathrm{T}}\hat{\mathbf{y}} = \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w} \arrow[d] \\
& \boxed{\mathbf{w} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{y}}}
\end{tikzcd}$$

Multi-Linear Regression

$$y = m_1 x_1 + m_2 x_2 + b$$



Multi-linear regression



$$\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}^1 \\ \hat{y}^2 \\ \hat{y}^3 \\ \hat{y}^4 \\ \hat{y}^5 \\ \hat{y}^6 \\ \hat{y}^7 \\ \hat{y}^8 \\ \hat{y}^9 \\ \hat{y}^{10} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_0^1 = 1 & x_1^1 & x_2^1 \\ x_0^2 = 1 & x_1^2 & x_2^2 \\ x_0^3 = 1 & x_1^3 & x_2^3 \\ x_0^4 = 1 & x_1^4 & x_2^4 \\ x_0^5 = 1 & x_1^5 & x_2^5 \\ x_0^6 = 1 & x_1^6 & x_2^6 \\ x_0^7 = 1 & x_1^7 & x_2^7 \\ x_0^8 = 1 & x_1^8 & x_2^8 \\ x_0^9 = 1 & x_1^9 & x_2^9 \\ x_0^{10} = 1 & x_1^{10} & x_2^{10} \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{w}$$

$$\mathbf{X}^{\mathrm{T}}\hat{\mathbf{y}} = \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w}$$

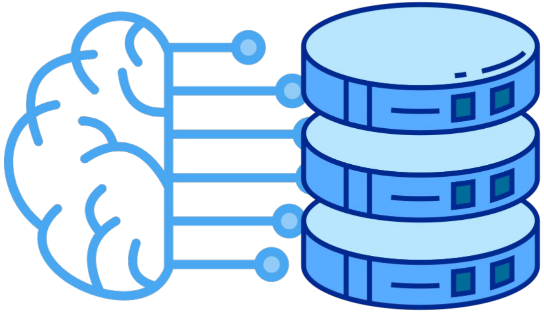
$$\boxed{\mathbf{w} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{y}}}$$

Data Split

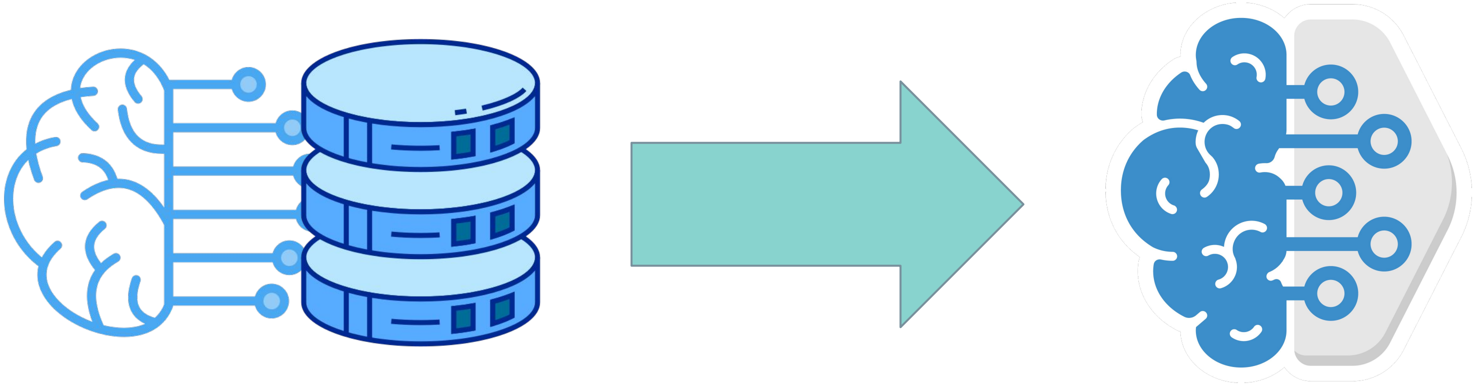
- Train – fit the model
  - Validation – optimize the model
- Test – Evaluate the model

How's it done?

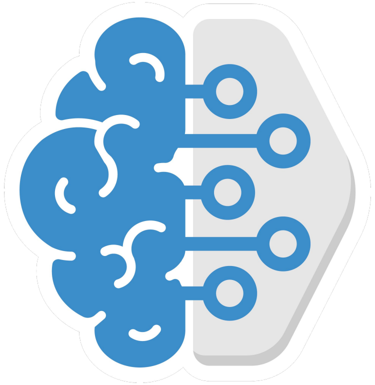
How’s it done?



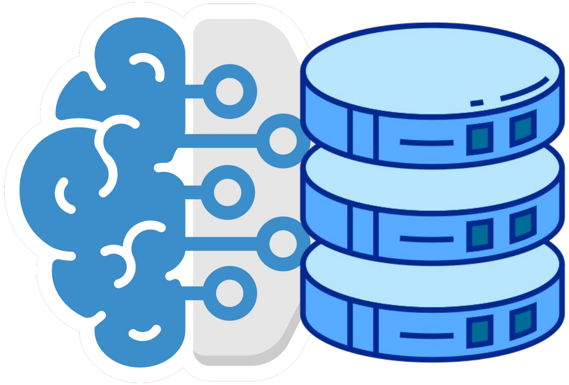
How's it done?



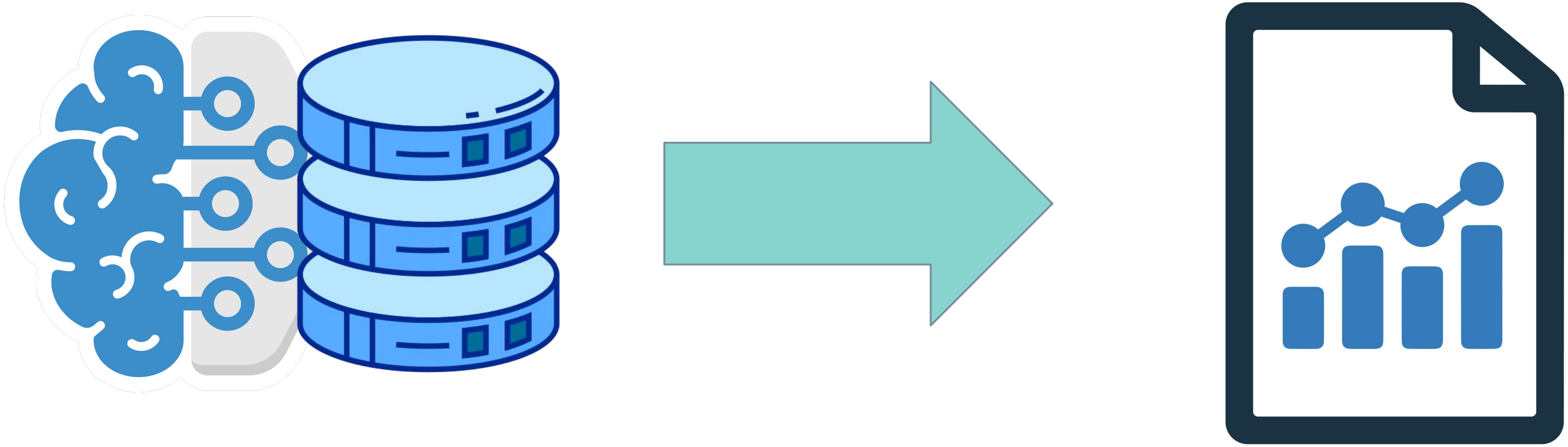
How’s it done?



How’s it done?



How's it done?



How effective was it?

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} \left| y_i - \hat{y}_i \right|$$

Better interpretability

Punishes high errors more

$$MSE = \frac{1}{n}\Sigma\left(y - \hat{y}\right)^2$$

Example of difference in MAE/MSE

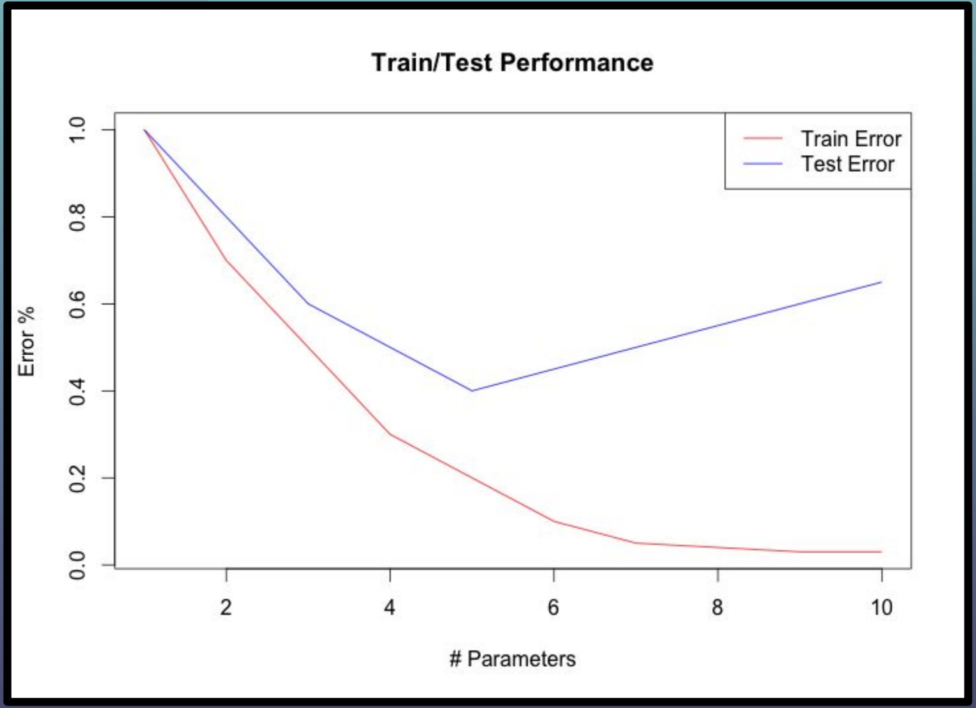
Situation 1:

Pred = 0.5
Actual = 0.7
MAE = 0.2
MSE = 0.16

Situation 2:

Pred = 5
Actual = 7
MAE = 2
MSE = 4

Overfitting

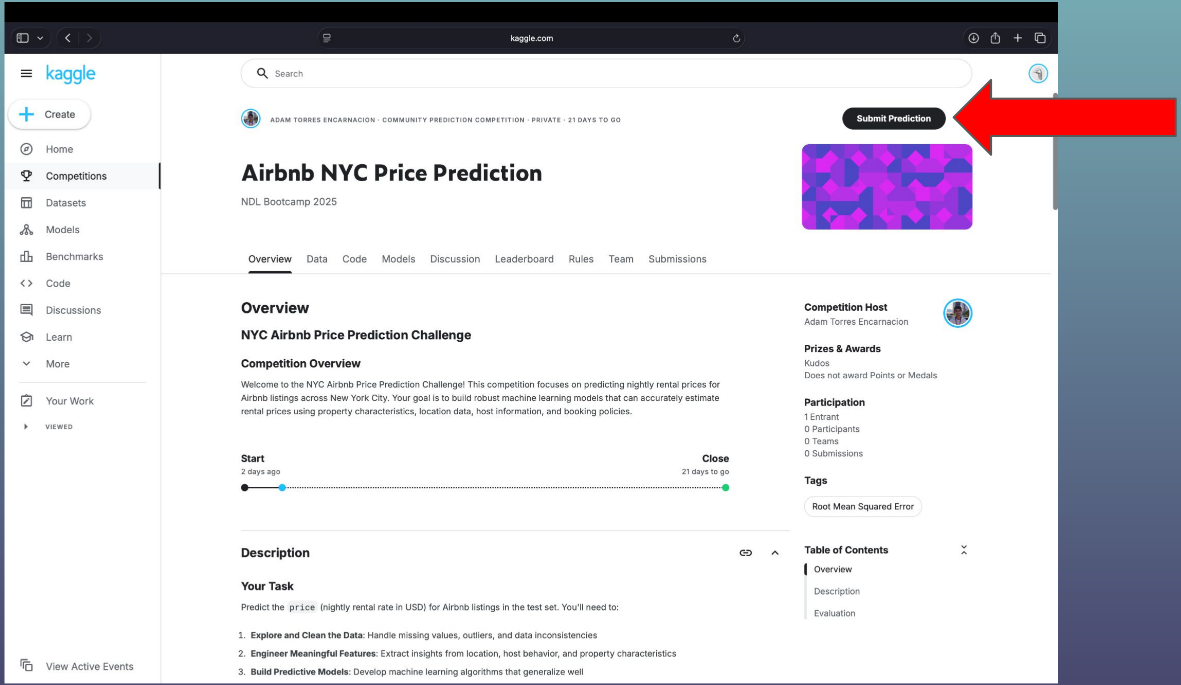


Model becomes too used to the training error that the testing error increases

Competition Details

- $100 to winner, $50 to 2nd, $25 to 3rd
- ~a week and a half
- Requirements for certificate:
  - Come to all three bootcamp days **OR**
  - Come to at least one day and get a error of less than 449.4 in the competition

Kaggle Submissions



ADAM TORRES ENCARNACION · COMMUNITY PREDICTION COMPETITION · PRIVATE · 21 DAYS TO GO

Airbnb NYC Price Prediction

NDL Bootcamp 2025

Overview

NYC Airbnb Price Prediction Challenge

Competition Overview

Welcome to the NYC Airbnb Price Prediction Challenge! This competition focuses on predicting nightly rental prices for Airbnb listings across New York City. Your goal is to build robust machine learning models that can accurately estimate rental prices using property characteristics, location data, host information, and booking policies.

Start 2 days ago

Close 21 days to go

Description

Your Task

Predict the `price` (nightly rental rate in USD) for Airbnb listings in the test set. You'll need to:

1. **Explore and Clean the Data**: Handle missing values, outliers, and data inconsistencies
2. **Engineer Meaningful Features**: Extract insights from location, host behavior, and property characteristics
3. **Build Predictive Models**: Develop machine learning algorithms that generalize well

Competition Host
Adam Torres Encarnacion

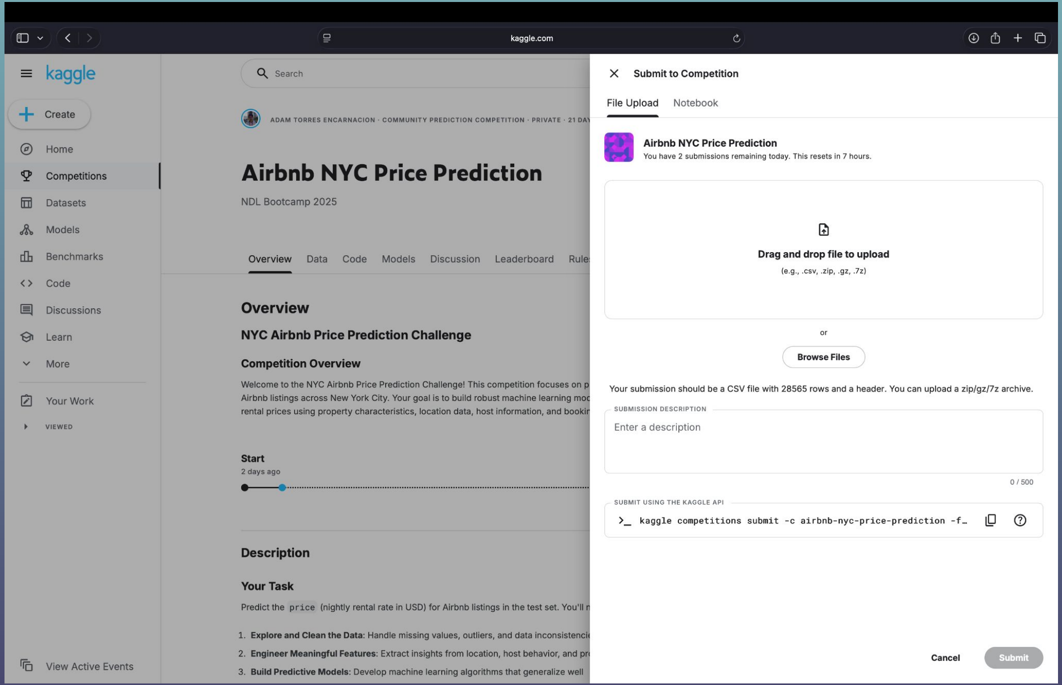
Prizes & Awards
Kudos
Does not award Points or Medals

Participation
1 Entrant
0 Participants
0 Teams
0 Submissions

Tags
Root Mean Squared Error

Table of Contents
Overview
Description
Evaluation

Kaggle Submissions



ADAM TORRES ENCARNACION · COMMUNITY PREDICTION COMPETITION · PRIVATE · 21 DAY

Airbnb NYC Price Prediction

NDL Bootcamp 2025

Overview

NYC Airbnb Price Prediction Challenge

Competition Overview

Welcome to the NYC Airbnb Price Prediction Challenge! This competition focuses on p
Airbnb listings across New York City. Your goal is to build robust machine learning moc
rental prices using property characteristics, location data, host information, and bookir

Start
2 days ago

Description

Your Task

Predict the `price` (nightly rental rate in USD) for Airbnb listings in the test set. You'll r

1. **Explore and Clean the Data**: Handle missing values, outliers, and data inconsistencie
2. **Engineer Meaningful Features**: Extract insights from location, host behavior, and pr
3. **Build Predictive Models**: Develop machine learning algorithms that generalize well

Submit to Competition

Airbnb NYC Price Prediction
You have 2 submissions remaining today. This resets in 7 hours.

Drag and drop file to upload
(e.g., .csv, .zip, .gz, .7z)

or

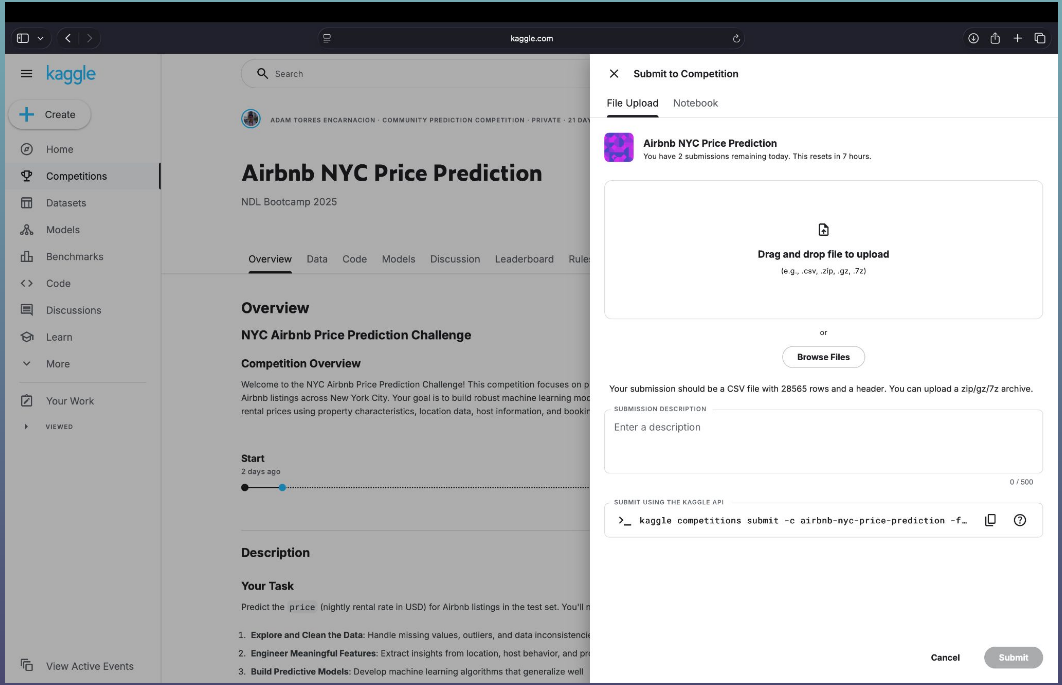
Your submission should be a CSV file with 28565 rows and a header. You can upload a zip/gz/7z archive.

SUBMISSION DESCRIPTION
Enter a description
0 / 500

SUBMIT USING THE KAGGLE API

```
>_ kaggle competitions submit -c airbnb-nyc-price-prediction -f…
```

Kaggle Submissions



ADAM TORRES ENCARNACION · COMMUNITY PREDICTION COMPETITION · PRIVATE · 21 DAY

Airbnb NYC Price Prediction

NDL Bootcamp 2025

Overview

NYC Airbnb Price Prediction Challenge

Competition Overview

Welcome to the NYC Airbnb Price Prediction Challenge! This competition focuses on p
Airbnb listings across New York City. Your goal is to build robust machine learning mod
rental prices using property characteristics, location data, host information, and bookir

Start
2 days ago

Description

Your Task

Predict the `price` (nightly rental rate in USD) for Airbnb listings in the test set. You'll r

1. **Explore and Clean the Data**: Handle missing values, outliers, and data inconsistencie
2. **Engineer Meaningful Features**: Extract insights from location, host behavior, and pr
3. **Build Predictive Models**: Develop machine learning algorithms that generalize well

Submit to Competition

Airbnb NYC Price Prediction
You have 2 submissions remaining today. This resets in 7 hours.

Drag and drop file to upload
(e.g., .csv, .zip, .gz, .7z)

or

Your submission should be a CSV file with 28565 rows and a header. You can upload a zip/gz/7z archive.

SUBMISSION DESCRIPTION
Enter a description
0 / 500

SUBMIT USING THE KAGGLE API

```
>_ kaggle competitions submit -c airbnb-nyc-price-prediction -f…
```

Epoch: Your Research Reading Club

Passionate about delving into the latest research? Epoch is a community where curious minds come together to explore, discuss, and learn.

What We Offer:

- In-depth modern paper discussions.
- Opportunities to expand your knowledge.
- A chance to present and prove yourself to your community!

Epoch provides a supportive environment for intellectual growth.

Join Our Community!

Scan the QR code to learn more and become a member.

Next Meeting

Will be in Westgate e203

Recap Lab

Help session