Machine learning, Deep learning, statistical inference

Simple setting

$$\begin{bmatrix} 0 & 7 & 21 & \dots \\ 1 & 3 & 22 & \dots \\ 0 & 2 & 26 & \dots \\ \vdots & \vdots & \vdots & \ddots \\ 1 & 7 & 29 & \dots \end{bmatrix}$$

i/p matrix

$X \rightarrow f \rightarrow y$

y: 0, 1, ... o/p

$\hat{h} \approx mx + \text{error}$

$\hat{h} \approx mx + mx^2 + \text{error}$

$f(X) = y$ → ML problem / DL problem too

Data

Statistical Inference
- ~ less data
- ~ Assumption

- ~ large datasets
- ~ not so many assumptions

sample (small dataset) → $\hat{f}$ → inference → f

Data

Problem framing → 1
- ML Problem framing
- Data Collection & Integration

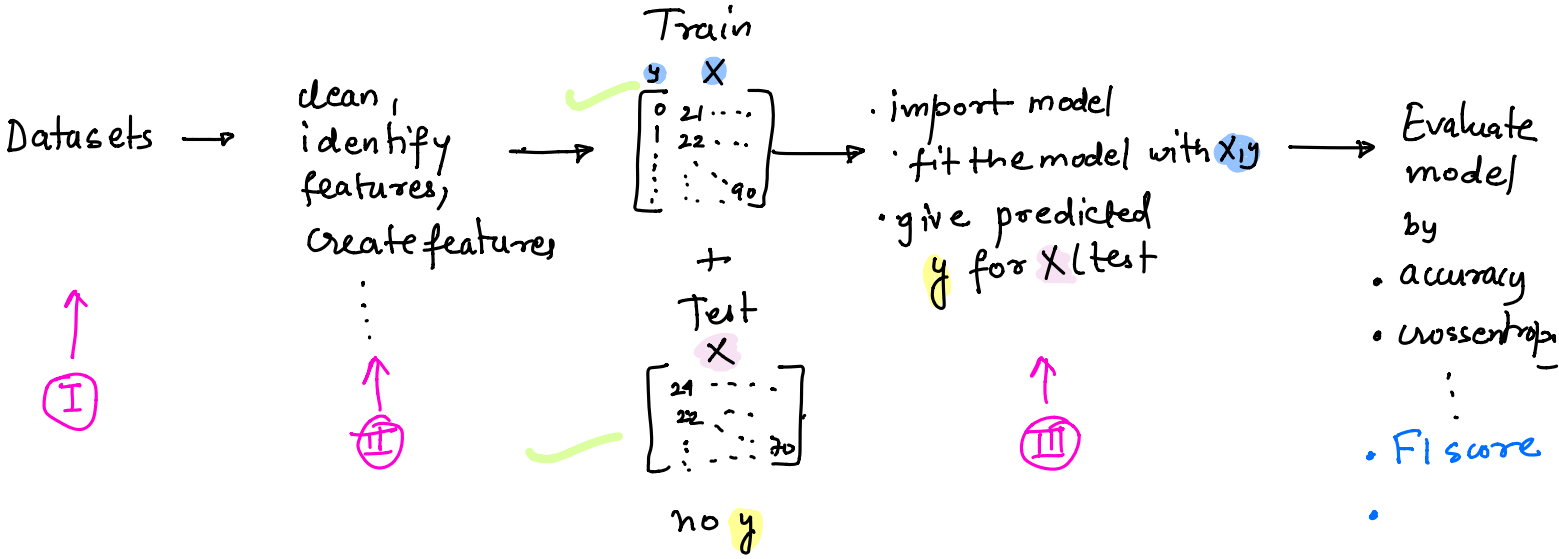
i/p Prep

o/p preparation → 2
- Data Preparation
- Data Visualization & Analytics
- Feature selection & Engineering

refine model

how good model is ? → 3
- Model Training
- Model Evaluation
- Prediction

Best model



How to improve model accuracy or evaluation measures values?

(I) Get more training

(II) Get new features

(III) Tune hyperparameters, change models

Here what we really care is prediction, Any model is fine, I want better prediction for my problem statement ⟹ ML/DL problem.

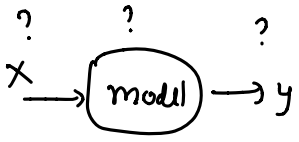
train with ik + look t

[Identify pattern in data & adapting with them]

Scientific method: google developer site

Step	Example
1. Set the research goal.	I want to predict how heavy traffic will be on a given day.
2. Make a hypothesis. $f \approx mX+c$	I think the weather forecast is an informative signal.
3. Collect the data.	Collect historical traffic data and weather on each day.
4. Test your hypothesis.	Train a model using this data.
5. Analyze your results.	Is this model better than existing systems?
6. Reach a conclusion.	I should (not) use this model to make predictions, because of X, Y, and Z.
7. Refine hypothesis and repeat.	Time of year could be a helpful signal.

1. what is X
2. what is y
3. How to evaluate themodel



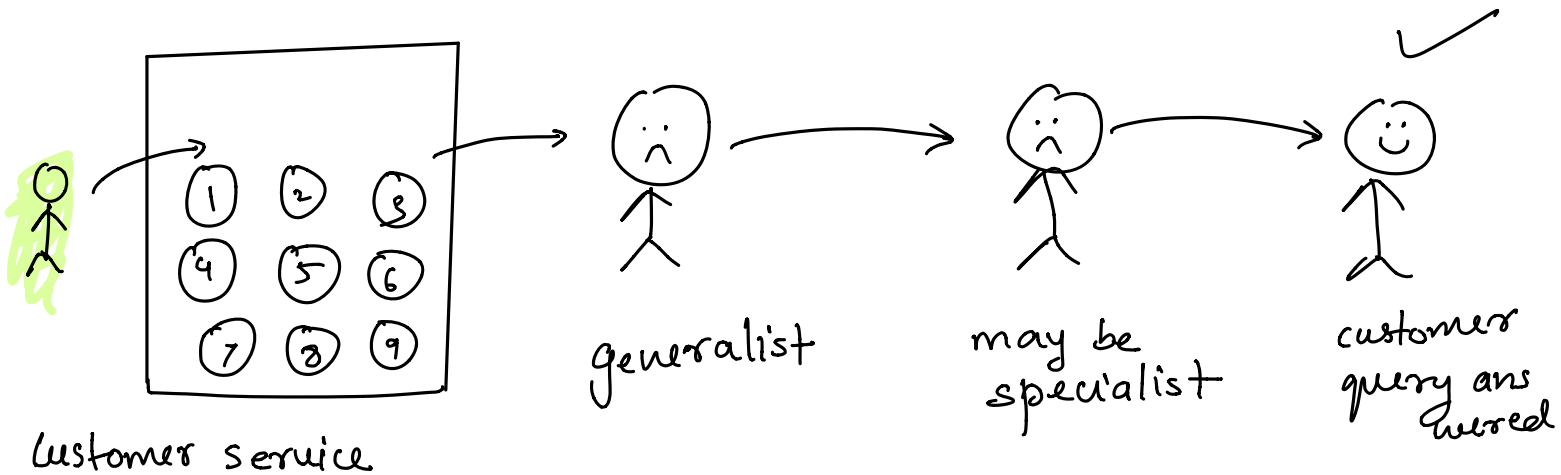
Toolsets

1. libraries : numpy, pandas, Matplotlib, Seaborn, Scikit, Tensorflow
   - multidimen
   - Datasets
   - plotting
   - advance plotting
   - ML
   - ML/DL
   - Keras → DL
2. Notebooks : Jupyter, Colab, AWS notebook
   - Jupyter: independ local
   - Colab → google
   - AWS notebook ← Sagemaker AWS ML service
3. Cloud services : Google ML, AWS Sagemaker ←
4. Big data : Pyspark *.

good for python → Pycharm — IDE

can write anycode → VSC — IDE

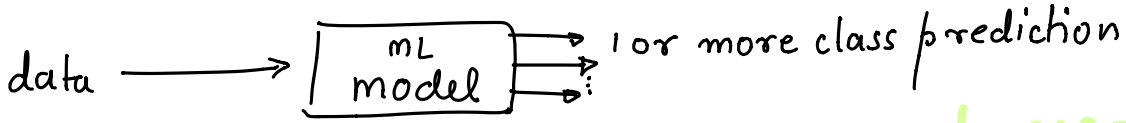
Example : Reference : Introduction to AWS coursera week 2

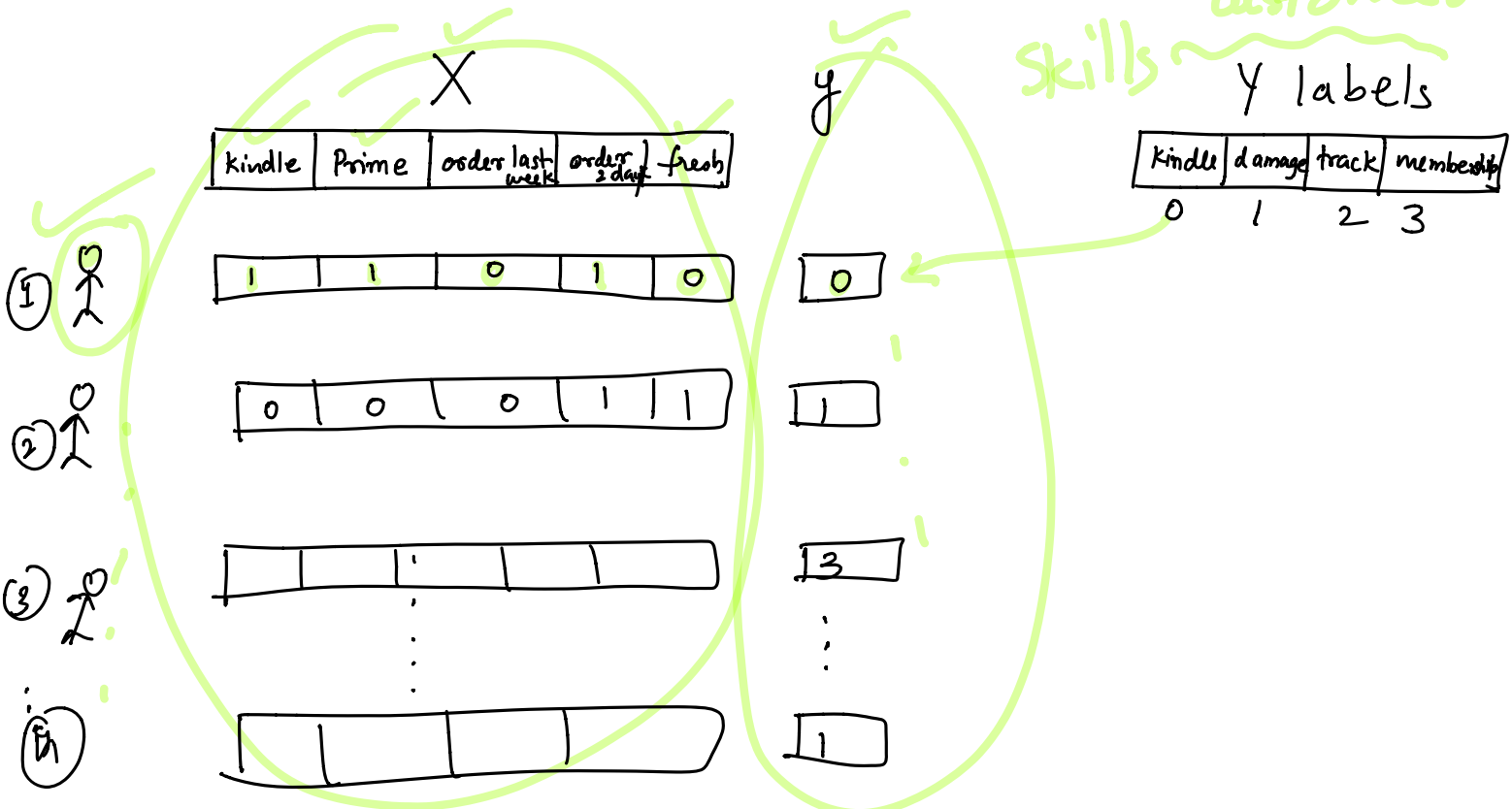


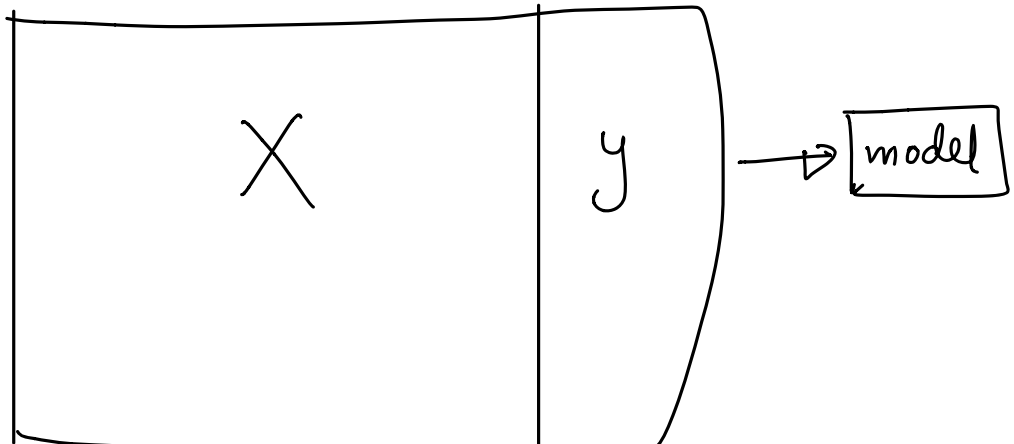
inefficient, costly, customer unhappy

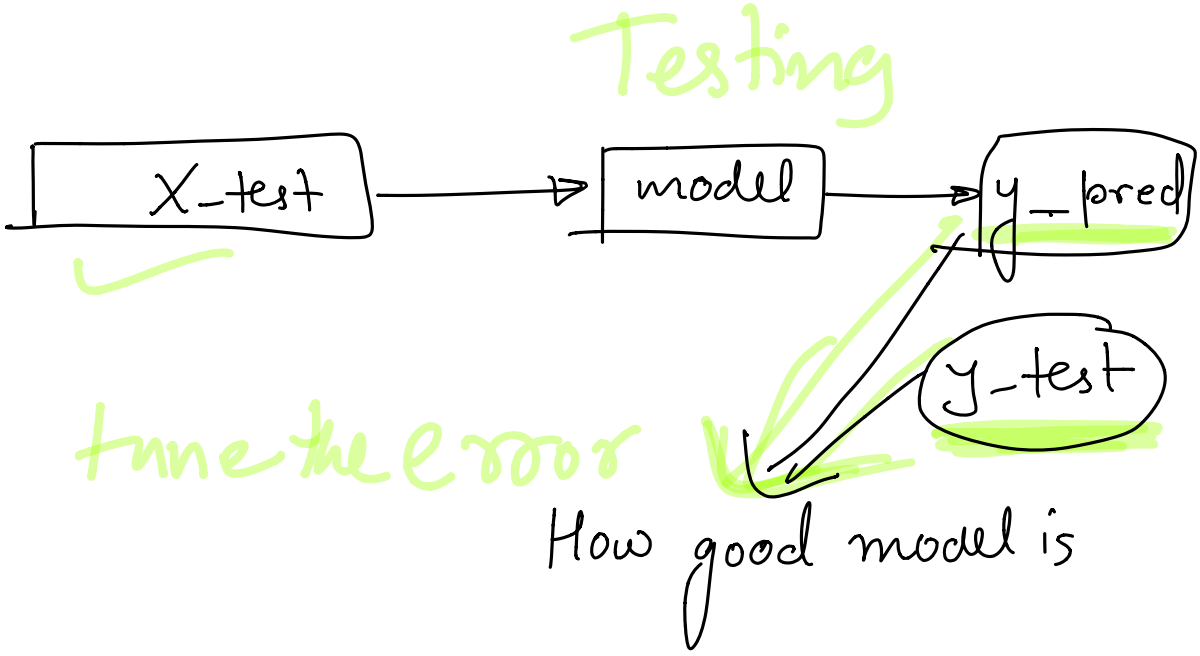
① Business problem to machine learning.

How to route → ?

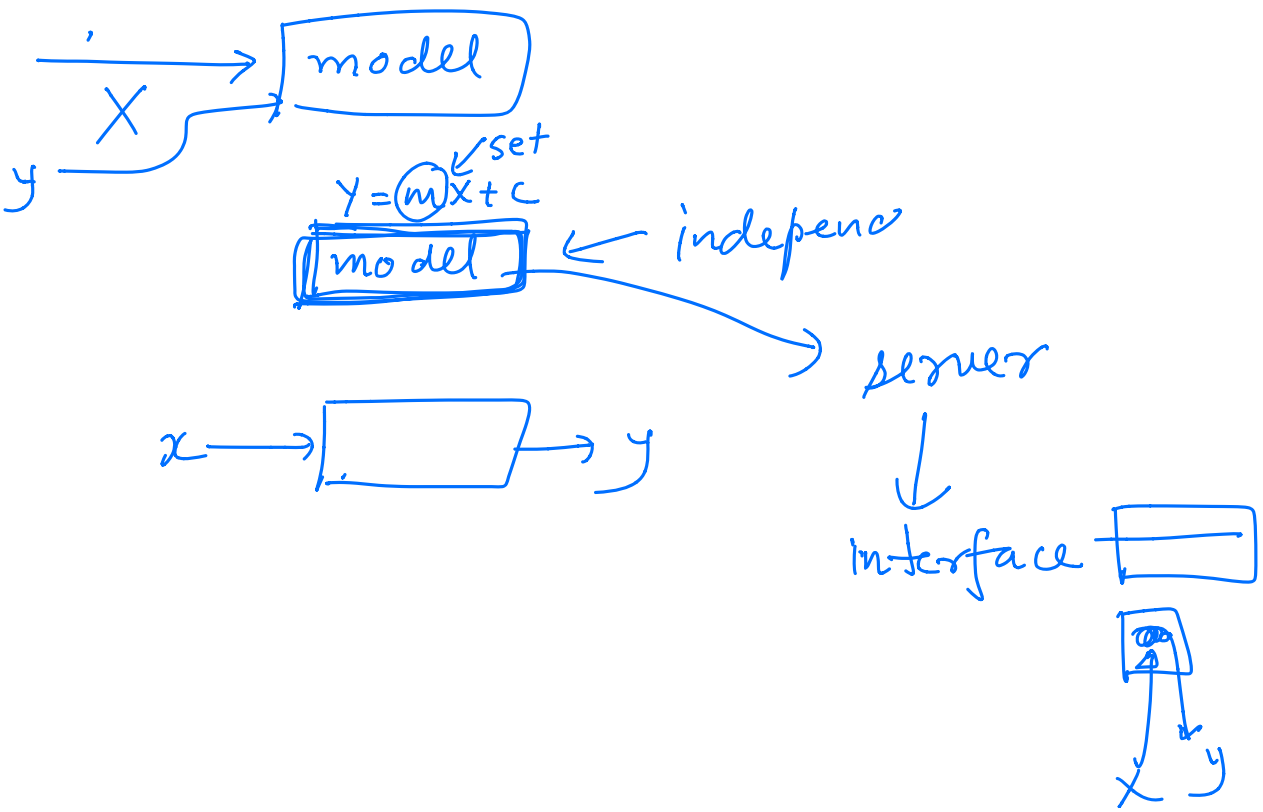







Testing
X_test
model
y_pred
y_test
tune the error
How good model is

final
model
deploy


model
X
y
set
Y = mX + C
model
indepen
server
x
y
interface
X
y