AI 2024 Online Summer Internship

Titanic Project

Lecture 1

Treating Titanic Passenger Survival Prediction Problem as a Machine Learning Problem using Train-Test Split Approach

Lecture Outline

- **Titanic Passenger Survival Prediction Problem**
- **Steps – Treating Titanic Passenger Survival Prediction Problem as a Machine Learning Problem using Train-Test Split Approach**

Titanic Passenger Survival Prediction Problem

SLIDE

Royal Mail Ship (RMS) Titanic – Brief Overview

- **RMS Titanic was a British passenger liner operated by the White Star Line that sank in the North Atlantic Ocean in the early morning hours of 15 April 1912, after striking an iceberg during her maiden voyage from Southampton to New York City**
- **Of the estimated 2,224 passengers and crew aboard, more than 1,500 died, making the sinking one of modern history's deadliest peacetime commercial marine disasters**
- **RMS Titanic was the largest ship afloat at the time she entered service[1]**

SLIDE

RMS Titanic – Main Features

- **Name of Ship**
  - **Royal Mail Ship (RMS) Titanic**
- **Manufactured**
  - **May 31, 1911**
- **Size**
  - **Length = 882 feet 9 inches (269.06 m)**
  - **Width = 92.5 feet (28.2 m)**
  - **Height = 104 feet (32 m)**
- **First Journey**
  - **April 10, 1912**
- **Source and Destination**
  - **Southampton, England to New York, USA**
- **Total Passengers On Board**
  - **2,208**
- **Passengers Survived**
  - **705**
- **Passengers Died**
  - **1,503**

[1] Last visited: 07-09-2020

SLIDE
RMS Titanic

Figure 01: Royal Mail Ship (RMS) Titanic [ ]

SLIDE
Lecture Focus

- **The main focus of this Lecture is developing a**
  - **Predictive System which can automatically predict whether a Passenger in RMS Titanic Survived or Not**

SLIDE
Titanic Passenger Survival Prediction System

- **Real-world World**
  - **Titanic Passenger Survival**
- **Treated as**
  - **Supervised Machine Learning Problem**
- **Note**
  - **Titanic Passenger Survival Prediction Problem is treated as a**

        - Binary Classification Problem because the
            - The main aim is to distinguish between Two Classes
                - Class 01 = Survived
                - Class 02 = Not Survived
- Goal
    - Learn an Input-Output Function
        - i.e. Learn from Input to predict the Output

SLIDE
Titanic Passenger Survival Prediction System – Task

- Given
    - A Passenger (Represented as Set of Attributes)
- Task
    - Automatically predict whether the Passenger Survived or Not

SLIDE
Titanic Passenger Survival Prediction System – Input and Output

- Input
    - A Passenger
- Output
    - Survived / Not Survived

SLIDE
Note

- In Kaggle Titanic Dataset, a Passenger is represented with many Attributes
- Kaggle Titanic Dataset
    - URL:
- For simplicity and to explain things more clearly
    - In this Lecture, we have represented a Passenger with Four Attributes

SLIDE
Titanic Passenger Survival Prediction System – Input Attributes

- In this Lecture, a Passenger is represented with the following Four Attributes
- Attribute 01 – PClass

- Possible Value 01 = First
- Possible Value 02 = Second
- Possible Value 03 = Third

- Attribute 02 – Gender
  - Possible Value 01 = Female
  - Possible Value 02 = Male
- Attribute 03 – Sibling
  - Possible Value 01 = Zero
  - Possible Value 02 = One
  - Possible Value 03 = Two
  - Possible Value 04 = Three
- Attribute 04 – Embarked
  - Possible Value 01 = Cherbourg
  - Possible Value 02 = Southampton
  - Possible Value 03 = Queenstown

SLIDE

Titanic Passenger Survival Prediction System – Output Attributes

- In Titanic Dataset, there is One Output Attribute
  - Attribute 01 – Survived
    - Possible Value 01 = Yes
    - Possible Value 02 = No

SLIDE

Titanic Passenger Survival Prediction System – Summary (Input and Output)

- The following Table summarizes the Input and Output Attributes for Titanic Dataset

Attribute No.	Attribute Names	Possible Values	Data Types
1	PClass	First, Second, Third	Categorical
2	Gender	Male, Female	Categorical

3	Sibling	Zero, One, Two, Three	Categorical
4	Embarked	Southampton, Queenstown, Cherbourg	Categorical
5	Survived	Yes, No	Categorical

Table 01: Attributes of Dataset

Steps – Treating Titanic Passenger Survival Prediction Problem as a Supervised Machine Learning Problem using Train-Test Split Approach

SLIDE
Titanic Passenger Survival Prediction Problem

- **Task**
  - **Develop a Titanic Passenger Survival Prediction System to Predict the Survival of a Passenger**
- **Input**
  - **Four Attributes**

1. **PClass**
2. **Gender**
3. **Sibling**
4. **Embarked**

- **Output**
  - **One Attribute**

1. **Survived**

- **Treated as a**
  - **Supervised Machine Learning Problem**
- **Goal**
  - **Learn an Input-Output Function**
    - **i.e. Learn from Input to predict the Output**

SLIDE
Titanic Passenger Survival Prediction System is a Classification Problem

- **Titanic Passenger Survival Prediction System is a Classification Problem because**
  - **Output is Categorical**

SLIDE

Titanic Passenger Survival Prediction Problem – Input and Output

- **Input**
  - **Categorical**
- **Output**
  - **Categorical**

SLIDE

Project Focus

Titanic Passenger Survival Prediction System

SLIDE

Steps – Treating Titanic Passenger Survival Prediction Problem as a Classification Problem

- **In Sha Allah (انشاء الله), I will follow the following steps to treat the Titanic Passenger Survival Prediction Problem as a Classification Problem**
  - **Step 1: Decide the Learning Settings**
  - **Step 2: Obtain Sample Data**
  - **Step 3: Understand and Pre-process Sample Data**
  - **Step 4: Represent Sample Data in Machine Understandable Format**
  - **Step 5: Select Suitable Machine Learning Algorithms**
  - **Step 6: Split Sample Data into Training Data and Testing Data**
  - **Step 7: Select Suitable Evaluation Measure(s)**
  - **Step 8: Execute First Two Phases of Machine Learning Cycle**
    - **Training Phase**
    - **Testing Phase**
  - **Step 9: Analyze Results**

If (Results are Good)
Then
Move to the Next Step
Else
Go to Step 1

- Step 10: Execute 3rd and 4th Phases of Machine Learning Cycle
  - Application Phase
  - Feedback Phase
- Step 11: Based on Feedback
  - Go to Step 1 and Repeat all the Steps

Step 1: Decide the Learning Setting

SLIDE
Step 1: Decide the Learning Setting

- In Sha Allah (انشاء الله), I will treat the Titanic Passenger Survival Prediction Problem as a
  - Supervised Machine Learning Problem
- Since Output is Categorical, it will be treated as a
  - Classification Problem

Step 2: Obtain Sample Data

SLIDE
Step 2: Obtain Sample Data

- Since I am Treating Titanic Passenger Survival Prediction Problem as a Supervised Machine Learning Problem, I will need
  - Annotated Data
- For more accurate learning, I need
  1. Large amount of Annotated Data
  2. High-quality Annotated Data
  3. Balanced Data
- Note
  - For simplicity, In Sha Allah (انشاء الله) I will use a toy Corpus / Dataset of 100 instances

SLIDE
Step 2: Obtain Sample Data Cont…

- Two Main Choices to Obtain Data
    1. Use an Existing Corpus
    2. Develop Your Corpus
- The Dataset to use is a subset of Kaggle Titanic Dataset
    - Corpus / Dataset
        - Dataset Link:
    - Paper Link:
        -
    - Paper Reference:
        - Cicoria, S., Sherlock, J., Muniswamaiah, M., and Clarke, L., 2014, Classification of titanic passenger data and chances of surviving the disaster, In Proceedings of Student-Faculty Research Day, CSIS (pp. 1-6)

SLIDE
Obtain Sample Data Cont…

- Total Instances in Sample Data = 100
    - Survived = 50
    - Not Survived = 50

SLIDE
Sample Data

- We obtained a Sample Data of 100 instances
    - See sample-data.csv File in Supporting Material
- The following Table shows the Sample Data

Instance No.	Input				Output
	PClass	Gender	Sibling	Embarked	Survived
x_1	Third	Male	One	Southampton	No
x_2	Second	Female	Zero	Southampton	Yes
x_3	Third	Male	Zero	Southampton	No
x_4	Third	Female	Three	Southampton	Yes
x_5	Third	Male	Zero	Queenstown	No
x_6	First	Female	Three	Southampton	Yes
x_7	Third	Male	Zero	Southampton	No

x_8	Third	Male	Zero	Southampton	Yes
x_9	First	Male	Zero	Southampton	No
x_{10}	Second	Male	Zero	Southampton	Yes
x_{11}	Third	Male	One	Queenstown	No
x_{12}	First	Male	Zero	Cherbourg	Yes
x_{13}	First	Male	Zero	Cherbourg	No
x_{14}	Second	Female	Zero	Southampton	Yes
x_{15}	Second	Male	Zero	Southampton	No
x_{16}	Third	Male	One	Cherbourg	Yes
x_{17}	Third	Male	Two	Cherbourg	No
x_{18}	Third	Male	Zero	Southampton	Yes
x_{19}	Third	Male	Zero	Southampton	No
x_{20}	Third	Female	One	Cherbourg	Yes
x_{21}	Third	Male	Zero	Cherbourg	No
x_{22}	First	Female	Zero	Southampton	Yes
x_{23}	First	Male	Zero	Cherbourg	No
x_{24}	Second	Female	Zero	Southampton	Yes
x_{25}	Third	Female	One	Southampton	No
x_{26}	Third	Female	Zero	Southampton	Yes
x_{27}	Third	Male	Zero	Southampton	No
x_{28}	Third	Male	Zero	Southampton	Yes
x_{29}	Third	Male	Zero	Southampton	No
x_{30}	Third	Female	One	Queenstown	Yes
x_{31}	Third	Female	Two	Southampton	No
x_{32}	First	Female	Zero	Cherbourg	Yes
x_{33}	Third	Male	Two	Southampton	No
x_{34}	First	Female	Zero	Southampton	Yes
x_{35}	First	Male	One	Cherbourg	No
x_{36}	First	Female	Zero	Southampton	Yes
x_{37}	First	Male	One	Southampton	No
x_{38}	Third	Female	One	Queenstown	Yes
x_{39}	Third	Female	One	Southampton	No
x_{40}	Second	Female	One	Southampton	Yes
x_{41}	Second	Female	One	Southampton	No
x_{42}	Third	Female	Zero	Queenstown	Yes

x_{43}	Third	Male	Zero	Cherbourg	No
x_{44}	First	Female	Zero	Cherbourg	Yes
x_{45}	Third	Male	Three	Southampton	No
x_{46}	Third	Female	One	Southampton	Yes
x_{47}	Third	Male	Zero	Southampton	No
x_{48}	Third	Male	Zero	Southampton	Yes
x_{49}	Third	Male	One	Southampton	No
x_{50}	Second	Female	Zero	Southampton	Yes
x_{51}	First	Male	Three	Southampton	No
x_{52}	Third	Male	Zero	Southampton	Yes
x_{53}	Third	Female	One	Southampton	No
x_{54}	Second	Female	Zero	Southampton	Yes
x_{55}	Third	Female	Three	Southampton	No
x_{56}	Third	Female	Zero	Southampton	Yes
x_{57}	Third	Female	Zero	Southampton	No
x_{58}	Third	Female	One	Southampton	Yes
x_{59}	Third	Male	Zero	Cherbourg	No
x_{60}	Third	Male	Zero	Southampton	Yes
x_{61}	Second	Male	Zero	Southampton	No
x_{62}	First	Female	One	Southampton	Yes
x_{63}	Third	Male	Three	Queenstown	No
x_{64}	Third	Female	Zero	Queenstown	Yes
x_{65}	Second	Female	Zero	Southampton	No
x_{66}	Third	Female	Zero	Cherbourg	Yes
x_{67}	Third	Male	Zero	Southampton	No
x_{68}	Second	Male	Zero	Southampton	Yes
x_{69}	Third	Male	Zero	Southampton	No
x_{70}	First	Male	Zero	Southampton	Yes
x_{71}	Third	Male	Zero	Southampton	No
x_{72}	Second	Male	Two	Southampton	Yes
x_{73}	Third	Male	Zero	Cherbourg	No
x_{74}	Third	Female	Zero	Southampton	Yes
x_{75}	Third	Female	Zero	Southampton	No
x_{76}	First	Female	One	Southampton	Yes
x_{77}	Third	Male	One	Southampton	No

x_{78}	Second	Female	Zero	Southampton	Yes
x_{79}	Third	Male	Zero	Southampton	No
x_{80}	Third	Female	Two	Southampton	Yes
x_{81}	Third	Male	Two	Southampton	No
x_{82}	Third	Female	Zero	Southampton	Yes
x_{83}	Third	Male	Zero	Southampton	No
x_{84}	First	Male	One	Southampton	Yes
x_{85}	Second	Male	Zero	Southampton	No
x_{86}	Third	Female	One	Queenstown	Yes
x_{87}	Second	Male	Zero	Southampton	No
x_{88}	First	Female	One	Southampton	Yes
x_{89}	Third	Male	Zero	Southampton	No
x_{90}	Second	Female	Zero	Southampton	Yes
x_{91}	Second	Male	Zero	Southampton	No
x_{92}	Second	Female	Zero	Southampton	Yes
x_{93}	Third	Male	Zero	Cherbourg	No
x_{94}	First	Male	One	Southampton	Yes
x_{95}	First	Male	One	Southampton	No
x_{96}	Third	Male	One	Cherbourg	Yes
x_{97}	Third	Male	One	Queenstown	No
x_{98}	Third	Male	One	Southampton	Yes
x_{99}	Second	Male	One	Southampton	No
x_{100}	Third	Male	One	Southampton	Yes

Step 03: Understand and Pre-process Data

SLIDE

Step 3: Understand and Pre-process Sample Data

- **Understanding Data**
  - **The Sample Data contains Five Attributes**
    - **PClass**
    - **Gender**
    - **Sibling**
    - **Embarked**

        - Survived
    - Separating Input from Output
        - Input comprises of Four Attributes
            - PClass
            - Gender
            - Sibling
            - Embarked
        - The Output comprises of a Single Attribute
            - Survived
- Pre-processing Data
    - Corpus is already pre-processed
        - Therefore, no pre-processing is needed 😊

Step 04: Represent Data in Machine Understandable Format

SLIDE
Step 4: Represent Sample Data in Machine Understandable Format

- Feature-based Classification Algorithms (implemented in Scikit-learn) can understand data in
    - Attribute-Value Pair
        - Values of Attributes / Features must be Numeric
- Problem
    - Our Sample Data is not in Attribute-Value Pair form
        - We need to transform our Sample Data into Machine Understandable Format
- Solution
    - There are many approaches to transform Sample Data into Machine Understandable Format

SLIDE
Feature Extraction

- Features are already extracted
    - Therefore, we will skip the Feature Extraction Step 😊

SLIDE

Important Note

- **In this Lecture, we are using Scikit-learn implementation of the Support Vector Classifier Machine Learning Algorithm**
- **Scikit-learn can only understand Data in Numerical Representation**
  - **Therefore, we will need to Convert the Categorical Values to Numerical Values**

SLIDE

Transforming Sample Data in Machine Understandable Format

- **In our Sample Data**
  - **Input is Categorical**
  - **Output is Categorical**
- **Considering Input (PClass, Gender, Sibling, Embarked) and Output (Survived), we will need to**
  - **Transform Input (Categorical) into Numerical Representation**
  - **Transform Output (Categorical) into Numerical Representation**

SLIDE

Converting Output into Numerical Representation

- **A Two-Step Process**
  - **Step 01: Define an Encoding Scheme**
  - **Step 02: Use Encoding Scheme defined in Step 01, to convert Categorical Output Values to Numerical Output Values for all instances in the Sample Data**

SLIDE

Converting Output into Numerical Representation Cont…

- **Step 01: Define an Encoding Scheme**
- **Encoding Scheme for Survived Attribute**
  - **No = 0**
  - **Yes = 1**

SLIDE

Converting Output into Numerical Representation Cont…

- **Step 02: Use Encoding Scheme defined in Step 01, to convert Categorical Output Values to Numerical Output Values for all instances in the Sample Data**
- **The Table below shows Sample Data after Encoding Categorical Output Values to Numerical Output Values**
  - **See sample-data-encoded-output.csv File in Supporting Material**

Instance No.	Input				Output
	PClass	Gender	Sibling	Embarked	Survived
x_1	Third	Male	One	Southampton	0
x_2	Second	Female	Zero	Southampton	1
x_3	Third	Male	Zero	Southampton	0
x_4	Third	Female	Three	Southampton	1
x_5	Third	Male	Zero	Queenstown	0
x_6	First	Female	Three	Southampton	1
x_7	Third	Male	Zero	Southampton	0
x_8	Third	Male	Zero	Southampton	1
x_9	First	Male	Zero	Southampton	0
x_{10}	Second	Male	Zero	Southampton	1
x_{11}	Third	Male	One	Queenstown	0
x_{12}	First	Male	Zero	Cherbourg	1
x_{13}	First	Male	Zero	Cherbourg	0
x_{14}	Second	Female	Zero	Southampton	1
x_{15}	Second	Male	Zero	Southampton	0
x_{16}	Third	Male	One	Cherbourg	1
x_{17}	Third	Male	Two	Cherbourg	0
x_{18}	Third	Male	Zero	Southampton	1
x_{19}	Third	Male	Zero	Southampton	0
x_{20}	Third	Female	One	Cherbourg	1
x_{21}	Third	Male	Zero	Cherbourg	0
x_{22}	First	Female	Zero	Southampton	1
x_{23}	First	Male	Zero	Cherbourg	0
x_{24}	Second	Female	Zero	Southampton	1

x_{25}	Third	Female	One	Southampton	0
x_{26}	Third	Female	Zero	Southampton	1
x_{27}	Third	Male	Zero	Southampton	0
x_{28}	Third	Male	Zero	Southampton	1
x_{29}	Third	Male	Zero	Southampton	0
x_{30}	Third	Female	One	Queenstown	1
x_{31}	Third	Female	Two	Southampton	0
x_{32}	First	Female	Zero	Cherbourg	1
x_{33}	Third	Male	Two	Southampton	0
x_{34}	First	Female	Zero	Southampton	1
x_{35}	First	Male	One	Cherbourg	0
x_{36}	First	Female	Zero	Southampton	1
x_{37}	First	Male	One	Southampton	0
x_{38}	Third	Female	One	Queenstown	1
x_{39}	Third	Female	One	Southampton	0
x_{40}	Second	Female	One	Southampton	1
x_{41}	Second	Female	One	Southampton	0
x_{42}	Third	Female	Zero	Queenstown	1
x_{43}	Third	Male	Zero	Cherbourg	0
x_{44}	First	Female	Zero	Cherbourg	1
x_{45}	Third	Male	Three	Southampton	0
x_{46}	Third	Female	One	Southampton	1
x_{47}	Third	Male	Zero	Southampton	0
x_{48}	Third	Male	Zero	Southampton	1
x_{49}	Third	Male	One	Southampton	0
x_{50}	Second	Female	Zero	Southampton	1
x_{51}	First	Male	Three	Southampton	0
x_{52}	Third	Male	Zero	Southampton	1
x_{53}	Third	Female	One	Southampton	0
x_{54}	Second	Female	Zero	Southampton	1
x_{55}	Third	Female	Three	Southampton	0
x_{56}	Third	Female	Zero	Southampton	1
x_{57}	Third	Female	Zero	Southampton	0
x_{58}	Third	Female	One	Southampton	1
x_{59}	Third	Male	Zero	Cherbourg	0

x_{60}	Third	Male	Zero	Southampton	1
x_{61}	Second	Male	Zero	Southampton	0
x_{62}	First	Female	One	Southampton	1
x_{63}	Third	Male	Three	Queenstown	0
x_{64}	Third	Female	Zero	Queenstown	1
x_{65}	Second	Female	Zero	Southampton	0
x_{66}	Third	Female	Zero	Cherbourg	1
x_{67}	Third	Male	Zero	Southampton	0
x_{68}	Second	Male	Zero	Southampton	1
x_{69}	Third	Male	Zero	Southampton	0
x_{70}	First	Male	Zero	Southampton	1
x_{71}	Third	Male	Zero	Southampton	0
x_{72}	Second	Male	Two	Southampton	1
x_{73}	Third	Male	Zero	Cherbourg	0
x_{74}	Third	Female	Zero	Southampton	1
x_{75}	Third	Female	Zero	Southampton	0
x_{76}	First	Female	One	Southampton	1
x_{77}	Third	Male	One	Southampton	0
x_{78}	Second	Female	Zero	Southampton	1
x_{79}	Third	Male	Zero	Southampton	0
x_{80}	Third	Female	Two	Southampton	1
x_{81}	Third	Male	Two	Southampton	0
x_{82}	Third	Female	Zero	Southampton	1
x_{83}	Third	Male	Zero	Southampton	0
x_{84}	First	Male	One	Southampton	1
x_{85}	Second	Male	Zero	Southampton	0
x_{86}	Third	Female	One	Queenstown	1
x_{87}	Second	Male	Zero	Southampton	0
x_{88}	First	Female	One	Southampton	1
x_{89}	Third	Male	Zero	Southampton	0
x_{90}	Second	Female	Zero	Southampton	1
x_{91}	Second	Male	Zero	Southampton	0
x_{92}	Second	Female	Zero	Southampton	1
x_{93}	Third	Male	Zero	Cherbourg	0
x_{94}	First	Male	One	Southampton	1

x_{95}	First	Male	One	Southampton	0
x_{96}	Third	Male	One	Cherbourg	1
x_{97}	Third	Male	One	Queenstown	0
x_{98}	Third	Male	One	Southampton	1
x_{99}	Second	Male	One	Southampton	0
x_{100}	Third	Male	One	Southampton	1

SLIDE
Note

- **Alhamdulillah (الحمدللہ), Output is transformed into Numerical Representation**
- **In Sha Allah (انشاءاللہ), in the next Slides, I will try to explain how to transform Input into Numerical Representation**

SLIDE
Converting Input into Numerical Representation

- **Step 01: Define an Encoding Scheme**
- **Step 02: Use Encoding Scheme defined in Step 01, to convert Categorical Input Values to Numerical Input Values for all instances in the Sample Data**

SLIDE
Converting Input into Numerical Representation Cont…

- **Step 01: Define an Encoding Scheme**
- **Encoding Scheme for PClass Attribute**
  - **First = 0**
  - **Second = 1**
  - **Third = 2**
- **Encoding Scheme for Gender Attribute**
  - **Female = 0**
  - **Male = 1**
- **Encoding Scheme for Sibling Attribute**
  - **One = 0**
  - **Three = 1**
  - **Two = 2**
  - **Zero = 3**

- **Encoding Scheme for Embarked Attribute**
  - **Cherbourg = 0**
  - **Queenstown = 1**
  - **Southampton = 2**

SLIDE
Converting Input into Numerical Representation Cont…

- **Step 02: Use Encoding Scheme defined in Step 01, to convert Categorical Input Values to Numerical Input Values for all instances in the Sample Data**
- **The Table below shows Sample Data after Encoding Categorical Input Values to Numerical Input Values**
  - **See sample-data-encoded.csv File in Supporting Material**

Instance No.	Input				Output
	PClass	Gender	Sibling	Embarked	Survived
x_1	2	1	0	2	0
x_2	1	0	3	2	1
x_3	2	1	3	2	0
x_4	2	0	1	2	1
x_5	2	1	3	1	0
x_6	0	0	1	2	1
x_7	2	1	3	2	0
x_8	2	1	3	2	1
x_9	0	1	3	2	0
x_{10}	1	1	3	2	1
x_{11}	2	1	0	1	0
x_{12}	0	1	3	0	1
x_{13}	0	1	3	0	0
x_{14}	1	0	3	2	1
x_{15}	1	1	3	2	0
x_{16}	2	1	0	0	1
x_{17}	2	1	2	0	0
x_{18}	2	1	3	2	1

x_{19}	2	1	3	2	0
x_{20}	2	0	0	0	1
x_{21}	2	1	3	0	0
x_{22}	0	0	3	2	1
x_{23}	0	1	3	0	0
x_{24}	1	0	3	2	1
x_{25}	2	0	0	2	0
x_{26}	2	0	3	2	1
x_{27}	2	1	3	2	0
x_{28}	2	1	3	2	1
x_{29}	2	1	3	2	0
x_{30}	2	0	0	1	1
x_{31}	2	0	2	2	0
x_{32}	0	0	3	0	1
x_{33}	2	1	2	2	0
x_{34}	0	0	3	2	1
x_{35}	0	1	0	0	0
x_{36}	0	0	3	2	1
x_{37}	0	1	0	2	0
x_{38}	2	0	0	1	1
x_{39}	2	0	0	2	0
x_{40}	1	0	0	2	1
x_{41}	1	0	0	2	0
x_{42}	2	0	3	1	1
x_{43}	2	1	3	0	0
x_{44}	0	0	3	0	1
x_{45}	2	1	1	2	0
x_{46}	2	0	0	2	1
x_{47}	2	1	3	2	0
x_{48}	2	1	3	2	1
x_{49}	2	1	0	2	0
x_{50}	1	0	3	2	1
x_{51}	0	1	1	2	0
x_{52}	2	1	3	2	1
x_{53}	2	0	0	2	0

x_{54}	1	0	3	2	1
x_{55}	2	0	1	2	0
x_{56}	2	0	3	2	1
x_{57}	2	0	3	2	0
x_{58}	2	0	0	2	1
x_{59}	2	1	3	0	0
x_{60}	2	1	3	2	1
x_{61}	1	1	3	2	0
x_{62}	0	0	0	2	1
x_{63}	2	1	1	1	0
x_{64}	2	0	3	1	1
x_{65}	1	0	3	2	0
x_{66}	2	0	3	0	1
x_{67}	2	1	3	2	0
x_{68}	1	1	3	2	1
x_{69}	2	1	3	2	0
x_{70}	0	1	3	2	1
x_{71}	2	1	3	2	0
x_{72}	1	1	2	2	1
x_{73}	2	1	3	0	0
x_{74}	2	0	3	2	1
x_{75}	2	0	3	2	0
x_{76}	0	0	0	2	1
x_{77}	2	1	0	2	0
x_{78}	1	0	3	2	1
x_{79}	2	1	3	2	0
x_{80}	2	0	2	2	1
x_{81}	2	1	2	2	0
x_{82}	2	0	3	2	1
x_{83}	2	1	3	2	0
x_{84}	0	1	0	2	1
x_{85}	1	1	3	2	0
x_{86}	2	0	0	1	1
x_{87}	1	1	3	2	0
x_{88}	0	0	0	2	1

x_{89}	2	1	3	2	0
x_{90}	1	0	3	2	1
x_{91}	1	1	3	2	0
x_{92}	1	0	3	2	1
x_{93}	2	1	3	0	0
x_{94}	0	1	0	2	1
x_{95}	0	1	0	2	0
x_{96}	2	1	0	0	1
x_{97}	2	1	0	1	0
x_{98}	2	1	0	2	1
x_{99}	1	1	0	2	0
x_{100}	2	1	0	2	1

SLIDE

Hoooooooorrrrrrrrraaaaaaaaayyyyyyyyyyy! 🚩

- **Alhamdulillah (الحمدلله), both Input and Output are transformed into Numerical Representation**

SLIDE

Recap – Original Sample Data

Instance No.	Input				Output
	PClass	Gender	Sibling	Embarked	Survived
x_1	Third	Male	One	Southampton	No
x_2	Second	Female	Zero	Southampton	Yes
x_3	Third	Male	Zero	Southampton	No
x_4	Third	Female	Three	Southampton	Yes
x_5	Third	Male	Zero	Queenstown	No
x_6	First	Female	Three	Southampton	Yes
x_7	Third	Male	Zero	Southampton	No
x_8	Third	Male	Zero	Southampton	Yes
x_9	First	Male	Zero	Southampton	No
x_{10}	Second	Male	Zero	Southampton	Yes

x_{11}	Third	Male	One	Queenstown	No
x_{12}	First	Male	Zero	Cherbourg	Yes
x_{13}	First	Male	Zero	Cherbourg	No
x_{14}	Second	Female	Zero	Southampton	Yes
x_{15}	Second	Male	Zero	Southampton	No
x_{16}	Third	Male	One	Cherbourg	Yes
x_{17}	Third	Male	Two	Cherbourg	No
x_{18}	Third	Male	Zero	Southampton	Yes
x_{19}	Third	Male	Zero	Southampton	No
x_{20}	Third	Female	One	Cherbourg	Yes
x_{21}	Third	Male	Zero	Cherbourg	No
x_{22}	First	Female	Zero	Southampton	Yes
x_{23}	First	Male	Zero	Cherbourg	No
x_{24}	Second	Female	Zero	Southampton	Yes
x_{25}	Third	Female	One	Southampton	No
x_{26}	Third	Female	Zero	Southampton	Yes
x_{27}	Third	Male	Zero	Southampton	No
x_{28}	Third	Male	Zero	Southampton	Yes
x_{29}	Third	Male	Zero	Southampton	No
x_{30}	Third	Female	One	Queenstown	Yes
x_{31}	Third	Female	Two	Southampton	No
x_{32}	First	Female	Zero	Cherbourg	Yes
x_{33}	Third	Male	Two	Southampton	No
x_{34}	First	Female	Zero	Southampton	Yes
x_{35}	First	Male	One	Cherbourg	No
x_{36}	First	Female	Zero	Southampton	Yes
x_{37}	First	Male	One	Southampton	No
x_{38}	Third	Female	One	Queenstown	Yes
x_{39}	Third	Female	One	Southampton	No
x_{40}	Second	Female	One	Southampton	Yes
x_{41}	Second	Female	One	Southampton	No
x_{42}	Third	Female	Zero	Queenstown	Yes
x_{43}	Third	Male	Zero	Cherbourg	No
x_{44}	First	Female	Zero	Cherbourg	Yes
x_{45}	Third	Male	Three	Southampton	No

x_{46}	Third	Female	One	Southampton	Yes
x_{47}	Third	Male	Zero	Southampton	No
x_{48}	Third	Male	Zero	Southampton	Yes
x_{49}	Third	Male	One	Southampton	No
x_{50}	Second	Female	Zero	Southampton	Yes
x_{51}	First	Male	Three	Southampton	No
x_{52}	Third	Male	Zero	Southampton	Yes
x_{53}	Third	Female	One	Southampton	No
x_{54}	Second	Female	Zero	Southampton	Yes
x_{55}	Third	Female	Three	Southampton	No
x_{56}	Third	Female	Zero	Southampton	Yes
x_{57}	Third	Female	Zero	Southampton	No
x_{58}	Third	Female	One	Southampton	Yes
x_{59}	Third	Male	Zero	Cherbourg	No
x_{60}	Third	Male	Zero	Southampton	Yes
x_{61}	Second	Male	Zero	Southampton	No
x_{62}	First	Female	One	Southampton	Yes
x_{63}	Third	Male	Three	Queenstown	No
x_{64}	Third	Female	Zero	Queenstown	Yes
x_{65}	Second	Female	Zero	Southampton	No
x_{66}	Third	Female	Zero	Cherbourg	Yes
x_{67}	Third	Male	Zero	Southampton	No
x_{68}	Second	Male	Zero	Southampton	Yes
x_{69}	Third	Male	Zero	Southampton	No
x_{70}	First	Male	Zero	Southampton	Yes
x_{71}	Third	Male	Zero	Southampton	No
x_{72}	Second	Male	Two	Southampton	Yes
x_{73}	Third	Male	Zero	Cherbourg	No
x_{74}	Third	Female	Zero	Southampton	Yes
x_{75}	Third	Female	Zero	Southampton	No
x_{76}	First	Female	One	Southampton	Yes
x_{77}	Third	Male	One	Southampton	No
x_{78}	Second	Female	Zero	Southampton	Yes
x_{79}	Third	Male	Zero	Southampton	No
x_{80}	Third	Female	Two	Southampton	Yes

x_{81}	Third	Male	Two	Southampton	No
x_{82}	Third	Female	Zero	Southampton	Yes
x_{83}	Third	Male	Zero	Southampton	No
x_{84}	First	Male	One	Southampton	Yes
x_{85}	Second	Male	Zero	Southampton	No
x_{86}	Third	Female	One	Queenstown	Yes
x_{87}	Second	Male	Zero	Southampton	No
x_{88}	First	Female	One	Southampton	Yes
x_{89}	Third	Male	Zero	Southampton	No
x_{90}	Second	Female	Zero	Southampton	Yes
x_{91}	Second	Male	Zero	Southampton	No
x_{92}	Second	Female	Zero	Southampton	Yes
x_{93}	Third	Male	Zero	Cherbourg	No
x_{94}	First	Male	One	Southampton	Yes
x_{95}	First	Male	One	Southampton	No
x_{96}	Third	Male	One	Cherbourg	Yes
x_{97}	Third	Male	One	Queenstown	No
x_{98}	Third	Male	One	Southampton	Yes
x_{99}	Second	Male	One	Southampton	No
x_{100}	Third	Male	One	Southampton	Yes

SLIDE

Recap - Sample Data in Numerical Representation

Instance No.	Input				Output
	PClass	Gender	Sibling	Embarked	Survived
x_1	2	1	0	2	0
x_2	1	0	3	2	1
x_3	2	1	3	2	0
x_4	2	0	1	2	1
x_5	2	1	3	1	0
x_6	0	0	1	2	1
x_7	2	1	3	2	0

x_8	2	1	3	2	1
x_9	0	1	3	2	0
x_{10}	1	1	3	2	1
x_{11}	2	1	0	1	0
x_{12}	0	1	3	0	1
x_{13}	0	1	3	0	0
x_{14}	1	0	3	2	1
x_{15}	1	1	3	2	0
x_{16}	2	1	0	0	1
x_{17}	2	1	2	0	0
x_{18}	2	1	3	2	1
x_{19}	2	1	3	2	0
x_{20}	2	0	0	0	1
x_{21}	2	1	3	0	0
x_{22}	0	0	3	2	1
x_{23}	0	1	3	0	0
x_{24}	1	0	3	2	1
x_{25}	2	0	0	2	0
x_{26}	2	0	3	2	1
x_{27}	2	1	3	2	0
x_{28}	2	1	3	2	1
x_{29}	2	1	3	2	0
x_{30}	2	0	0	1	1
x_{31}	2	0	2	2	0
x_{32}	0	0	3	0	1
x_{33}	2	1	2	2	0
x_{34}	0	0	3	2	1
x_{35}	0	1	0	0	0
x_{36}	0	0	3	2	1
x_{37}	0	1	0	2	0
x_{38}	2	0	0	1	1
x_{39}	2	0	0	2	0
x_{40}	1	0	0	2	1
x_{41}	1	0	0	2	0
x_{42}	2	0	3	1	1

x_{43}	2	1	3	0	0
x_{44}	0	0	3	0	1
x_{45}	2	1	1	2	0
x_{46}	2	0	0	2	1
x_{47}	2	1	3	2	0
x_{48}	2	1	3	2	1
x_{49}	2	1	0	2	0
x_{50}	1	0	3	2	1
x_{51}	0	1	1	2	0
x_{52}	2	1	3	2	1
x_{53}	2	0	0	2	0
x_{54}	1	0	3	2	1
x_{55}	2	0	1	2	0
x_{56}	2	0	3	2	1
x_{57}	2	0	3	2	0
x_{58}	2	0	0	2	1
x_{59}	2	1	3	0	0
x_{60}	2	1	3	2	1
x_{61}	1	1	3	2	0
x_{62}	0	0	0	2	1
x_{63}	2	1	1	1	0
x_{64}	2	0	3	1	1
x_{65}	1	0	3	2	0
x_{66}	2	0	3	0	1
x_{67}	2	1	3	2	0
x_{68}	1	1	3	2	1
x_{69}	2	1	3	2	0
x_{70}	0	1	3	2	1
x_{71}	2	1	3	2	0
x_{72}	1	1	2	2	1
x_{73}	2	1	3	0	0
x_{74}	2	0	3	2	1
x_{75}	2	0	3	2	0
x_{76}	0	0	0	2	1
x_{77}	2	1	0	2	0

x_{78}	1	0	3	2	1
x_{79}	2	1	3	2	0
x_{80}	2	0	2	2	1
x_{81}	2	1	2	2	0
x_{82}	2	0	3	2	1
x_{83}	2	1	3	2	0
x_{84}	0	1	0	2	1
x_{85}	1	1	3	2	0
x_{86}	2	0	0	1	1
x_{87}	1	1	3	2	0
x_{88}	0	0	0	2	1
x_{89}	2	1	3	2	0
x_{90}	1	0	3	2	1
x_{91}	1	1	3	2	0
x_{92}	1	0	3	2	1
x_{93}	2	1	3	0	0
x_{94}	0	1	0	2	1
x_{95}	0	1	0	2	0
x_{96}	2	1	0	0	1
x_{97}	2	1	0	1	0
x_{98}	2	1	0	2	1
x_{99}	1	1	0	2	0
x_{100}	2	1	0	2	1

Step 05: Select Suitable Machine Learning Algorithms

SLIDE

Step 05: Select Suitable Machine Learning Algorithms

- **Previous students have shown that Good Starting Points for Classification Problems are**
  - **Random Forest Classifier**
  - **Support Vector Classifier**
  - **Naïve Bayes**
  - **Gradient Boosting Classifier**

SLIDE
Lecture Focus

- In Sha Allah, in this Lecture, we will use

Support Vector Classifier

Step 06: Split Sample Data into Training Data and Testing Data

---

SLIDE
Step 6: Split Sample Data into Training and Testing

- We Split the Sample Data using
  - Train-Test Split Ratio of
    - 80% - 20%
- Training Data
  - Total Instances = 80
    - Survived = 40
    - Not Survived = 40
- Testing Data
  - Total Instances = 20
    - Survived = 10
    - Not Survived = 10

SLIDE
Training Data

- The following Table shows the Training Data
  - See training-data-encoded.csv File in Supporting Material

Instance No.	Input				Output
	PClass	Gender	Sibling	Embarked	Survived
x_1	2	1	0	2	0

x_2	1	0	3	2	1
x_3	2	1	3	2	0
x_4	2	0	1	2	1
x_5	2	1	3	1	0
x_6	0	0	1	2	1
x_7	2	1	3	2	0
x_8	2	1	3	2	1
x_9	0	1	3	2	0
x_{10}	1	1	3	2	1
x_{11}	2	1	0	1	0
x_{12}	0	1	3	0	1
x_{13}	0	1	3	0	0
x_{14}	1	0	3	2	1
x_{15}	1	1	3	2	0
x_{16}	2	1	0	0	1
x_{17}	2	1	2	0	0
x_{18}	2	1	3	2	1
x_{19}	2	1	3	2	0
x_{20}	2	0	0	0	1
x_{21}	2	1	3	0	0
x_{22}	0	0	3	2	1
x_{23}	0	1	3	0	0
x_{24}	1	0	3	2	1
x_{25}	2	0	0	2	0
x_{26}	2	0	3	2	1
x_{27}	2	1	3	2	0
x_{28}	2	1	3	2	1
x_{29}	2	1	3	2	0
x_{30}	2	0	0	1	1
x_{31}	2	0	2	2	0
x_{32}	0	0	3	0	1
x_{33}	2	1	2	2	0
x_{34}	0	0	3	2	1
x_{35}	0	1	0	0	0
x_{36}	0	0	3	2	1

x_{37}	0	1	0	2	0
x_{38}	2	0	0	1	1
x_{39}	2	0	0	2	0
x_{40}	1	0	0	2	1
x_{41}	1	0	0	2	0
x_{42}	2	0	3	1	1
x_{43}	2	1	3	0	0
x_{44}	0	0	3	0	1
x_{45}	2	1	1	2	0
x_{46}	2	0	0	2	1
x_{47}	2	1	3	2	0
x_{48}	2	1	3	2	1
x_{49}	2	1	0	2	0
x_{50}	1	0	3	2	1
x_{51}	0	1	1	2	0
x_{52}	2	1	3	2	1
x_{53}	2	0	0	2	0
x_{54}	1	0	3	2	1
x_{55}	2	0	1	2	0
x_{56}	2	0	3	2	1
x_{57}	2	0	3	2	0
x_{58}	2	0	0	2	1
x_{59}	2	1	3	0	0
x_{60}	2	1	3	2	1
x_{61}	1	1	3	2	0
x_{62}	0	0	0	2	1
x_{63}	2	1	1	1	0
x_{64}	2	0	3	1	1
x_{65}	1	0	3	2	0
x_{66}	2	0	3	0	1
x_{67}	2	1	3	2	0
x_{68}	1	1	3	2	1
x_{69}	2	1	3	2	0
x_{70}	0	1	3	2	1
x_{71}	2	1	3	2	0

x_{72}	1	1	2	2	1
x_{73}	2	1	3	0	0
x_{74}	2	0	3	2	1
x_{75}	2	0	3	2	0
x_{76}	0	0	0	2	1
x_{77}	2	1	0	2	0
x_{78}	1	0	3	2	1
x_{79}	2	1	3	2	0
x_{80}	2	0	2	2	1

SLIDE

Testing Data

- The following Table shows the Testing Data
  - See testing-data-encoded.csv File in Supporting Material

Instance No.	Input				Output
	PClass	Gender	Sibling	Embarked	Survived
x_1	2	1	2	2	0
x_2	2	0	3	2	1
x_3	2	1	3	2	0
x_4	0	1	0	2	1
x_5	1	1	3	2	0
x_6	2	0	0	1	1
x_7	1	1	3	2	0
x_8	0	0	0	2	1
x_9	2	1	3	2	0
x_{10}	1	0	3	2	1
x_{11}	1	1	3	2	0
x_{12}	1	0	3	2	1
x_{13}	2	1	3	0	0
x_{14}	0	1	0	2	1
x_{15}	0	1	0	2	0
x_{16}	2	1	0	0	1

x_{17}	2	1	0	1	0
x_{18}	2	1	0	2	1
x_{19}	1	1	0	2	0
x_{20}	2	1	0	2	1

Step 07: Select Suitable Evaluation Measure(s)

SLIDE

Step 07: Select Suitable Evaluation Measure(s)

- **I will use the Accuracy Evaluation Measure to evaluate the performance of the Model**
- **Accuracy**
  - **Accuracy is defined as the proportion of correctly classified Test Instances**

$$Accuracy = \frac{Number\ of\ Correctly\ Classified\ Test\ Instances}{Total\ Number\ of\ Test\ Instances}$$

- **Note**
  - **Error = 1 - Accuracy**

Step 08: Execute First Two Phases of Machine Learning Cycle

SLIDE

Step 8: Execute First Two Phases of Machine Learning Cycle

- **Recall the Equation**

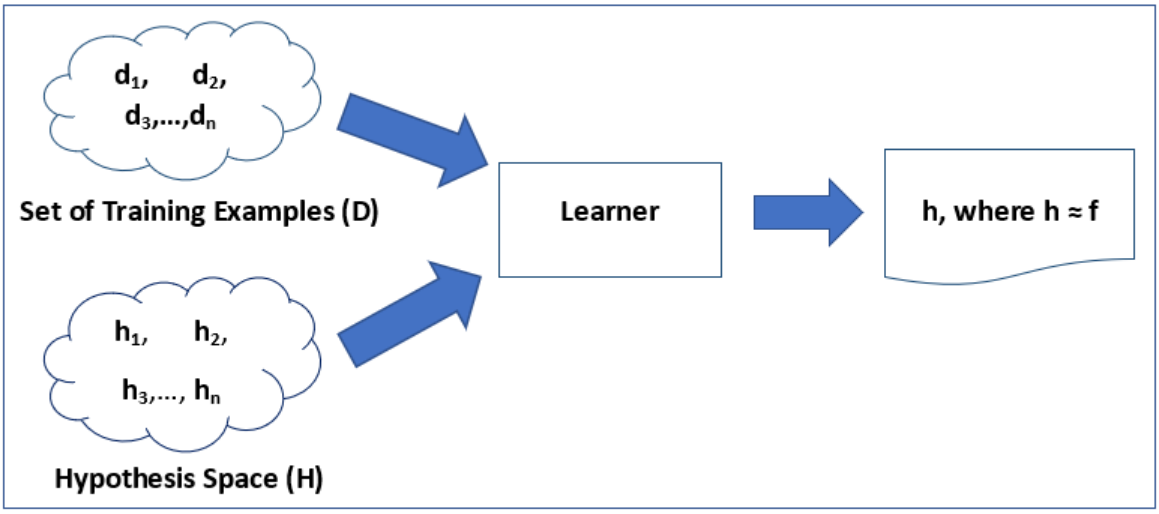
$$Data = Model + Error$$

- Training Phase
  - Use Training Data to build the Model
- Testing Phase
  - Use Testing Data to evaluate the performance of the Model
- Note that we aim to
  - Learn an Input-Output Function

SLIDE

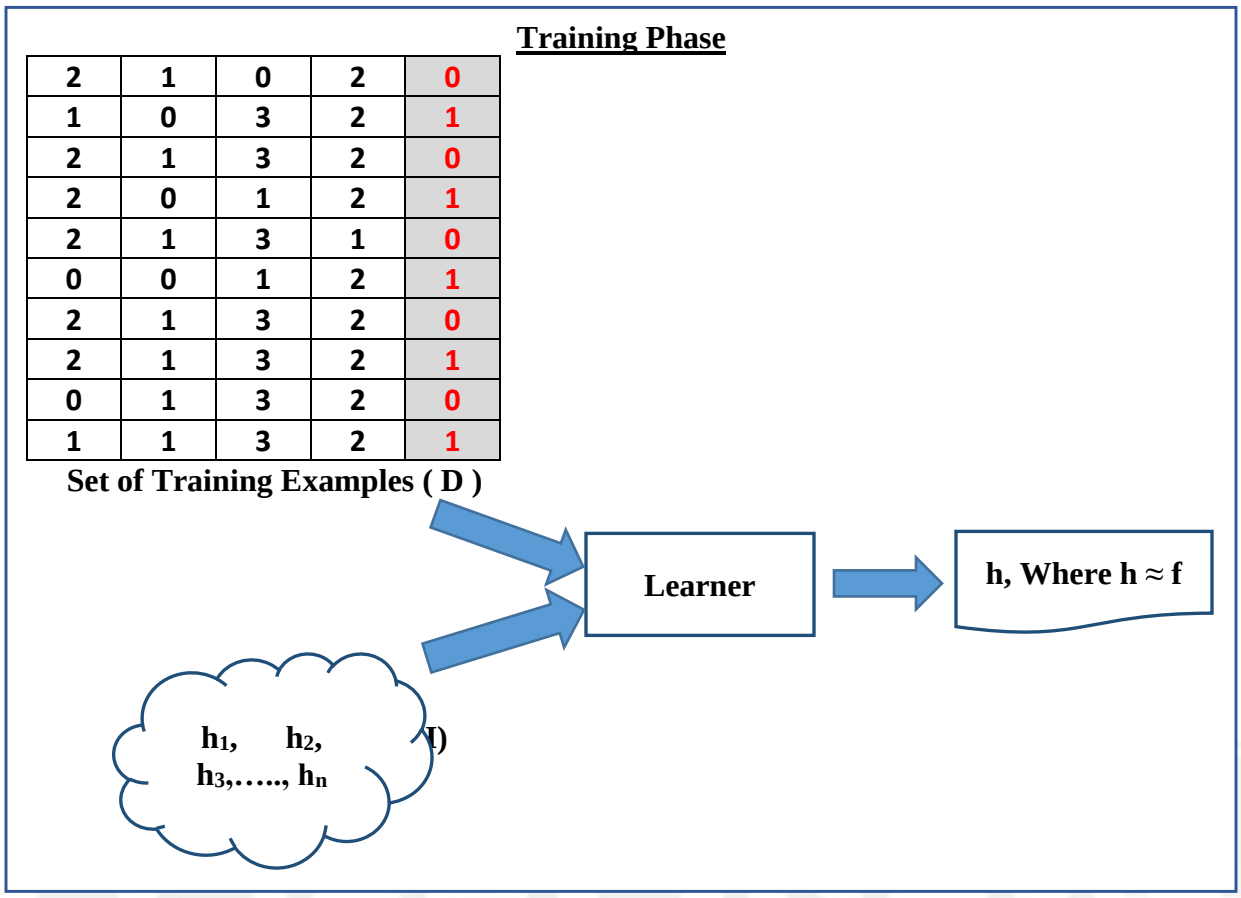
General Settings - Learning Input-Output Function

- Recall – Our goal is to
  - Learn an Input-Output Function



SLIDE

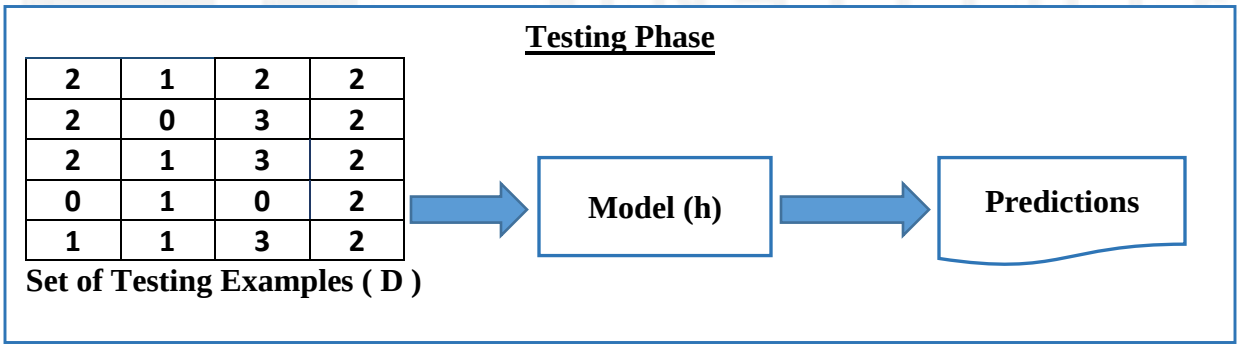
Training Phase



SLIDE

Testing Phase

- Apply Model on the Testing Data



SLIDE

Testing Phase Cont…

- The following Table shows the Predictions Returned by the Model (h)
  - See model-predictions.csv File in Supporting Material

Instance No.	Input				Output	
	PClass	Gender	Sibling	Embarked	Survived	Predictions
x_1	2	1	2	2	0	0
x_2	2	0	3	2	1	1
x_3	2	1	3	2	0	0
x_4	0	1	0	2	1	0
x_5	1	1	3	2	0	0
x_6	2	0	0	1	1	1
x_7	1	1	3	2	0	0
x_8	0	0	0	2	1	1
x_9	2	1	3	2	0	0
x_{10}	1	0	3	2	1	1
x_{11}	1	1	3	2	0	0
x_{12}	1	0	3	2	1	1
x_{13}	2	1	3	0	0	0
x_{14}	0	1	0	2	1	0
x_{15}	0	1	0	2	0	0
x_{16}	2	1	0	0	1	1
x_{17}	2	1	0	1	0	0
x_{18}	2	1	0	2	1	0
x_{19}	1	1	0	2	0	0
x_{20}	2	1	0	2	1	0

SLIDE

Testing Phase, Continue

- Calculating Accuracy
  - To calculate Accuracy, we will compare

- Actual Values with Predicted Values

- Note
  - To explain calculations more clearly, I have converted Numerical Predicted Values to Categorical Predicted Values

Instance No.	Input				Output		
	PClass	Gender	Sibling	Embarked	Actual Values	Predicted Values	Score
x_1	Third	Male	Two	Southampton	No	No	1
x_2	Third	Female	Zero	Southampton	Yes	Yes	1
x_3	Third	Male	Zero	Southampton	No	No	1
x_4	First	Male	One	Southampton	Yes	No	0
x_5	Second	Male	Zero	Southampton	No	No	1
x_6	Third	Female	One	Queenstown	Yes	Yes	1
x_7	Second	Male	Zero	Southampton	No	No	1
x_8	First	Female	One	Southampton	Yes	Yes	1
x_9	Third	Male	Zero	Southampton	No	No	1
x_{10}	Second	Female	Zero	Southampton	Yes	Yes	1
x_{11}	Second	Male	Zero	Southampton	No	No	1
x_{12}	Second	Female	Zero	Southampton	Yes	Yes	1
x_{13}	Third	Male	Zero	Cherbourg	No	No	1
x_{14}	First	Male	One	Southampton	Yes	No	0
x_{15}	First	Male	One	Southampton	No	No	1
x_{16}	Third	Male	One	Cherbourg	Yes	Yes	1
x_{17}	Third	Male	One	Queenstown	No	No	1
x_{18}	Third	Male	One	Southampton	Yes	No	0
x_{19}	Second	Male	One	Southampton	No	No	1
x_{20}	Third	Male	One	Southampton	Yes	No	0

$$Accuracy = \frac{16}{20} = 0.8$$

Step 09: Analyze Results

SLIDE
Step 9: Analyze Results

- The assumption for this Example
  - Here, I am assuming that the Model
    - performed well on large Test Data and we can apply it in the real-world 😊

Step 10: Execute 3rd and 4th Phases of Machine Learning Cycle

SLIDE
Step 10: Execute 3rd and 4th Phases of Machine Learning Cycle

- Application Phase
  - Model is deployed in Real-world to make predictions on Real-time Data
- Steps – Make Predictions on Real-time Data
  - Step 1: Take Input from User
  - Step 2: Convert User Input into Feature Vector
    - The same as Feature Vectors of Sample Data
  - Step 3: Apply Model on the Feature Vector of the unseen instance
  - Step 4: Return Prediction to the User

SLIDE
Example – Making Predictions on Real-time Data

- Step 1: Take Input from User
  - User Input

Please enter PClass: First
Please enter your Gender: Female

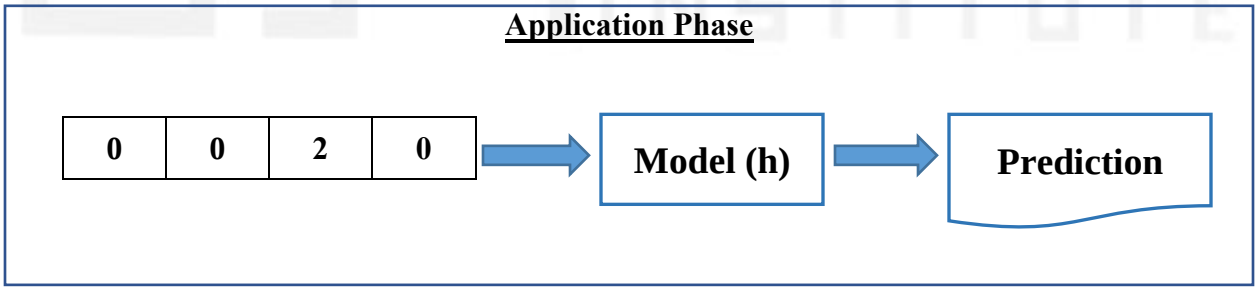
Please enter your Sibling: Two
Please enter Embarked: Cherbourg

- Step 2: Convert User Input into Feature Vector
  - Feature Vector

<First, Female, Two, Cherbourg>

- Feature Vector after Label Encoding
  - Exactly same as Label Encoded Feature Vectors of Sample Data
  - Label Encoded Feature Vector
    - < 0, 0, 2, 0 >
- Step 3: Apply Model on the Label Encoded Feature Vector of unseen instance
  - Model (h) is applied on < 0, 0, 2, 0 >
- Step 4: Return Prediction to the User
  - 1 (Yes)

SLIDE
Application Phase



SLIDE
Feedback Phase

- **A Two-Step Process**
- **Step 1: After some time, take Feedback from**
  - **Domain Experts and Users on deployed Titanic Passenger Survival Prediction System**
- **Step 2: Make a List of Possible Improvements based on Feedback receive**

Step 11: Improve Titanic Passenger Survival Prediction System based on Feedback

---

SLIDE

Step 11: Improve Titanic Passenger Survival Prediction System based on Feedback

- **Go to Step 1 and improve the Titanic Passenger Survival Prediction System based on**
  - **List of Possible Improvements made in Step 10**

TODO and Your Turn

SLIDE
TODO

- **Task**
  - **Consider the Heart Disease Classification Problem. The main aim is to predict whether a patient has Heart Disease or Not (i.e. Binary Classification Problem)?**
  - **Heart Disease Dataset Link**
    - **URL:**
  - **For simplicity, I have taken a sample of 100 instances from the Original Heart Disease Dataset**
    - **See heart-disease-sample-data.csv File in Supporting Material**
- **Note**
  - **Your answer should be**
    - **Well Justified**
- **Question**
  - **Write down the Input and Output of the Heart Disease Classification Problem?**
  - **Follow the Steps mentioned in this Lecture and show**
    - **How will you treat the Heart Disease Classification Problem as a Supervised Machine Learning Problem using Train-Test Split Approach?**

SLIDE
Your Turn

- **Task**
  - **Select a Problem (similar to the one given in TODO) and answer the questions given below**
- **Note**
  - **Your answer should be**
    - **Well Justified**
- **Questions**

- Write Input and Output for the selected Machine Learning Problem?
- Follow the Steps mentioned in this Lecture and show
  - How will you treat the selected Machine Learning Problem as a Supervised Machine Learning Problem using Train-Test Split Approach?

Lecture Summary

SLIDE
Lecture Summary

- Titanic Passenger Survival Prediction System – Task
  - Given
    - A Passenger (Represented as Set of Attributes)
  - Task
    - Automatically predict whether the Passenger Survived or Not
- Titanic Passenger Survival Prediction System – Input and Output
  - Input
    - A Passenger
  - Output
    - Survived / Not Survived
- The Problem of Titanic Passenger Survival Prediction is treated as a
  - Supervised Machine Learning Task
- The main goal of Titanic Passenger Survival Prediction System is to
  - Learn an Input-Output Function
    - i.e. Learn from Input to predict the Output
- Learning Input-Output Function – General Settings
  - Input to Learner
    - Set of Training Examples (D)
    - Set of Hypothesis (a.k.a. Hypothesis Space (H))
  - Job of Learner
    - The main job of a Learner is to search the Hypothesis Space (H) using the Set of Training Examples (D) to find out a Hypothesis (h) from Hypothesis Space (H), which best fits the Set of Training Examples (D)
  - Output of Learner
    - A Learner outputs a Hypothesis (h) from Hypothesis Space (H), which best fits the Set of Training Examples (D)
- Steps to treat the Titanic Passenger Survival Prediction System Problem as a Classification Problem

- Step 01: Decide the Learning Settings
- Step 02: Obtain Sample Data
- Step 03: Understand and Pre-process Sample Data
- Step 04: Represent Sample Data in Machine Understandable Format
- Step 05: Select Suitable Machine Learning Algorithms
- Step 06: Split Sample Data into Training Data and Testing Data
- Step 07: Select Suitable Evaluation Measure(s)
- Step 08: Execute First Two Phases of Machine Learning Cycle
  - Training Phase
  - Testing Phase
- Step 09: Analyze Results

```
If (Results are Good )
Then
 Move to the Next Step
Else
 Go to Step 01
```

- Step 10: Execute 3rd and 4th Phases of Machine Learning Cycle
  - Application Phase
  - Feedback Phase
- Step 11: Based on Feedback
  - Go to Step 01 and Repeat all the Steps