Machine Learning

AIML CZG565

M2 : Machine Learning Workflow

A. Raghunathan

BITS - CSIS

BITS Pilani

Pilani Campus

Machine Learning

Disclaimer and Acknowledgement

- These content of modules & context under topics are prepared by the course owner instructor Dr.Sugata, instructors Rajavadhana , lab material by Dr.Pankaj with grateful acknowledgement to many others who made their course materials freely available online.

- The content for these slides has been obtained from books and various other source on the Internet

- We here by acknowledge all the contributors for their material and inputs.

- We have provided source information wherever necessary

- Students are requested to refer to the textbook w.r.t detailed content of the presentation deck shared over canvas

- We have reduced the slides from canvas and modified the content flow to suit the requirements of the course and for ease of class presentation

Course Plan

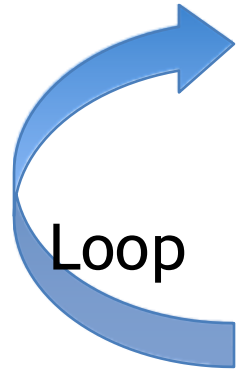|		
M1	Introduction	
M2	Machine learning Workflow	
M3	Linear Models for Regression	
M4	Linear Models for Classification	
M5	Decision Tree	
M6	Instance Based Learning	
M7	Support Vector Machine	
M8	Bayesian Learning	
M9	Ensemble Learning	
M10	Unsupervised Learning	
M11	Machine Learning Model Evaluation/Comparison	

Agenda

- Role of Data

- Data Preprocessing / wrangling

- Data skewness removal (sampling)

- Model Training *(will be covered in subsequent modules)*

- Model Testing and performance metrics *(will be covered in subsequent modules)*

ML in a Nutshell

- Tens of thousands of machine learning algorithms
  - Hundreds new every year

- Every ML algorithm has three components
  - **Data Representation**
  - **Parameter Optimization**
  - **Model Evaluation, Selection**

Slide credit: Pedro Domingos

BITS Pilani, Pilani Campus

ML in Practice

- Understand domain, prior knowledge, and goals
- Data integration, selection, cleaning, pre-processing, etc.
- Learn optimal parameter of the models
- Interpret results
- Consolidate and deploy discovered knowledge

Loop

Based on a slide by Pedro Domingos

Data

Definition of Data

Collection of *data objects* and their *attributes*

An *attribute* is a property or characteristic of an object
- – Examples: eye color of a person, temperature, etc.
- – *aka* variable, field, characteristic, dimension, or feature

A collection of attributes describe an *object*
- – Object is also known as record, point, case, sample, entity, or instance

Attributes

Objects

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Types of Attributes

There are different types of attributes

- Nominal
  - Examples: ID numbers, zip codes
- Ordinal
  - Examples: rankings (e.g., taste of potato chips on a scale from 1-10), grades, height {tall, medium, short}
- Interval
  - Examples: calendar dates, temperatures in Celsius or Fahrenheit.
- Ratio
  - Examples: temperature in Kelvin, length, counts, elapsed time (e.g., time to run a race)

Discrete and Continuous Attributes

Discrete Attribute
- Has only a finite or countably infinite set of values
- Examples: zip codes, counts, or the set of words in a collection of documents
- Often represented as integer variables.
- Note: binary attributes are a special case of discrete attributes

Continuous Attribute
- Has real numbers as attribute values
- Examples: temperature, height, or weight.
- Practically, real values can only be measured and represented using a finite number of digits.
- Continuous attributes are typically represented as floating-point variables.

Important Characteristics of Data

– Dimensionality (number of attributes)
- • High dimensional data brings a number of challenges

– Sparsity
- • Only presence counts

– Resolution
- • Patterns depend on the scale

– Size
- • Type of analysis may depend on size of data

Data Types

- Relational/Object

- Transactional Data

- Document Data

- Web & Social

  Network Data

- Spatial Data

- Time Series

- Sequence Data

Discrete Numeric   Binary

Binary   Ordinal   Continuous Numeric

	ServiceRating	IsPriority Customer	CardType	Credit Score	isMultipleAccount Holder
Jack	5	Yes	Platinum	7.5	Yes
Jill	2	Yes	Gold	8.2	No
John	9	No	Gold	7	Yes

Data Types

- Relational/Object

- Transactional Data

- Document Data

- Web & Social

Network

- Spatial Data

- Time Series

- Sequence Data

	Purchase 1	Purchase 2	......	...	......
Jack	Paper, Pen, Medicine	Milk, Bread, Egg, Milk			
Jill	Rice, Medicine, Vegetable, Milk	Rice, Egg, Vegetable, Milk			
John	Bread, Jam, Butter , Jam	Milk, Bread, Pasta, Medicine			

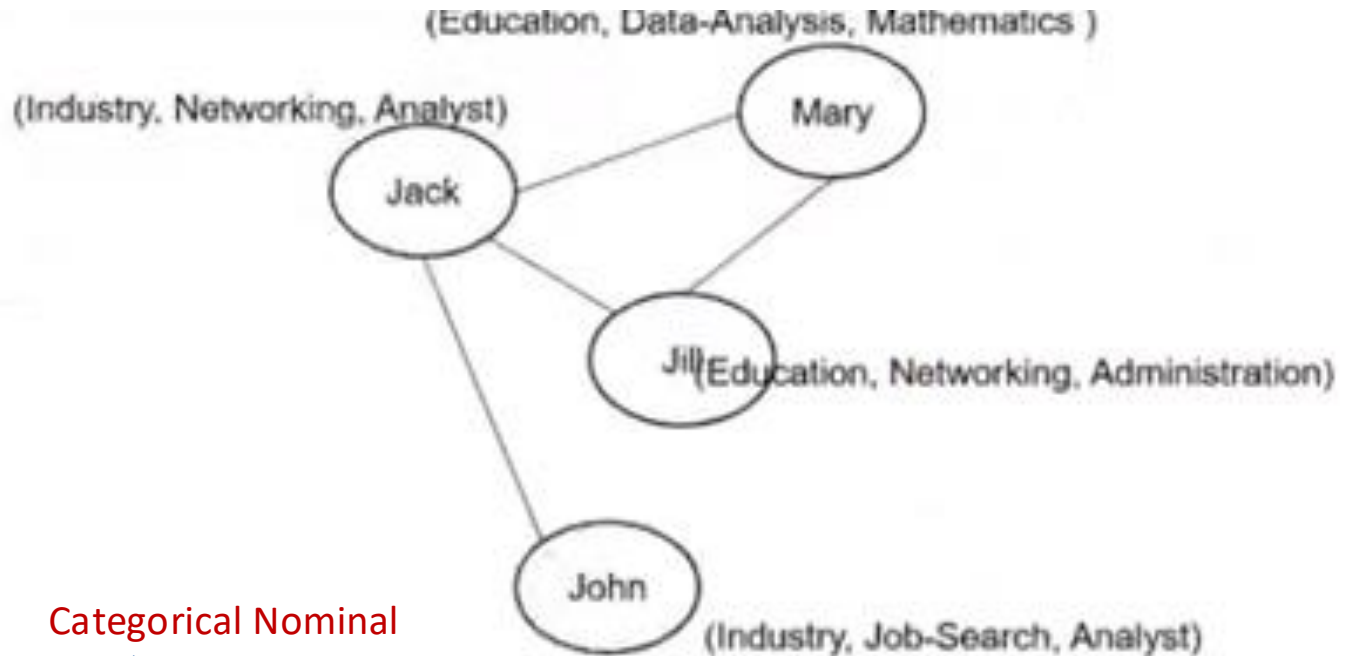|	Items Bought
Transaction 1	Paper, Pen, Medicine
Transaction 2	Rice, Medicine, Vegetable, Milk
Transaction 3	Milk, Bread, Egg, Milk
Transaction 4	Bread, Jam, Butter , Jam

Data Types

- Relational/Object

- Transactional
  Data

- Document Data

- Web & Social
  Network Data

- Spatial Data

- Time Series

- Sequence Data

	Trend	Data	Story	Mining	Cloth
Document 1	5	10	4	8	0
Document 2	5	5	8	0	7
Document 3	2	8	2	4	0

Data Types

- Relational/Object

- Transactional Data

- Document Data

- Web & Social Network Data

- Spatial Data

- Time Series

Categorical Nominal

(Education, Data-Analysis, Mathematics)

(Industry, Networking, Analyst)

Jack

Mary

Jill (Education, Networking, Administration)

John

(Industry, Job-Search, Analyst)

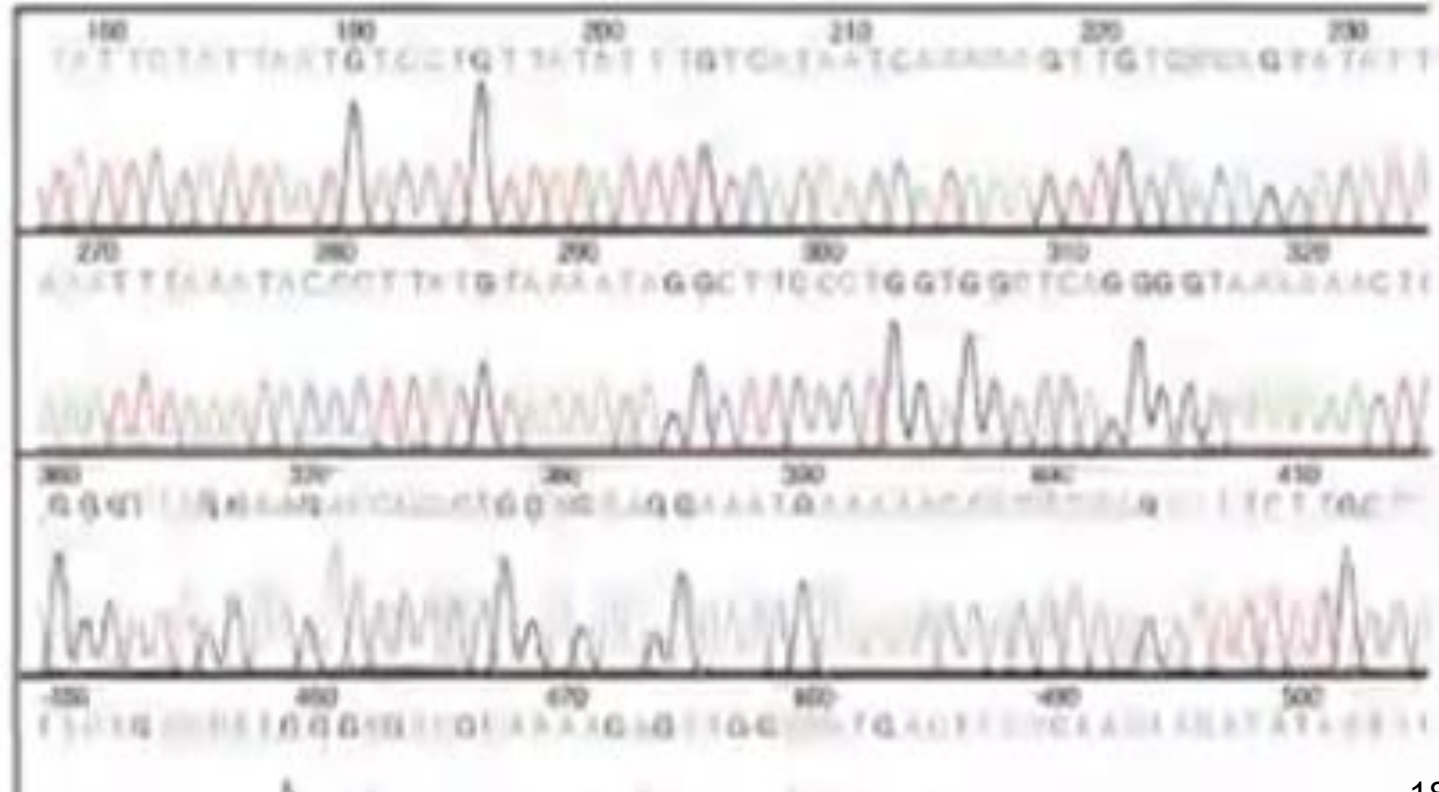|	Work-Field	Purpose of Connect	Domain of work	No.of. Connections	Link to parent	...
John	Industry	Job-Search	Analyst	1	Jack	
Mary	Education	Data-Analysis	Mathematics	2	Jack, Jill	

15

Data Types

- Relational/Object

- Transactional Data

- Document Data

- Web & Social Network Data

- Spatial Data

- Time Series

- Sequence Data

User	Call Type	Call Duration	Time Stamp	Tower Cell ID	Latitude	Longitude
9341959679262440000	Voice	10	2019-11-20 14:15 :01	123456	12.97	77.58
9341959679262440000	Text	0	2019-11-19 11:10 :09	123456	12.73	77.82
9221959659362440000	Voice	10	2019-11-20 14:15 :01	324576	19.07	72.87

Data Types

- Relational/Object

- Transactional Data

- Document Data

- Web & Social Network Data

- Spatial Data

- Time Series

- Sequence Data

Sample Time Series Data & Visualisation

	Profit%	EmployeeAttrition Rate	CreditScore
Jan 18	28.5	10	20
April 18	35.9	42	50
July 18	46.8	45	40
Oct 18	50.7	5	45
Jan 19	70.25	8	45
April 19	75.85	45	20
July 19	90.5	47	50

Data Types

- Relational/Object

- Transactional Data

- Document Data

- Web & Social
  Network Data

- Spatial Data

- Time Series

- Sequence Data

Identify the Data Types and attribute types

A bank wishes to analyze its customer base for targeted marketing and needs to segment the customers based on its account information with its branch. Post analysis it might be interested to target potential customers of high income level possessing Titanium card types.

Name	Gender	Service Rating	Is Priority Customer?	Card Type	Credit Score	Is Multiple Account Holder	Income Level	Region
Jack	Male	5	Yes	Platinum	7.5	Yes	Upper	BGLR
Jill	Female	2	Yes	Gold	8.2	No	Middle	DELHI
John	Male	9	No	Gold	7	Yes	Lower	BGLR
Mary	Male	6	No	Gold	6.0	No	Lower	BGLR

Data Quality

BITS Pilani, Pilani Campus

Data Quality …

What kinds of data quality problems?

How can we detect problems with the data?

What can we do about these problems?
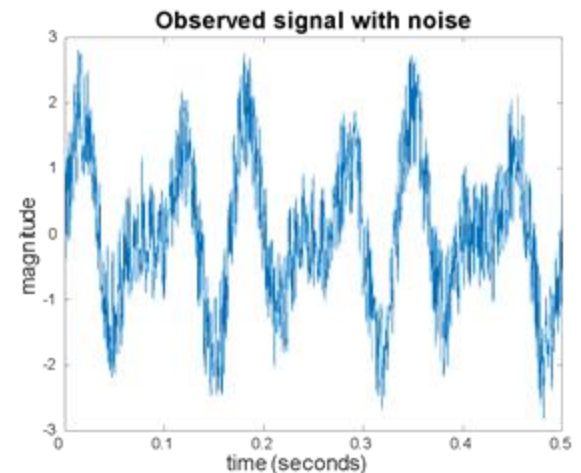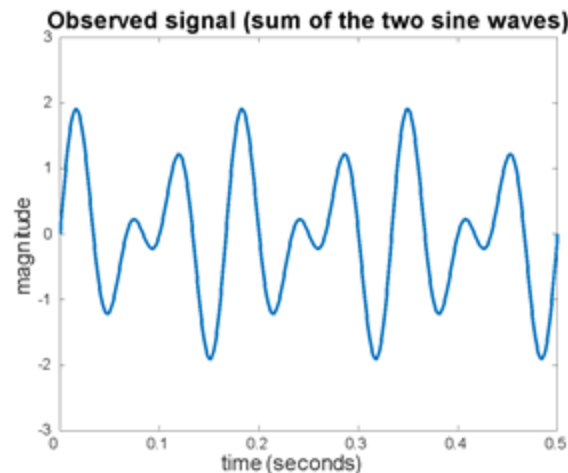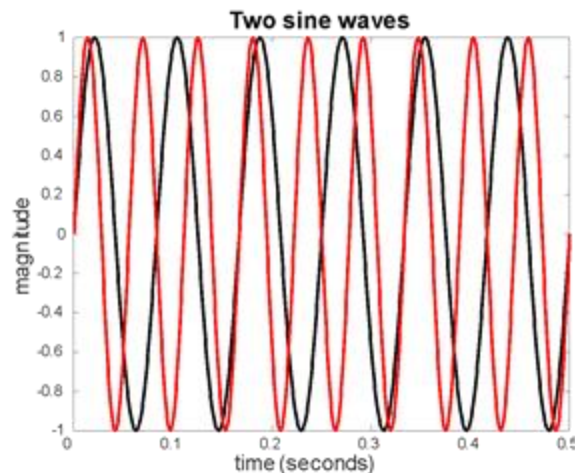
Examples of data quality problems:
- Noise and outliers
- Wrong data
- Fake data
- Missing values
- Duplicate data

Noise

For objects, noise is an extraneous object

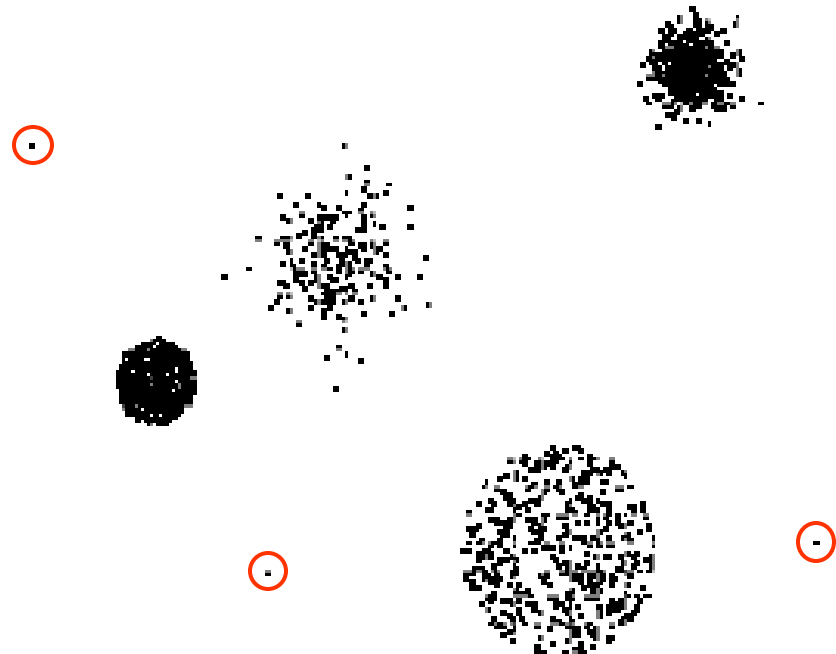For attributes, noise refers to modification of original values

- Examples: distortion of a person's voice when talking on a poor phone and "snow" on television screen
- The figures below show two sine waves of the same magnitude and different frequencies, the waves combined, and the two sine waves with random noise
    - The magnitude and shape of the original signal is distorted

Two sine waves | Observed signal (sum of the two sine waves) | Observed signal with noise

Outliers

Outliers are data objects with characteristics that are considerably different than most of the other data objects in the data set

- **Case 1:** Outliers are noise that interferes with data analysis

- **Case 2:** Outliers are the goal of our analysis
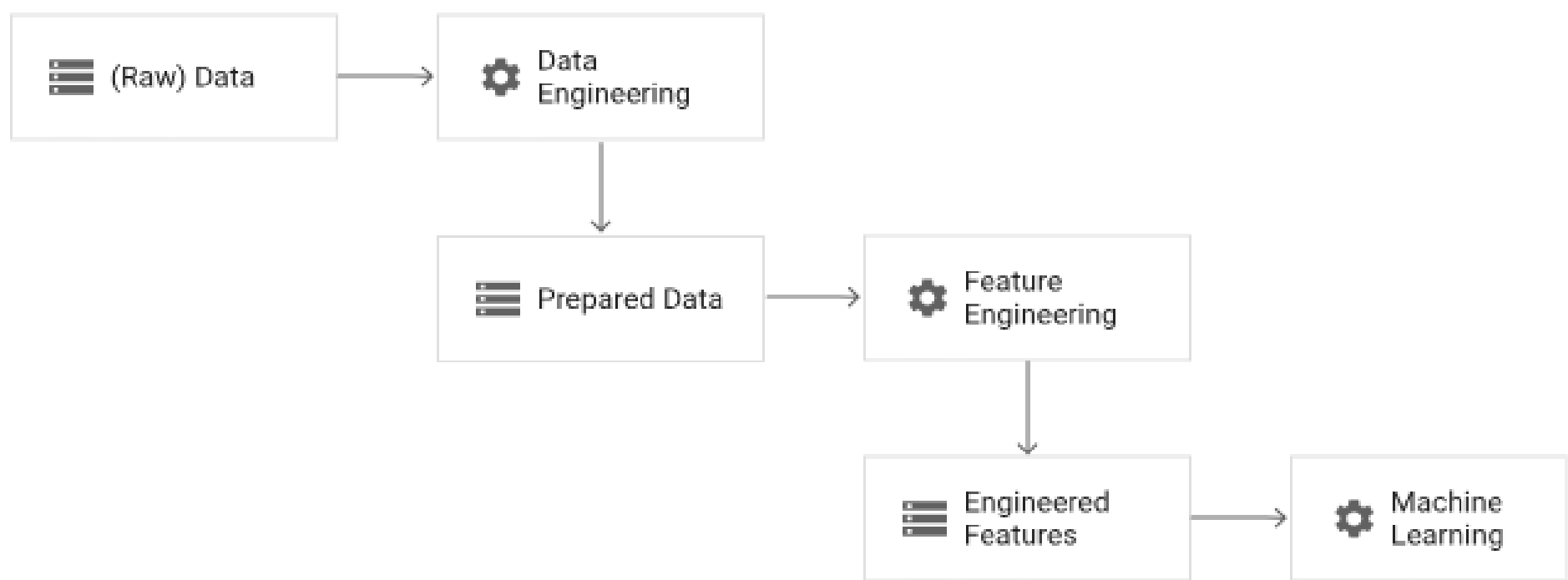
  - Credit card fraud

  - Intrusion detection

Data Pre-Processing

Preprocessing

- Preprocessing the data for ML involves both data engineering and feature engineering
- Data engineering : process of converting raw data into prepared data.
- Feature engineering : tunes the prepared data to create the features that are expected by the ML model

https://www.tensorflow.org/tfx/guide/tft_bestpractices

BITS Pilani, Pilani Campus

Case study

BITS WILP is in collaboration with multiple IT companies interested to upskill and level skill their employee through inducting them in tailored Mtech AIML program. Over a year of successful completion , the student are yet to complete another one semester and enroll in Dissertation to complete the program with certification. ***Students of similar academic background irrespective of the time of enrollment, seems to score more or less in same range in every semester***. **Accounting department requires to complete few academic year closure documentation for which , they would have to bill the collaborative organization based on the prospective no. of students who might be eligible for project semester**. As of current semester the students have completed their exams but the process is pending for grading. As Data analyst help accounts team to get necessary information with the given available data across all the collaborative program.

Challenge 1 : Insufficient Training Data.
Idea : Trade-off algorithm vs Data readiness

AttributesOfInterest
Name
Gender
Age
DataOfBirth
Organisation
JobTitle
NatureOfJob
EntranceScore
EligibilityScore
PreviousDegree
WILPBatch
Section
ISM
MFML
ACI
ML
NLP
……….

Data Pre-processing

- Data Aggregation
- Data cleansing
- Instances selection and partitioning
- Feature tuning

Data Aggregation

- Consider a data set consisting of transactions (data objects) recording the daily sales of products in various store locations (Minneapolis, Chicago, Paris, ...) for different days over the course of a year

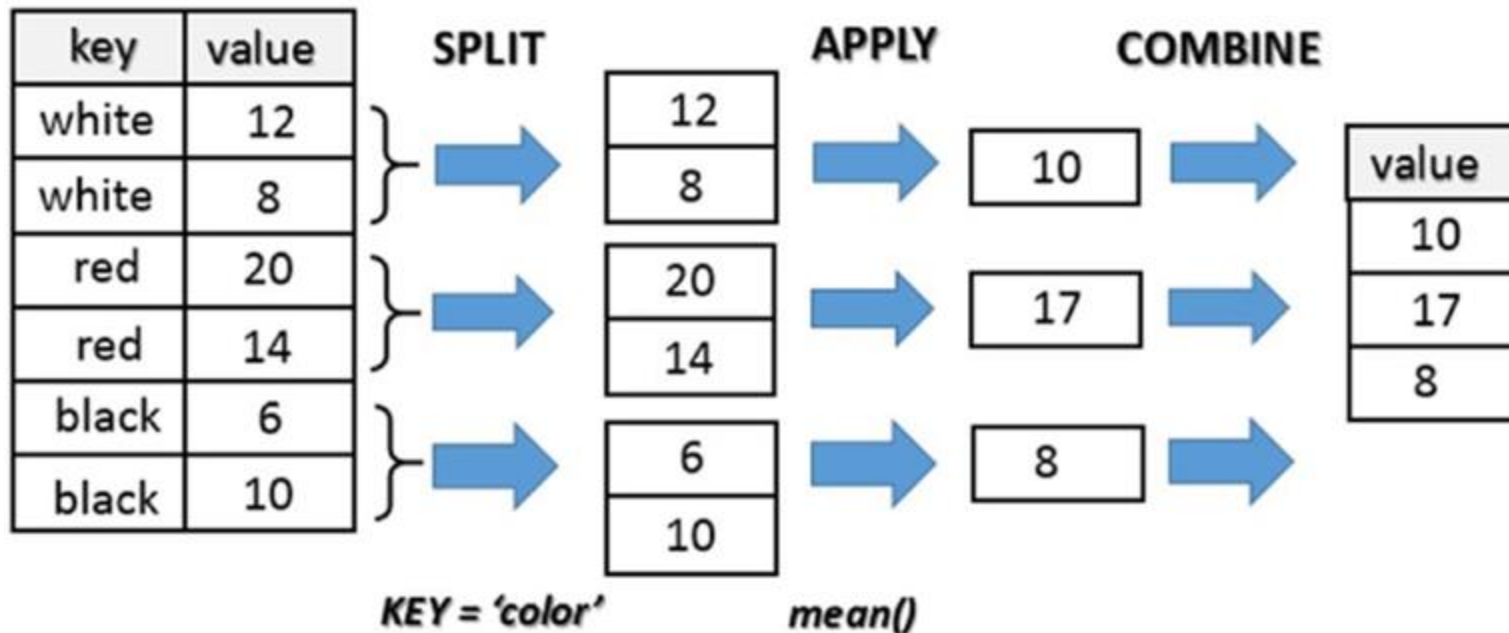- How can we aggregate this data?

Data set containing information about customer purchases.

Transaction ID	Item	Store Location	Date	Price	...
⋮	⋮	⋮	⋮	⋮	
101123	Watch	Chicago	09/06/04	$25.99	...
101123	Battery	Chicago	09/06/04	$5.99	...
101124	Shoes	Minneapolis	09/06/04	$75.00	...
⋮	⋮	⋮	⋮	⋮	

Data Aggregation

Python Group By Example

Source: Python for Data Analytics by Fabio Nelli

BITS Pilani, Pilani Campus

Data cleansing

- Removing or correcting records of corrupted or invalid values from raw data
  - NOISY: containing noise, errors, or outliers .
    - e.g., Salary="−10" (an error)
  - INCONSISTENT: containing discrepancies in codes or names, e.g.,
    - Age="42", Birthday="03/07/2010"
    - Was rating "1, 2, 3", now rating "A, B, C"
    - discrepancy between duplicate records
  - INTENTIONAL (e.g., disguised missing data)
    - Jan. 1 as everyone's birthday
- Removing records that are missing a large number of columns
- Duplicate data

Data cleansing

A mistake or a millionaire?

Missing values

Inconsistent duplicate entries

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	10000K	Yes
6	No	NULL	60K	No
7	Yes	Divorced	220K	NULL
8	No	Single	85K	Yes
9	No	Married	90K	No
9	No	Single	90K	No

Data cleansing

Imputing Missing values

	DATE	air_mv	air_mv_zero	air_mv_previous	air_mv_mean	air_expand
1	JAN49	112	112	112	112	112
2	FEB49	118	118	118	118	118
3	MAR49	132	132	132	132	132
4	APR49	129	129	129	129	1.
5	MAY49		0	129	284.54385965	128.29783049
6	JUN49	135	135	135	135	135
7	JUL49		0	135	284.54385965	144.73734152
8	AUG49	148	148	148	148	148
9	SEP49	136	136	136	136	136
10	OCT49	119	119	119	119	119
11	NOV49		0	119	284.54385965	116.19900978
12	DEC49	118	118	118	118	118
13	JAN50	115	115	115	115	115
14	FEB50	126	126	126	126	126
15	MAR50	141	141	141	141	141

Labels: Insert missing records; Replace with 0; Replace with last known value; Replace with mean; Interpolate based on splines

Data cleansing

- IQR
    - Outliers are usually, a value higher/lower than 1.5 x IQR
- Z-score method (3 sigma)

BITS Pilani, Pilani Campus

Data cleansing

Handling outliers (univariate) using IQR

Interquartile Range (IQR):

- IQR = Q3 - Q1 (where Q1 is the 25th percentile and Q3 is the 75th percentile)

Outlier Detection:

- **Lower Bound**: Q1 - 1.5 * IQR
- **Upper Bound**: Q3 + 1.5 * IQR

Example:

- If Q1 = 10 and Q3 = 20, then IQR = 10
- Lower Bound = 10 - 1.5 * 10 = -5
- Upper Bound = 20 + 1.5 * 10 = 35
- Data points < -5 or > 35 are outliers

Exercise

Find the outlier in the following data using Inter-Quartile Range.

Data = 10,2, 11, 15,11,14,13,17,12,22,14,11.

1. Sort :10, 11, 11, 11, 12, 12, 13, 14, 14, 15, 17, 22
2. Median: (12+13)/2=12.5=Q2
3. Q1=11(25th percentile)
4. Q3=14.5(75th percentile)
5. IQR=Q3-Q1=3.5
6. Min=Q1-1.5IQR=5.75
7. Max=Q3+1.5IQR=19.75

Outlier=22

Quartile Formula

The Quartile Formula For Q1 $= \frac{1}{4}(n+1)^{th}$ term

The Quartile Formula For Q3 $= \frac{3}{4}(n+1)^{th}$ term

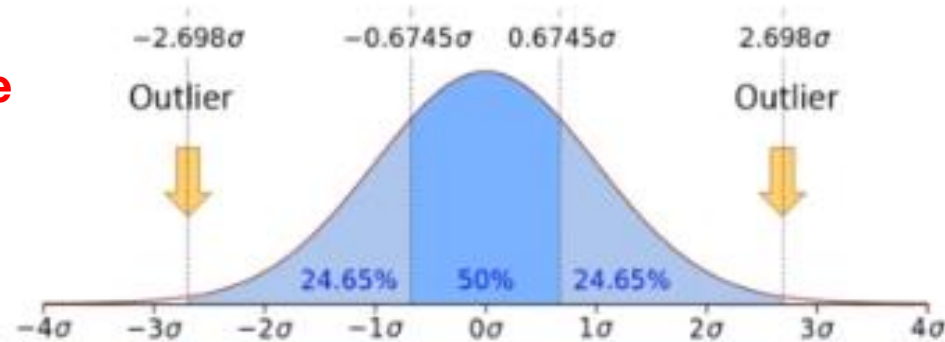The Quartile Formula For Q2 $= Q3 - Q1$ (Equivalent to Median)

Data cleansing

3 Sigma Rule: Based on the properties of a normal distribution

– **Mean (μ) and Standard Deviation (σ**

➢ **99% of the observations of a variable**

following a normal distribution lie

within mean +/- 3 X standard deviation

Outlier Detection:

– **Lower Bound**: $\mu - 3\sigma$
– **Upper Bound**: $\mu + 3\sigma$

Example Calculation:

– If $\mu = 50$ and $\sigma = 5$, then:
  - Lower Bound = 50 - 3 * 5 = 35
  - Upper Bound = 50 + 3 * 5 = 65
– Data points < 35 or > 65 are outliers

Instances selection and partitioning

Challenge 2 : Non-representative Training Data .

Idea : Training Data be representative of the new cases we want to generalize

➢ Small sample size leads to sampling noise. Increase sampling size.

➢ If sampling process is flawed, even large sample size can lead to sampling bias

The key principle for effective sampling is the following:

➢ Using a sample will work almost as well as using the entire data set, if the sample is representative

➢ A sample is representative if it has approximately the same properties (of interest) as the original set of data
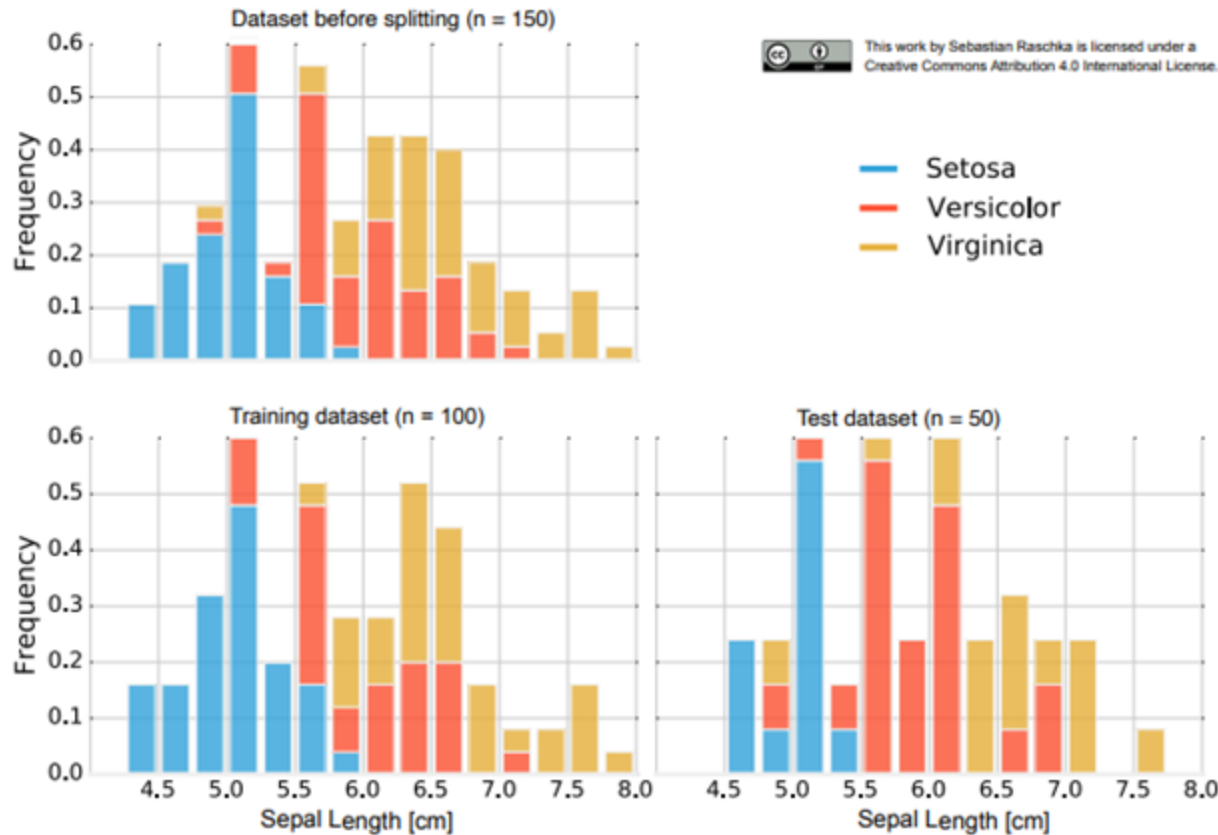
| 8000 points | 2000 Points | 500 Points | 37 |

Instances selection and partitioning

Sampling

Issues with Subsampling (Independence Violation)

IRIS Dataset of Flowers
50 Setosa,
50 Versicolor,
 50 Virginica

- **Random subsampling can assign 2/3 (100) to training set and 1/3 (50) to the test set**
- **Training set → 38 x Setosa, 28 x Versicolor, 34 x Virginica**
- **Test set → 12 x Setosa, 22 x Versicolor, 16 x Virginica**
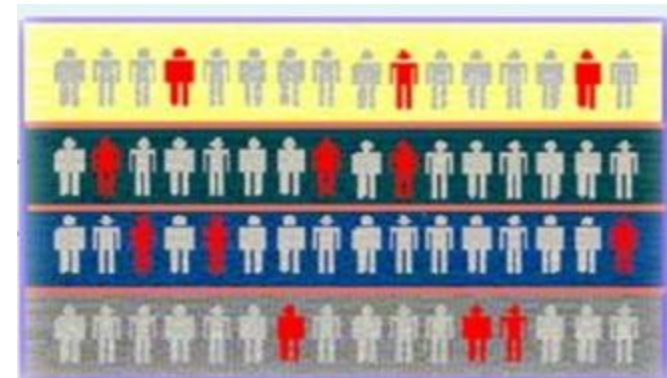
Instances selection and partitioning

Sampling - Frequently Used

Simple Random Type

Stratified Sampling Type

Clustered Sampling Type

Instances selection and partitioning

Sampling - Imbalanced Training Set

Scenario : Building Classifiers with Imbalanced Training Set

Modify the distribution of training data so that rare class is

well-represented in training set

- Under sample the majority class

- Over sample the rare class

Feature tuning

Feature Scaling

To map the continuous values from one range to target range to easily compare and fit in apt distribution to enable statistical processing
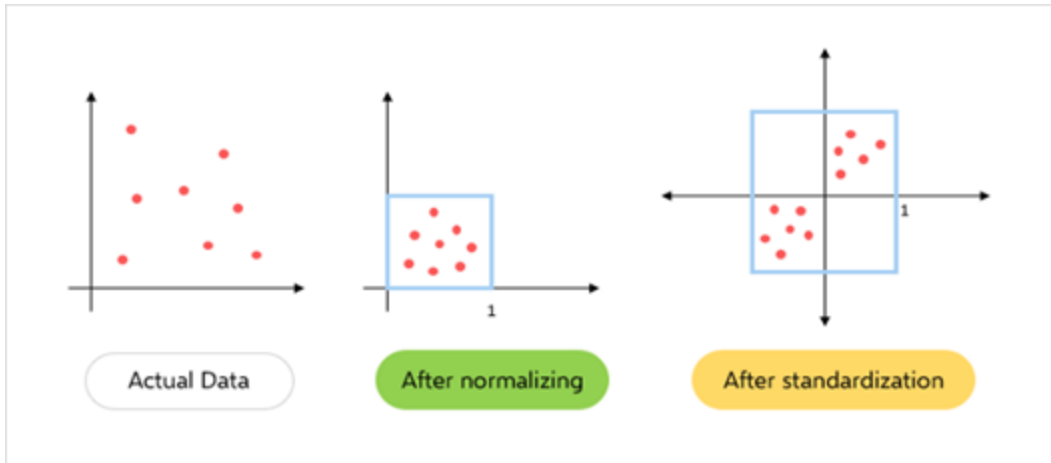
Feature scaling

Normalization

Standardization

$$X_{new} = \frac{X - min_A}{max_A - min_A}(new\_max_A - new\_min_A) + new\_min_A$$

$$X' = \frac{X - Mean}{Standard\ deviation}$$

Simple feature scaling

length
0.93
1
0.93

length
168.8
180.0
168.8

Min-max

length
0.84
0.9
0.84

Z-score

length
7.2
12
7.2

Actual Data

After normalizing

After standardization

Note: Scaling the target values is generally not required

Feature tuning

Feature Scaling - Normalization Vs Standardization

- Normalization
  - when approximate upper and lower bounds on data is known
  - When data is approximately uniformly distributed across that range. E.g age. Not to be used on skewed attribute e.g. income
  - when the algorithms do not make assumptions about the data distribution e.g. (KNN,NN)
  - scales in a range of [0,1] or [-1,1]
- Standardization
  - used when algorithms make assumptions about the data distribution (Gaussian distribution)
  - not bounded by range
  - less affected by outliers

Note:
 Fit the scalers to the training data only
Use them to transform the training set and the test set

Feature tuning
Feature Scaling - Normalization Vs Standardization

Min-max normalization: to [new_min$_A$, new_max$_A$]

$$v' = \frac{v - min_A}{max_A - min_A}(new\_max_A - new\_min_A) + new\_min_A$$

Ex. Let income range $12,000 to $98,000 normalized to [0.0, 1.0]. Then $73,000 is mapped to

$$\frac{73,600 - 12,000}{98,000 - 12,000}(1.0 - 0) + 0 = 0.716$$

Z-score normalization/Standardization (μ: mean, σ: standard deviation):

$$v' = \frac{v - \mu_A}{\sigma_A}$$

Ex. Let μ = 54,000, σ = 16,000. Then

$$\frac{73,600 - 54,000}{16,000} = 1.225$$

Normalization by decimal scaling

$$v' = \frac{v}{10^j}$$

Where j is the smallest integer such that Max($|v'|$) < 1

Feature Engineering

Feature Engineering

Feature engineering needed for coming up with a good set of features - Irrelevant Features

Feature extraction
- Dimensionality reduction

Feature selection
- more useful features to train on among existing features.

Feature Construction
- Combine existing features to produce a more useful one

Feature Transformation

Case study

Input:

WILP student details enrolled in Mtech AIML program.

Analysis:

Predict the GPA of the AIML students in Semester3 to estimate the no. of students who might enroll in dissertation

Observation:

Students with similar educational background tend to perform same in the exams

AttributesOfInterest
Name
Gender
Age
DataOfBirth
Organisation
JobTitle
NatureOfJob
EntranceScore
EligibilityScore
PreviousDegree
WILPBatch
Section
ISM
MFML
ACI
ML
NLP
……….

Feature Engineering - Extraction

Curse of Dimensionality

AttributesOfInterest
Name
Gender
Age
DataOfBirth
Organisation
JobTitle
NatureOfJob
EntranceScore
EligibilityScore
PreviousDegree
WILPBatch
Section
ISM
MFML
ACI
ML
NLP
……….

- Reducing the number of features by creating lower-dimension

- When dimensionality increases, data becomes increasingly sparse in the space that it occupies

- Solution : Dimensionality Reduction techniques: e.g Principal Components Analysis (PCA)

Feature Engineering - Selection

- Selecting a subset of the input features for training the model
  - Handle Redundant features
  - Remove Irrelevant feature
  - dropping features (missing a large number of value)

dataframe= dataframe.drop(['*COLNAME-1*','*COLNAME-2*'],axis=1)

AttributesOfInterest
~~Name~~
~~Gender~~
~~Age~~
~~DataOfBirth~~
~~Organisation~~
~~JobTitle~~
~~NatureOfJob~~
~~EntranceScore~~
~~EligibilityScore~~
PreviousDegree
~~WILPBatch~~
~~Section~~
ISM
MFML
ACI
ML
NLP
……….

Feature Engineering - construction

- Creating new features by using techniques
  - Polynomial expansion (by using univariate mathematical functions)
  - Feature crossing (to capture feature interactions)
  - Features can also be constructed by using business logic from the domain of the ML use case.

Feature Engineering - Transform

AttributesOfInterest
PreviousDegree SEM-1-Total SEM-2-Total SEM-3-Total CGPA isEligibleForDissertation

AttributesOfInterest
PreviousDegree SEM-1-GPA SEM-2-GPA SEM-3-GPA CGPA isEligibleForDissertation

AttributesOfInterest
PreviousDegree S1-isComplete S2-isComplete S3-isComplete CGPA isEligibleForDissertation

Feature Engineering - Transform
Encoding Numerical Features

- **Discretization** : Convert continuous attribute into a discrete attribute
  - Naive Bayes, decision trees and their ensembles including Random forest, Minimum distance classifiers or KNN prefer discrete features.
  - Also known as binning' or 'bucketing'
  - To handle outliers.
  - To improve the value spread i.e., spread of data
- Discretization involves converting the raw values of a numeric attribute (e.g., age) into
  - interval labels (e.g., 0–10, 11–20, etc.) OR
  - conceptual labels (e.g., youth, adult, senior)

Feature Engineering - Transform
Encoding Numerical Features

Simple Discretization: Binning

- Equal-width (distance) partitioning

  o Divides the range into N intervals of equal size: uniform grid

  o if A and B are the lowest and highest values of the attribute, the width of intervals will be: $W = (B - A)/N$.

  o The most straightforward, but outliers may dominate presentation

  o Skewed data is not handled well

- Equal-depth (frequency) partitioning

  o Divides the range into N intervals, each containing approximately same number of samples

  o Good data scaling

Feature Engineering - Transform

- Binarization maps a categorical attribute into one or more binary variables - One Hot/ Dummy Encoding

Car	Fuel
A	Gas
B	Diesel
C	Gas
D	gas

→

Car	Fuel		Gas	Diesel
A	Gas	....	1	0
B	Diesel	....	0	1
C	Gas	....	1	0
D	gas	....	1	0

- Categorical features to a numeric representation - Label Encoding

Fuel
Gas
Diesel
Gas
gas

→

Fuel
1
2
1
3

Self Practice

Case Study: California Housing Prices Prediction

Link to dataset:

https://www.kaggle.com/datasets/camnugent/california-housing-prices

Try to implement

- Data Quality and Cleaning
- Feature Construction and Transformation
- Scaling and Normalization
- Anything else that you can think of, to prepare the data and remove any issues from it.

ML Workflow

ML workflow

1. Should I use ML on this problem?
   - Is there a pattern to detect?
   - Can I solve it analytically?
   - Do I have data?
2. Gather and organize data.
3. Preprocessing, cleaning, visualizing.
4. Choosing a model, loss, regularization, ...
5. Optimization
6. Hyperparameter search
7. Analyze performance and mistakes, and iterate back to step 5 (or 3)

Example: Marks prediction

Data: survey, marks from previous years

Process the data

- training set; test set

- representation of input features; output

Choose form of model: linear regression

system's performance evaluation: objective function

optimize performance by setting appropriate parameters:
Optimization

Evaluate on test set: generalization

Few Terminologies
(To interpret the jargons in the prescribed text book)

Terminologies

Amount taken	Period	Credit Score	Defaulter
40 lakhs	5 years	1000	No
10 Lakhs	5 months	550	YES
80 Lakhs	3 years	950	No
20 Lakhs	4 years	1500	No

- **Training example.** An example of the form $\langle \mathbf{x}, f(\mathbf{x}) \rangle$.

- **Target function (target concept).** The true function f.

- **Hypothesis**. A proposed function h believed to be similar to f.

- **Concept**. A boolean function. Examples for which $f(\mathbf{x}) = 1$ are called **positive examples** or **positive instances** of the concept. Examples for which $f(\mathbf{x}) = 0$ are called **negative examples** or **negative instances**.

- **Classifier**. A discrete-valued function. The possible values $f(\mathbf{x}) \in \{1, \ldots, K\}$ are called the **classes** or **class labels**.

- **Hypothesis Space**. The space of all hypotheses that can, in principle, be output by a learning algorithm.

- **Version Space**. The space of all hypotheses in the hypothesis space that have not yet been ruled out by a training example.

Inductive Learning Hypothesis

Any hypothesis found to approximate the target function well over a sufficiently large set of training examples will also approximate the target function well over other unobserved examples

Inductive learning or "**Prediction**":
– **Given** examples of a function *(X, F(X))*
– **Predict** function *F(X)* for new examples *X*

- Classification
  F(X) = Discrete

- Regression
  F(X) = Continuous

- Probability estimation
  F(X) = Probability*(X):*

Inductive Learning Hypothesis

- Target Concept

- Discrete   : f(x) ∈ {Yes, No, Maybe}   Classification
                                                                Regression

- Continuous   : f(x) ∈ [20-100]

- Probability Estimation   : f(x) ∈ [0-1]

Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport?
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes

Inductive Learning Hypothesis

- Target Concept

- Discrete : $f(x) \in \{Yes, No, Maybe\}$ Classification

- Continuous : $f(x) \in [20-100]$ Regression

- Probability Estimation : $f(x) \in [0-1]$

Sky	AirTemp	Altitude	Wind	Water	Forecast	Humidity
Sunny	Warm	Normal	Strong	Warm	Same	60
Sunny	Warm	High	Strong	Warm	Same	75
Rainy	Cold	High	Strong	Warm	Change	70
Sunny	Warm	High	Strong	Cool	Change	45

62

Inductive Learning Hypothesis

- Target Concept

- Discrete $\quad\quad\quad$: $f(x) \in \{Yes, No, Maybe\}$

- Continuous $\quad\quad$: $f(x) \in [20\text{–}100]$ Classification

- Probability Estimation : $f(x) \in [0\text{–}1]$ Regression

Sky	AirTemp	Humidity	Wind	Water	Forecast	P(EnjoySport =Yes)
Sunny	Warm	Normal	Strong	Warm	Same	0.95
Sunny	Warm	High	Strong	Warm	Same	0.7
Rainy	Cold	High	Strong	Warm	Change	0.5
Sunny	Warm	High	Strong	Cool	Change	0.6

References

- Chapter 1 – Machine Learning, Tom Mitchell

- Chapter 1, 2 – Introduction to Machine Learning, 2nd edition, Ethem Alpaydin

- Chapter 1 - Pattern Recognition & Machine Learning Christopher M. Bhishop

End of
CS-2