CS4104 Machine Learning

Course Outline

Course Learning Outcomes (CLOs)

1. Understand the basic concepts and algorithms in machine learning (ML). Supervised learning, unsupervised learning, reinforcement learning.
2. Understand and Apply statistical models to solve problems in ML, with a focus on how the vector-space, Bayesian and decision tree models are implemented and applied to classification, regression and clustering.
3. ML models for combination in terms of ensembles, Boosting and stacking and ML algorithms' components including hypothesis, loss functions and optimization.
4. Evaluations of the ML algorithms with variance and bias trade-off.

zeshan.khan@nu.edu.pk

Topics

- Introduction, Supervised Learning, Unsupervised Learning, Semi-Supervised Learning, Reinforcement learning
- Data types for ML, Classification and Regression Algorithms
  - KNN
  - Baysian
  - SVM
  - Decision Tree
  - Linear Regression
  - Logistic Regression
  - Evaluations
- Clustering
  - K-Means Clustering
  - Hierarchical Clustering
- Feature Selections
  - Principal Component Analysis (PCA)
- Ensemble and Combination of ML Algorithms
- Neural Network (Back Propagation)

zeshan.khan@nu.edu.pk

Books

- Machine Learning: An Algorithmic Perspective, Stephen Marsland, Second Edition
- Introduction to Data Mining by Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar. Second Edition

Assessments

- Quizzes/Assignments (10 Marks)
  - 5 Programming Assignments
  - 5 Quizzes (short questions Non-Programming)
- Class Project (10 Marks)
  - Assessment will be in three phases
    - Proposal
    - Results prediction
    - Reporting and Visualizations
- Midterms (30 Marks)
  - Two midterms 15 marks each
- Final Exam (50 Marks)
  - One Exam of 50 marks from all course contents

zeshan.khan@nu.edu.pk

Course Instructor

- Zeshan Khan
- Office
  - Office # 32, Ground Floor, Academic Block
- Contacts
  - zeshan.khan@nu.edu.pk
  - http://zeshankhanalvi.com

CS4104 Machine Learning

An Introduction

Machine Learning

Field of study that gives computers ability to learn

- Machine learning is programming computers to optimize a performance criterion using example data or past experience.
- There is no need to "learn" to calculate payroll
- Learning is used when:
  - Human expertise does not exist (navigating on Mars),
  - Humans are unable to explain their expertise (speech recognition)
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted to particular cases (user biometrics)

zeshan.khan@nu.edu.pk

Learning

- Tom Mitchell (1998): A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with E

zeshan.khan@nu.edu.pk

Examples

- Task: Face Recognition
  - Experience: Images or Examples
  - Performance Measure: Accuracy

Applications

- Predict whether a patient, hospitalized due to heart attack, will have a second heart attack
- Predict the price of a stock in 6 months from now
- Pace detection: find faces in images (or indicate if a face is present)
- Spam filtering: identify email messages as spam or non-spam
- Fraud-detection applications that seek patterns in jumbo size data sets

zeshan.khan@nu.edu.pk

Commercial Viewpoint

- Lots of data is being collected and warehoused
  - Web data, e-commerce
  - purchases at department/grocery stores
  - Bank/Credit Card transactions

zeshan.khan@nu.edu.pk

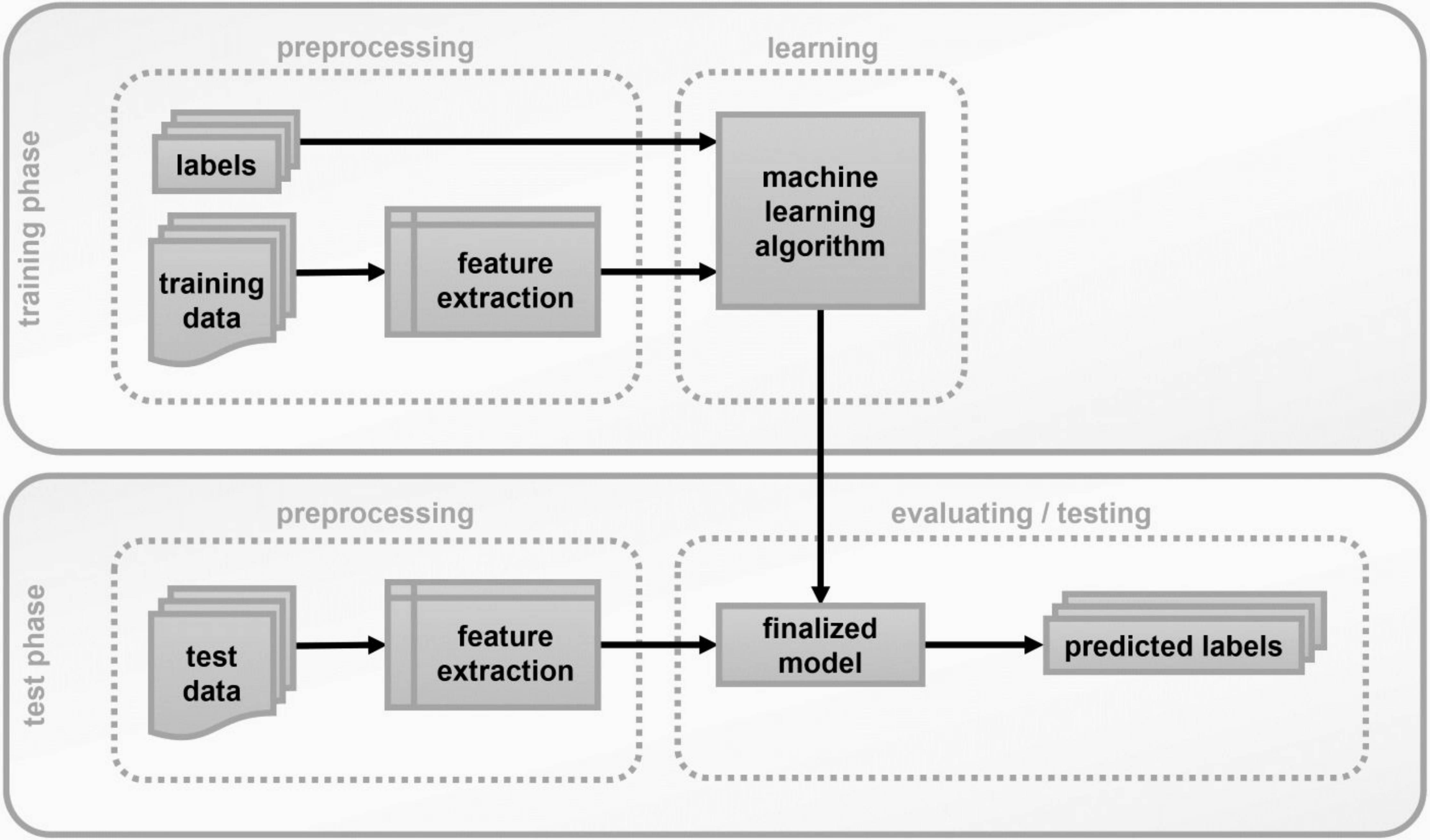
Machine Learning

- Supervised
  - Classification
  - Regression
- Un-Supervised
- Semi Supervised
- Reinforcement
- Association

zeshan.khan@nu.edu.pk

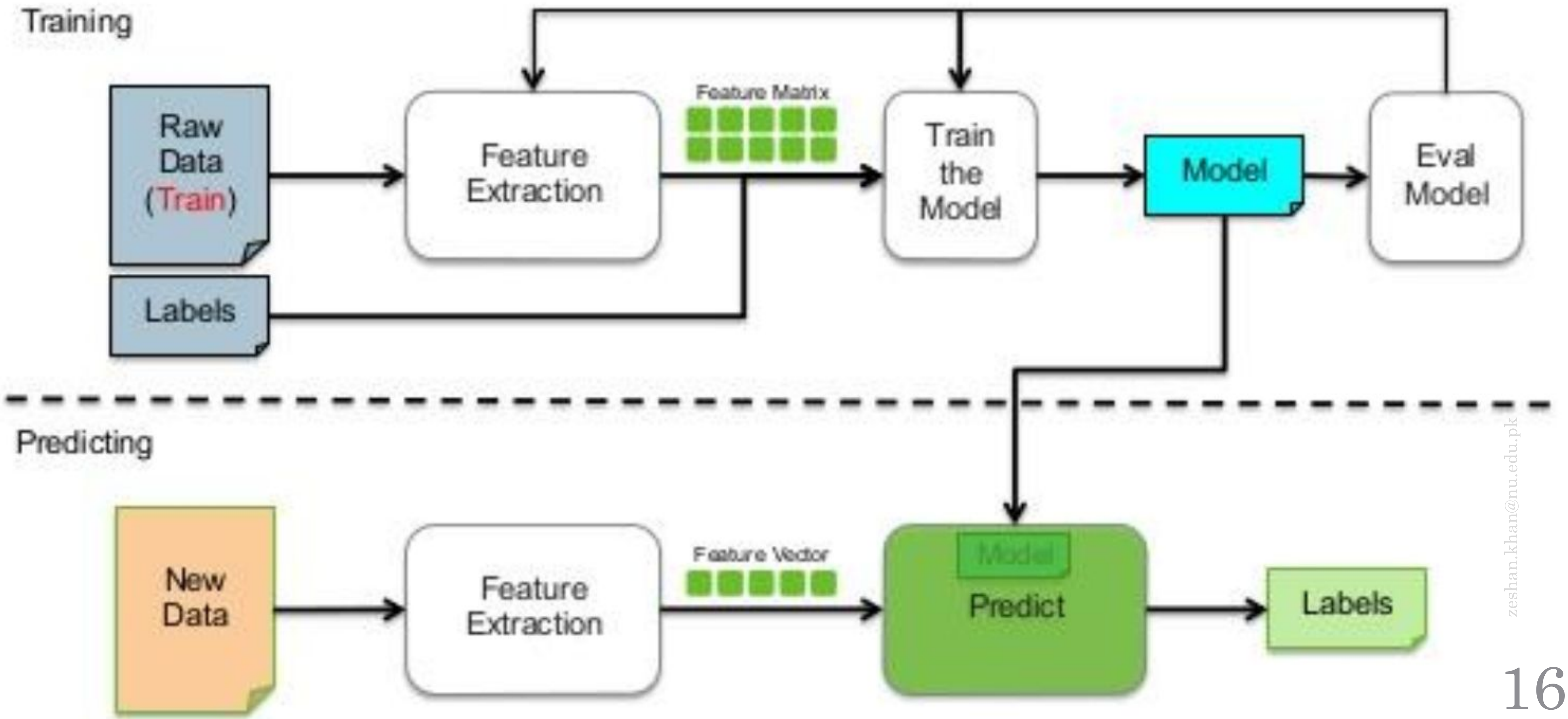
Supervised Learning

- Learn from Supervised Training Data
- For example; spam filtering where Large number of email messages labelled as either
  - Spam
  - non-spam
- New email message will then be classified as spam or non-spam



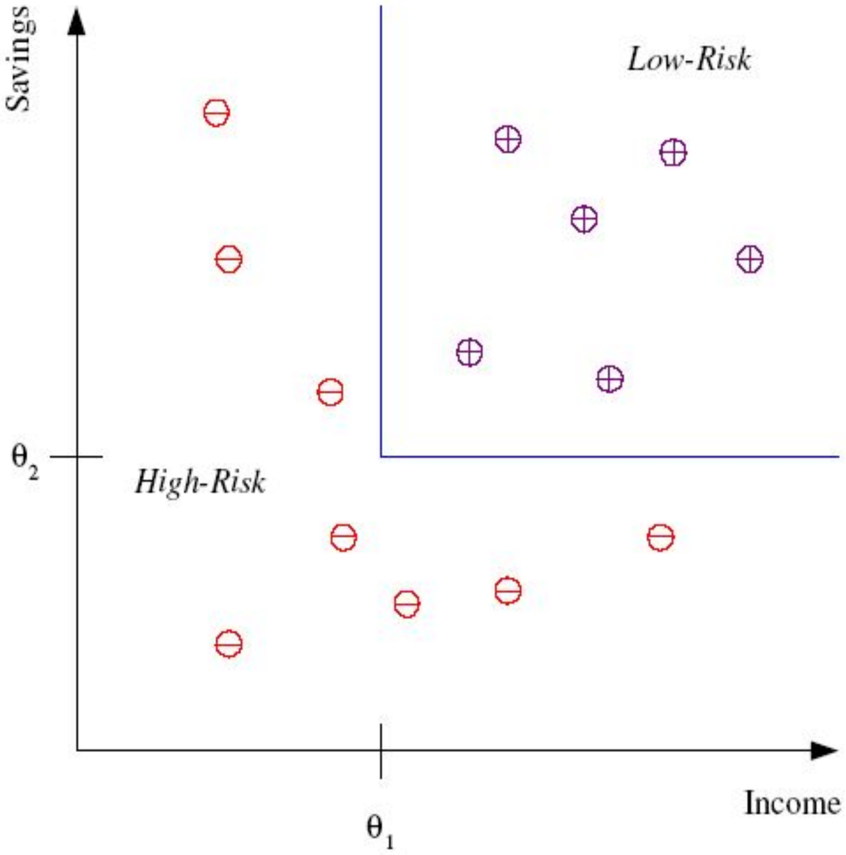
training phase
preprocessing
learning
labels
training data
feature extraction
machine learning algorithm
test phase
preprocessing
evaluating / testing
test data
feature extraction
finalized model
predicted labels

Supervised Learning Workflow



Classification

- Risk Analysis:
- The risk is higher for the values of income and saving less than the θs.
  - $(I > \theta_1 \wedge S > \theta_2)?\, L : H$
  - I: Income
  - S: Saving
  - L: Low-Risk
  - H: High-Risk



zeshan.khan@nu.edu.pk

Applications

- Pattern recognition
- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
- Medical diagnosis: From symptoms to illnesses
- Biometrics: Recognition/authentication using physical and/or behavioral characteristics: Face, iris, signature, etc
- Outlier/novelty detection

zeshan.khan@nu.edu.pk

Regression

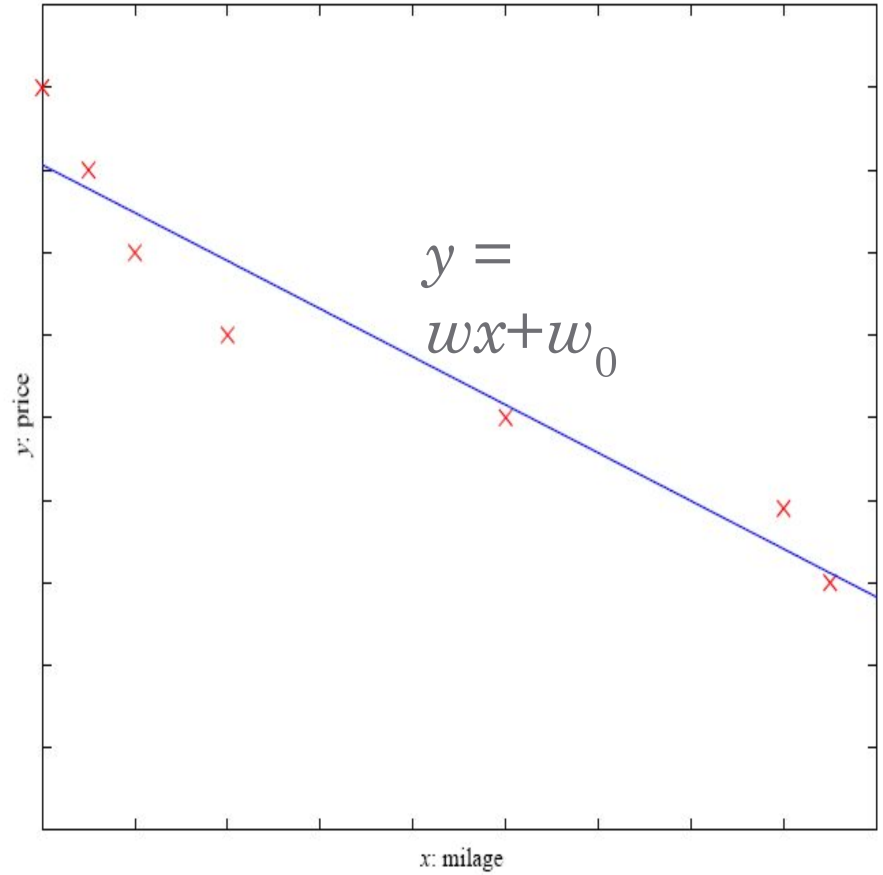
- Example: Price of a used car
- x : car attributes

  y : price

  $y = g(x \mid \theta)$

  $g(\,)$ model,

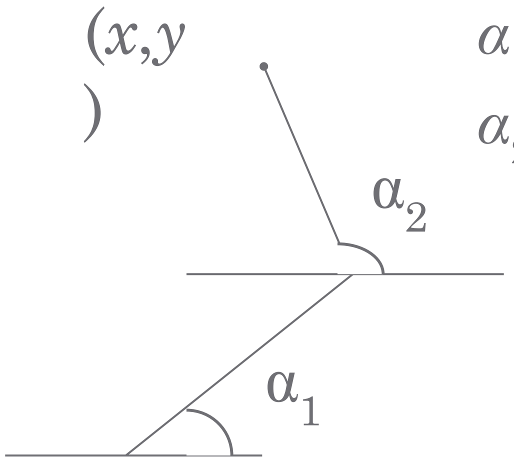
  θ parameters



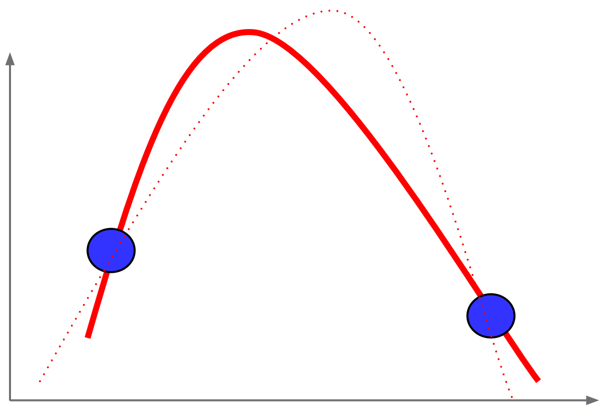
zeshan.khan@nu.edu.pk

Regression Applications

- Navigating a car: Angle of the steering
- Kinematics of a robot arm



$$\alpha_1 = g_1(x,y)$$
$$\alpha_2 = g_2(x,y)$$



- Response surface design

Supervised Learning: Uses

- Prediction of future cases: Use the rule to predict the output for future inputs
- Knowledge extraction: The rule is easy to understand
- Compression: The rule is simpler than the data it explains
- Outlier detection: Exceptions that are not covered by the rule, e.g., fraud

Data Types

- Categorical
  - A variable that can take on one of a limited number of possible values. E.g. Grades A to F, Olympic games (GameA, GameB etc.)
- Numerical
  - A variable that can take a real value. E.g. 1, 2, 3, 4.
- Some more data types…

zeshan.khan@nu.edu.pk

Classification Vs Regression

Classification

- $Y = f(x): x \in X \text{ and } Y \in \{c_1, c_2, c_3, \dots, c_n\}$
- Where X is a vector representation of features.
- Y is a class label from a finite set of labels.

Regression

- $Y = f(x): x \in X \text{ and } Y \in R$
- Where X is a vector representation of features.
- Y is a regression value which can be any real value.

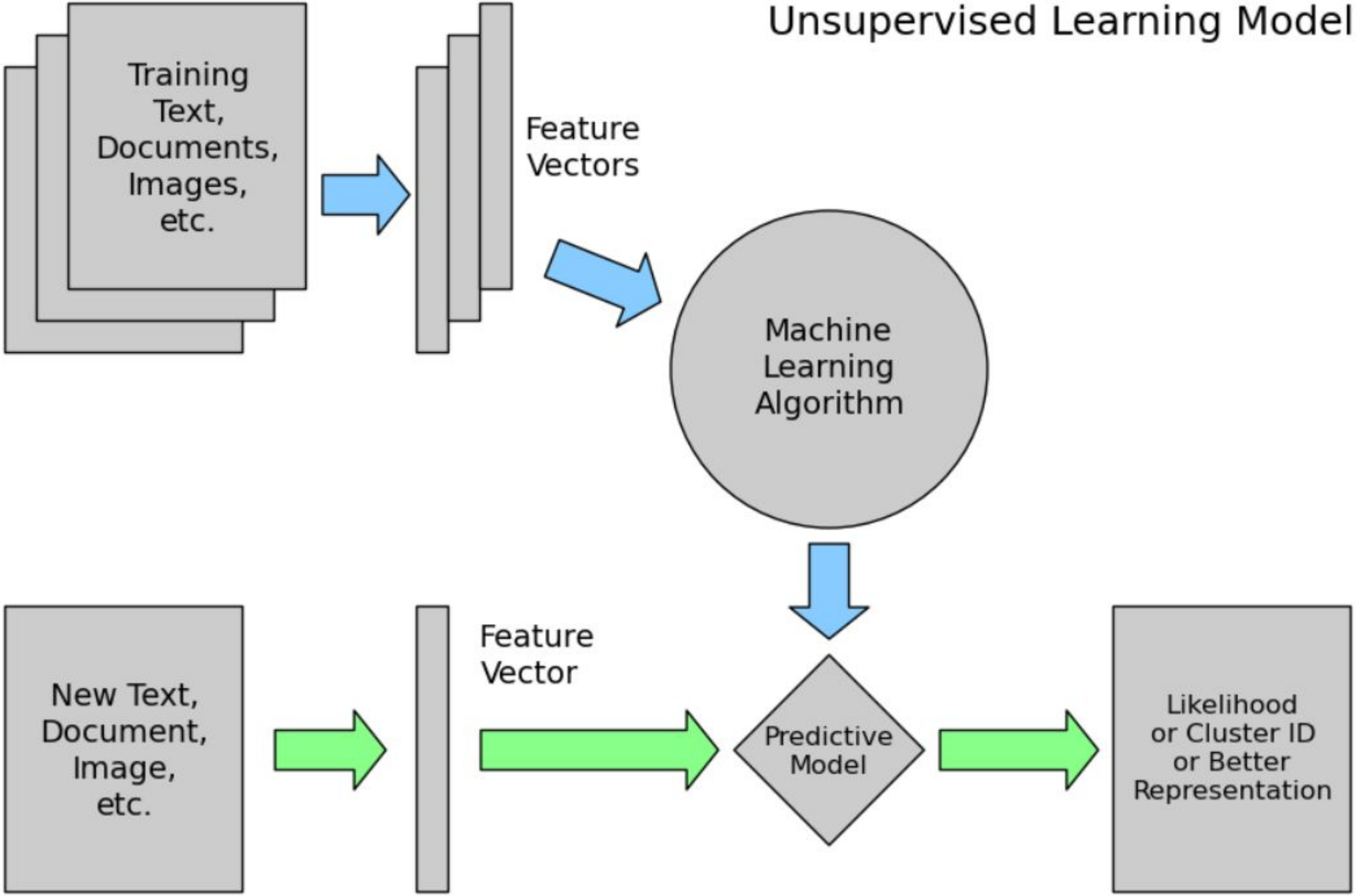
zeshan.khan@nu.edu.pk

Unsupervised Learning

- The correct classes of training data are not known
- Applications
  - Fraud Detection: Identify groups of motor insurance policy holders with a high average claim cost
  - Social Networks: Recognize communities within large groups of people

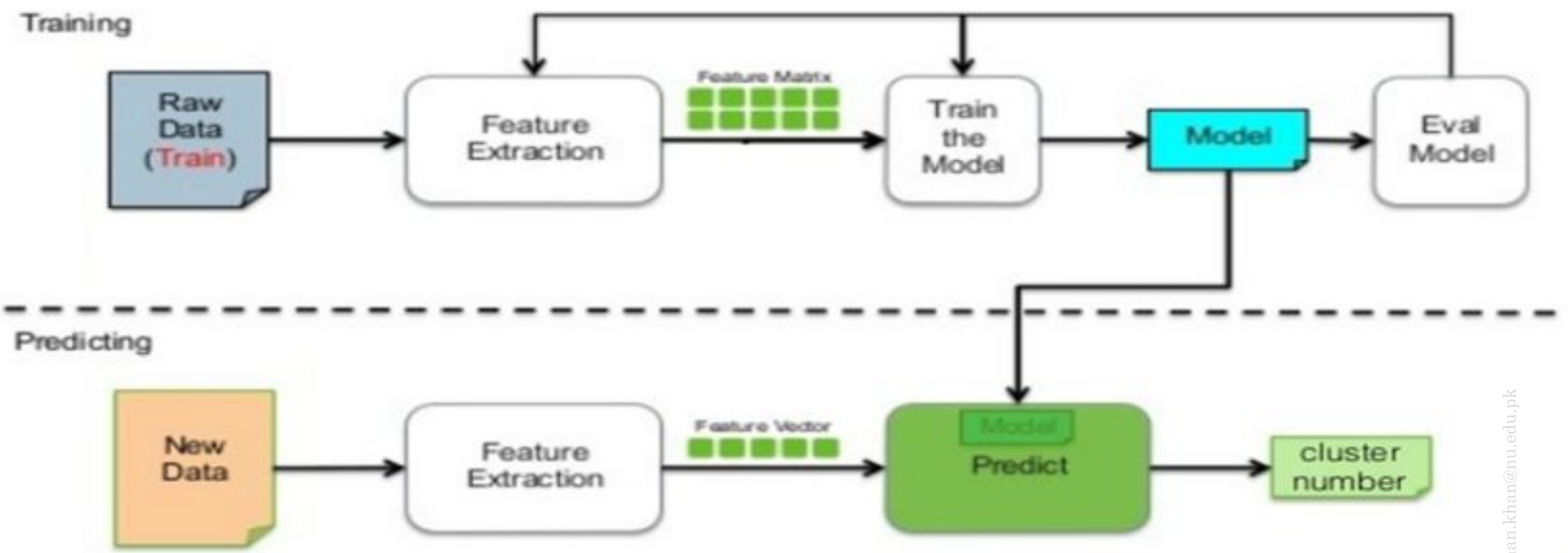
zeshan.khan@nu.edu.pk

Unsupervised Learning Model

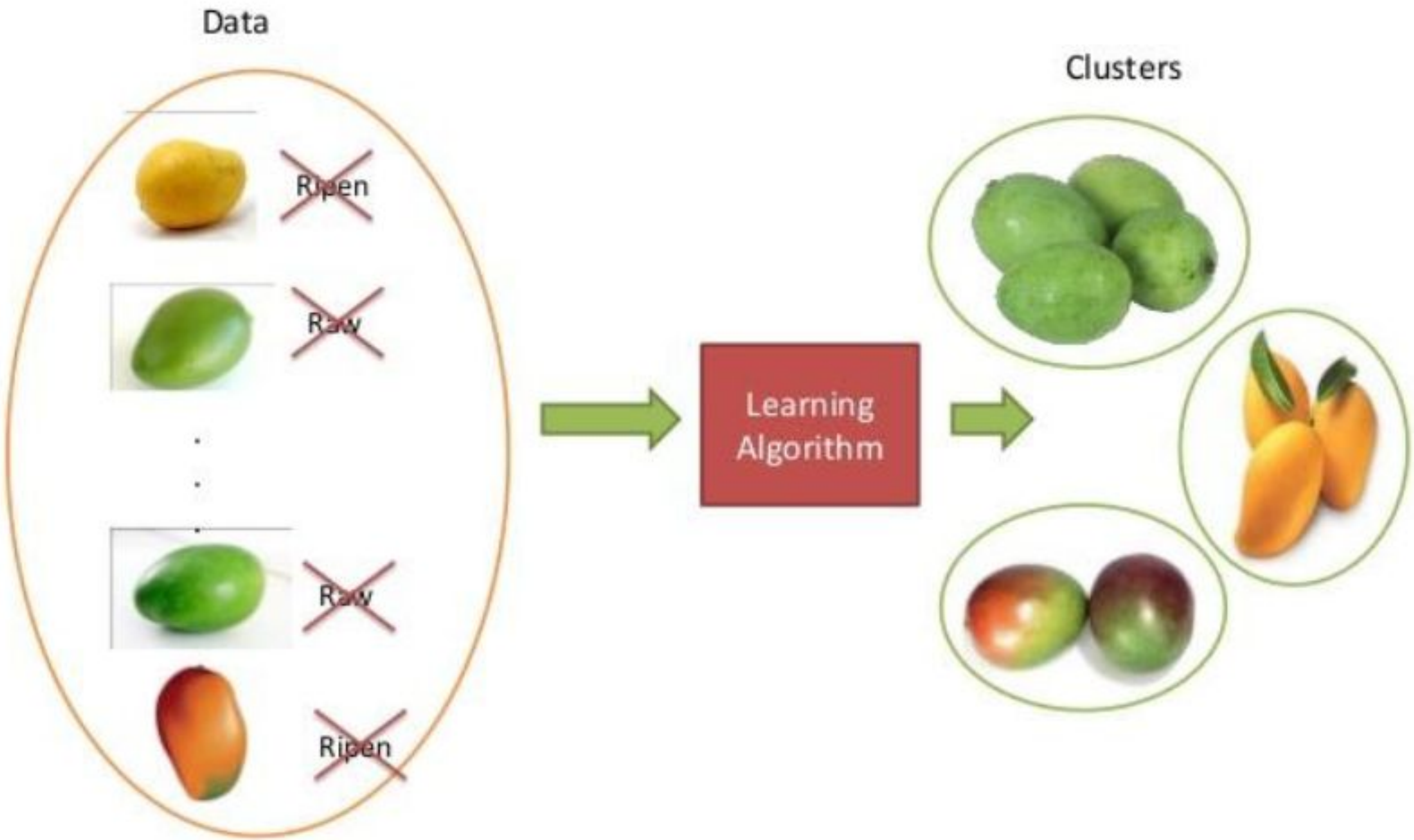


zeshan.khan@nu.edu.pk

Unsupervised Learning Workflow

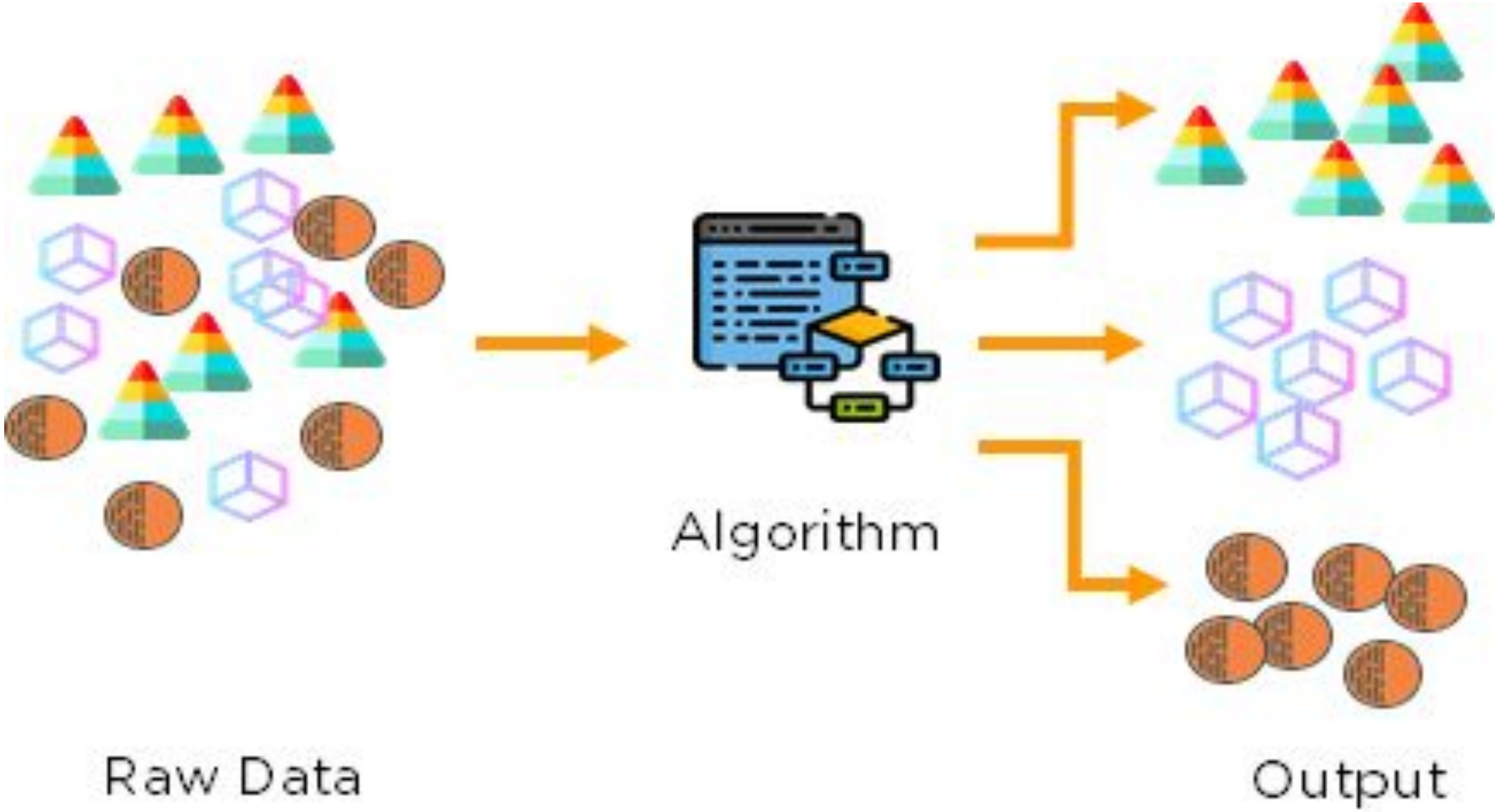


Example



zeshan.khan@nu.edu.pk

Example…



zeshan.khan@nu.edu.pk

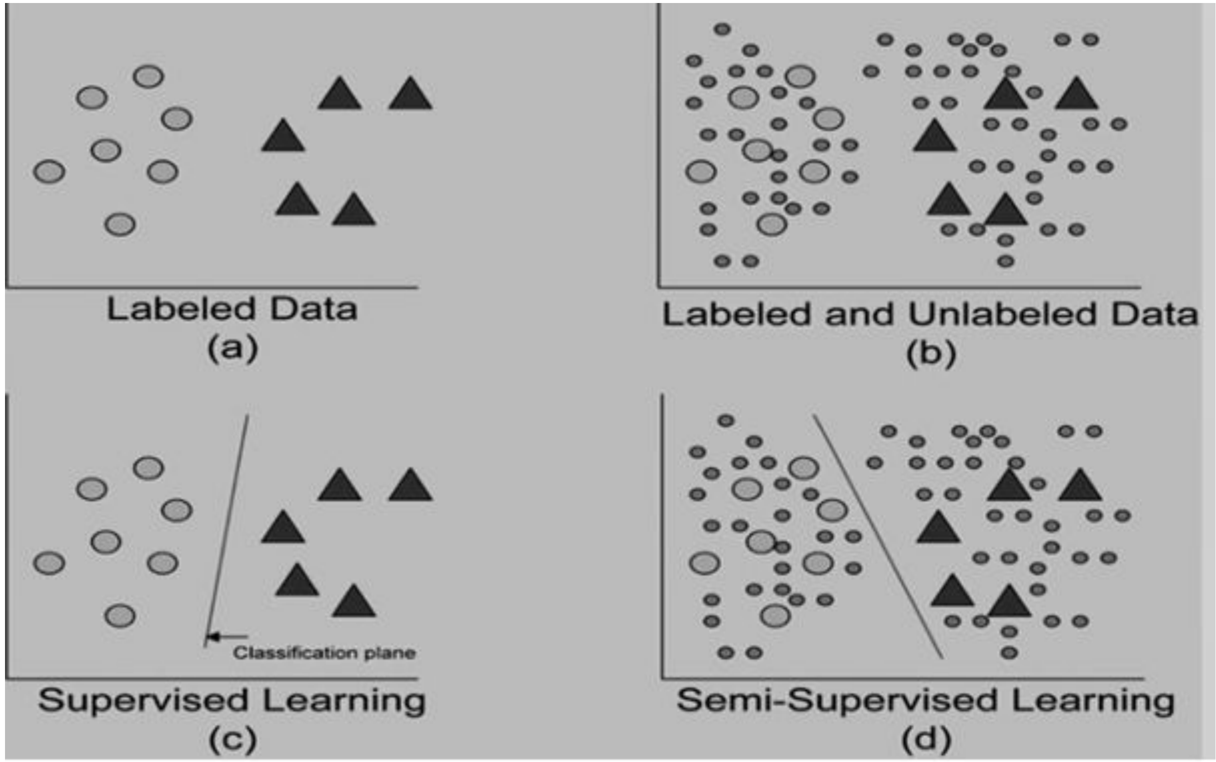
Unsupervised Learning

- Learning "what normally happens"
- No output
- Clustering: Grouping similar instances
- Example applications
  - Customer segmentation in CRM
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs

zeshan.khan@nu.edu.pk

Semi-Supervised Learning

- Learning from a small amount of labeled data and a large amount of un-labeled data.



Reinforcement Learning

- Learning a policy: A sequence of outputs
- No supervised output but delayed reward
- Credit assignment problem
- Game playing
- Robot in a maze
- Multiple agents, partial observability, ...

zeshan.khan@nu.edu.pk

Association

- Basket analysis:
  - P (Y | X) probability that somebody who buys X also buys Y where X and Y are products/services.
  - Example: P (chips | cold drink) = 0.7

zeshan.khan@nu.edu.pk

CS4104 Machine Learning

K Nearest Neighbors Classifier

zeshan.khan@nu.edu.pk

Instance Based Learning

- First Example of Supervised Classification
- Rote-learner
  - Memorizes entire training data and performs classification only if attributes of record match one of the training examples exactly
- Nearest neighbor
  - Uses k “closest” points (nearest neighbors) for performing classification

Instance Based Learning

Labeled Data

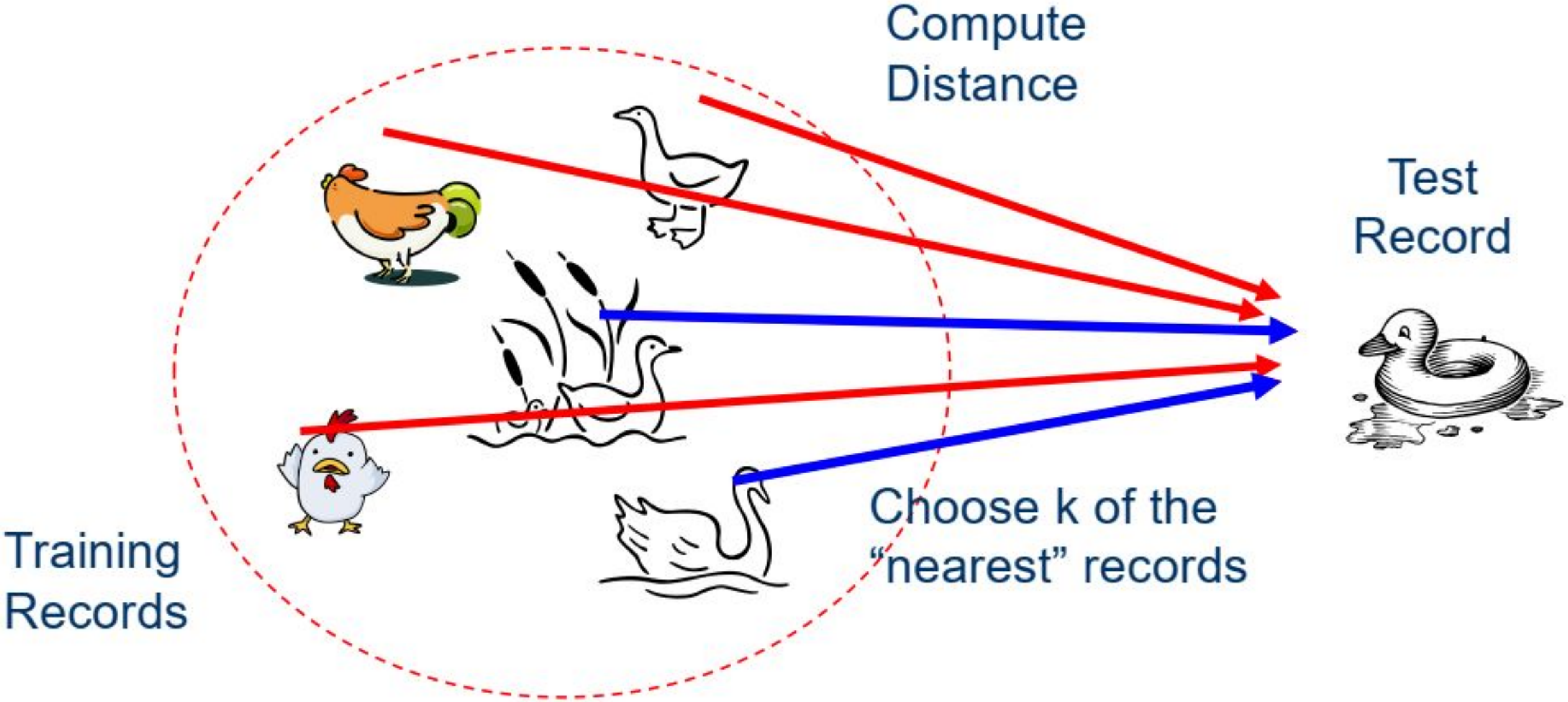
Att1	Att2	Class
1	2	A
5	7	B
2	5	A
4	2	B

Unlabeled Data

Att1	Att2	Class
1	2	?
2	6	?
3	4	?

zeshan.khan@nu.edu.pk

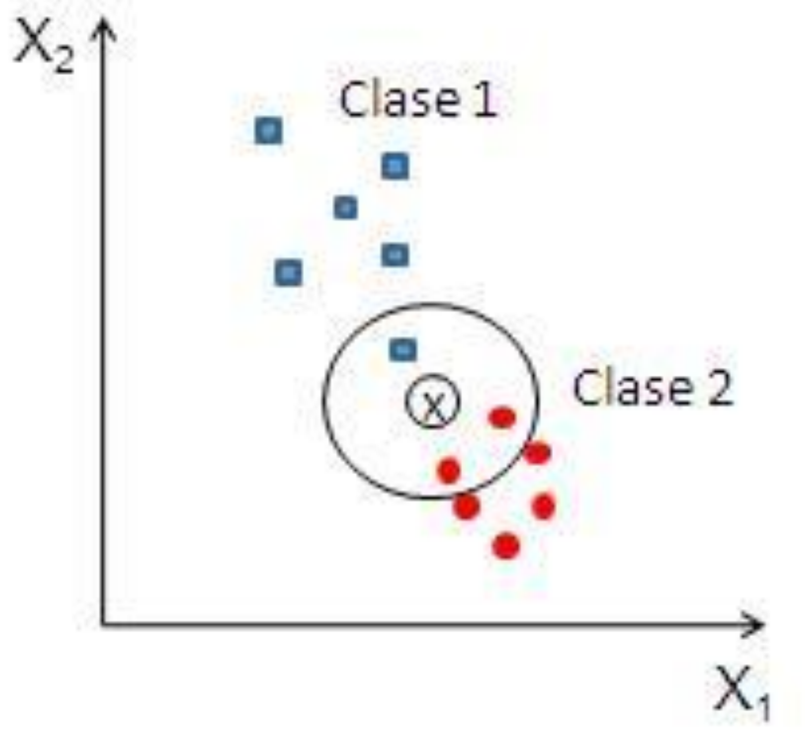
Nearest Neighbors



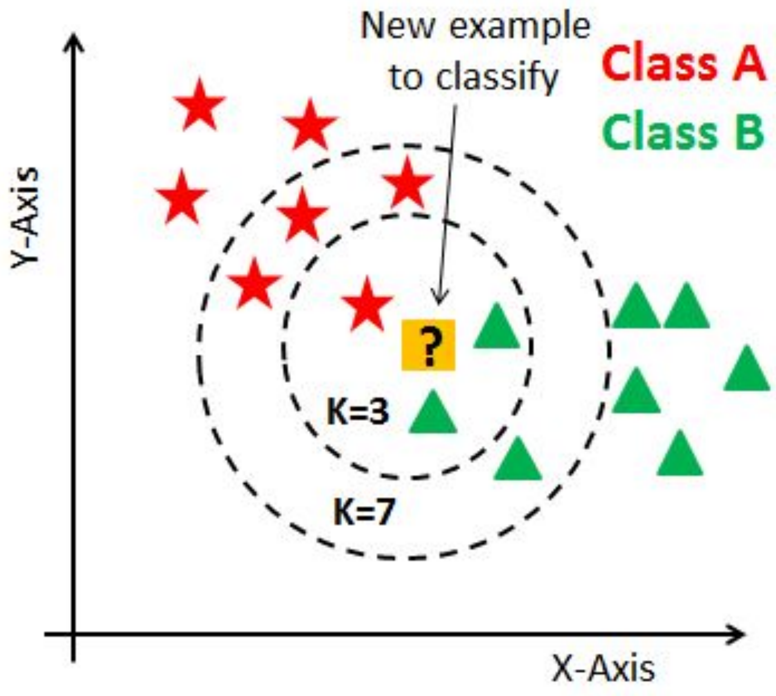
K Nearest Neighbors

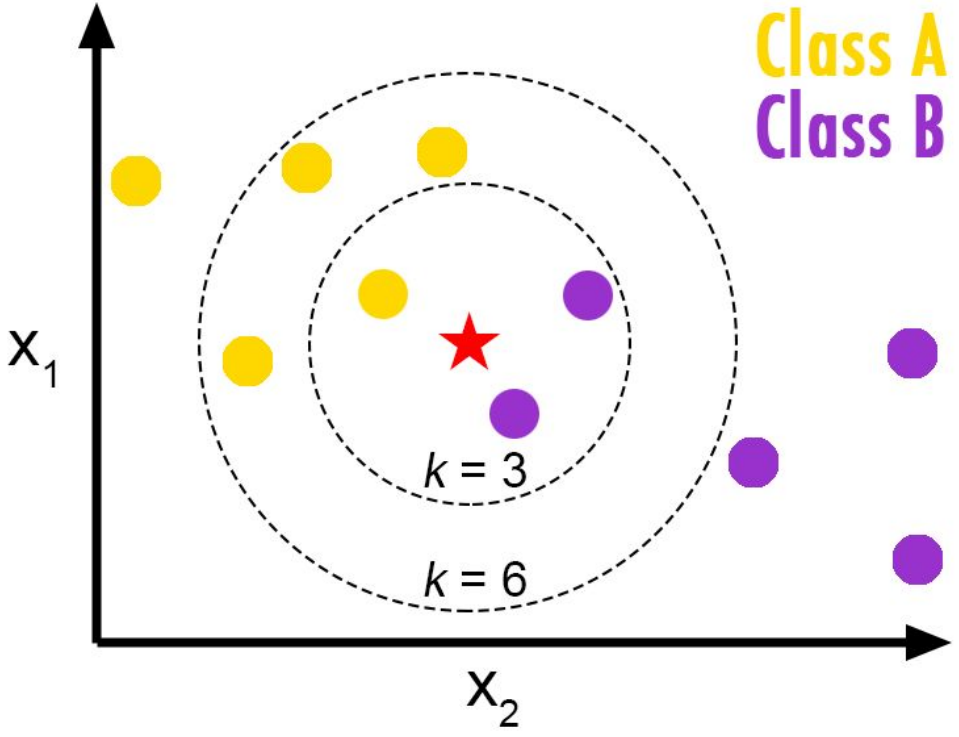
- Requires three things
  - The set of stored records
  - Distance Metric to compute distance between records
  - The value of k, the number of nearest neighbors to retrieve

- To classify an unknown record:
  1. Compute distance to other training records
  2. Identify k nearest neighbors
  3. Use class labels of nearest neighbors to determine the class label of unknown record (e.g., by taking majority vote)



K Nearest Neighbors (KNN)





K Nearest Neighbors

- 1. Compute distance to other training records
    1. $d(X1, X2) = \sqrt{\sum_{i=0}^{dim} (X1_i - X2_i)^2}$
    2. $d(X1, X2) = \sum_{i=0}^{dim} |(X1_i - X2_i)|$
  2. Identify k nearest neighbors
  3. Use class labels of nearest neighbors to determine the class label of unknown record (e.g., by taking majority vote)

Example

X1	X2	Class
1	3	B
2	4	B
3	2	A
5	4	A
2	5	?

- Assuming Distance as city block distance

zeshan.khan@nu.edu.pk

Example

X1	X2	Class	Distance
1	3	B	|2-1|+|5-3|=3
2	4	B	|2-2|+|5-4|=1
3	2	A	|2-3|+|5-2|=4
5	4	A	|2-5|+|5-4|=4
2	5	?	

1. **Compute distance to other training records**
2. Identify k nearest neighbors
3. Use class labels of nearest neighbors to determine the class label of unknown record (e.g., by taking majority vote)

zeshan.khan@nu.edu.pk

Example (k=1)

X1	X2	Class	Distance
1	3	B	|2-1|+|5-3|=3
2	4	B	|2-2|+|5-4|=1
3	2	A	|2-3|+|5-2|=4
5	4	A	|2-5|+|5-4|=4
2	5	?	

1. Compute distance to other training records
2. Identify k nearest neighbors
3. Use class labels of nearest neighbors to determine the class label of unknown record (e.g., by taking majority vote)

Example (k=2)

X1	X2	Class	Distance
1	3	B	|2-1|+|5-3|=3
2	4	B	|2-2|+|5-4|=1
3	2	A	|2-3|+|5-2|=4
5	4	A	|2-5|+|5-4|=4
2	5	?	

1. Compute distance to other training records
2. Identify k nearest neighbors
3. Use class labels of nearest neighbors to determine the class label of unknown record (e.g., by taking majority vote)

zeshan.khan@nu.edu.pk

Example (k=1,2)

X1	X2	Class	Distance
1	3	B	|2-1|+|5-3|=3
2	4	B	|2-2|+|5-4|=1
3	2	A	|2-3|+|5-2|=4
5	4	A	|2-5|+|5-4|=4
2	5	B	

1. Compute distance to other training records
2. Identify k nearest neighbors
3. Use class labels of nearest neighbors to determine the class label of unknown record (e.g., by taking majority vote)

zeshan.khan@nu.edu.pk

Issues in KNN

- Scaling Issues (Data Normalization)
- Attributes may have to be scaled to prevent distance measures from being dominated by one of the attributes. E.g. The Salary and GPA have different ranges.

zeshan.khan@nu.edu.pk

KNN Pseudo Code

- Constructor
- Train
- Test