17th Behavioral Modelling Summer School
14-16 September, 2018

Application of Machine Learning in Travel Behavior Analysis

Muhammad Awais Shafique
Assistant Professor
University of Central Punjab, Pakistan

What is Machine Learning?

"Field of study that gives computers the ability to learn without being explicitly programmed."

Arthur Samuel (1959)

"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 experience E."

Tom Mitchell (1998)

Human Learning

Diagnose Tumor
Malignant or Benign

- X-ray
- Ultrasound
- Biopsy
- MRI

Tests

- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape etc.

Observations

Machine Learning

"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 experience E."

- Task T: Determining whether Tumor is Malignant or Benign
- Experience E: Medical records of Tumor Diagnosis
- Performance Measure P: Number of correct Detections

Types of Learning Algorithms

- Supervised learning
  - Teach the computer and then let it use the learning to do the task (labelled data).
- Unsupervised learning
  - Let the computer determine structure and patterns in the data (unlabeled data).
- Semi-supervised learning
  - Let the computer determine patterns and then teach the computer to do the intended task (unlabeled and labelled data).
- Reinforcement learning
  - Let the computer determine the ideal behavior in an environment to maximize its performance.

Problems with Data Collection

Travel Information

- Starting time
- Ending time
- Route taken
- Mode used
- Accompanying persons
- Trip purpose

Traditional Travel Surveys

- Face-to-face interviews
- Paper questionnaires
- Telephone surveys
- Computer-assisted surveys

Shortcomings reported by Brog (1982), Richardson (1996), Zmud (1997), Hato (2010) etc.

- Time-consuming
- Laborious
- Dependence on memory
- Non-reporting of short trips
- Biased responses
- Variable perception of time

Automatic Detection of Travel Information

Evolution of Collection Methods

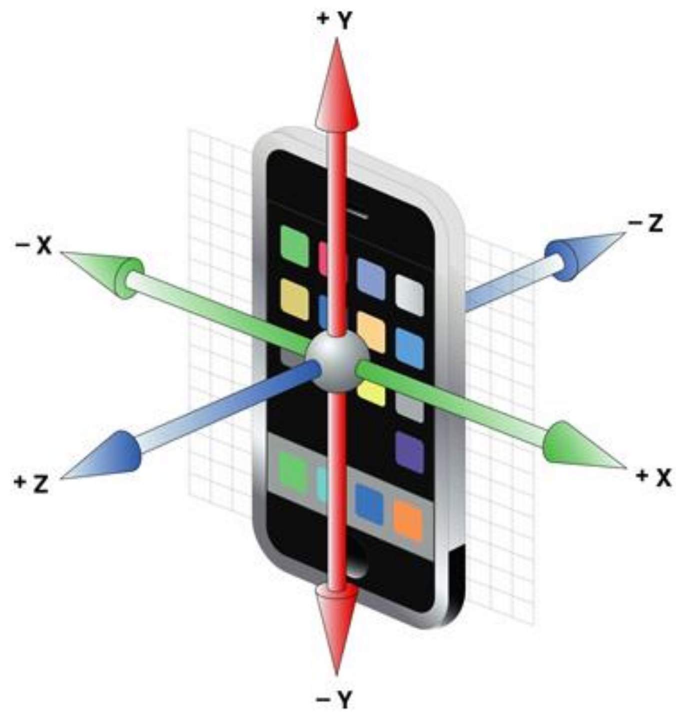
Data collection method	Advantages	Disadvantages
Questionnaire (paper-based, internet-based, telephone surveys) **Forrest and Pearson (2005)**	• Economy • Uniformity • Standardization	• Biased responses • Inaccurate perception of time • Small percentage of returned forms • Dependence on memory
Toll Tag Readers **Ferman et al. (2005)** 	• Automatic collection of average travel time and speed	• Limited only to tagged vehicles. • Limited only to tolled roads • Detailed O-D information unavailable • Cost of sensors
Bluetooth **Puckett and Vickich (2010)** 	• Automatic collection of average travel time and speed	• Limited only to Bluetooth enabled devices. • Limited only to roads • Detailed O-D information unavailable • Cost of sensors • Stops are unnoticed

Evolution of Collection Methods

Data collection method	Advantages	Disadvantages
Wireless Location Technology **Fontaine et al. (2007)** 	• Cell phones can be anonymously located when an action is taken (e.g. phone call or sms)	• Tracking cannot be done at regular intervals • Biased sample due to limited number of participating mobile companies • Stops are unnoticed
Smartphones (GPS + accelerometer) **Reddy et al. (2010)** 	• GPS enabled cell phones tracked at regular intervals • Detailed movement pattern recorded • Precise acceleration data along three axes available at regular intervals • Detailed O-D information available	• Privacy of the mobile user is an issue due to the data available at micro level

Lets Start from the Beginning

- Equipped with multiple sensors
  - GPS
  - Accelerometer
  - Gyroscope
  - Barometer etc.



Accelerometer Data

User ID	Trip ID	Date	Time	Acceleration		
				X axis	Y axis	Z axis
gf001	1	12/25/2010	11:33:39 AM	0.96	5.32	8.43
gf001	1	12/25/2010	11:33:44 AM	0.69	5.98	7.78
gf001	1	12/25/2010	11:33:47 AM	0.73	6.21	7.16
gf001	1	12/25/2010	11:33:49 AM	0.84	6.4	6.63
gf001	1	12/25/2010	11:33:54 AM	0.46	6.47	7.05
gf001	1	12/25/2010	11:33:59 AM	1.11	7.01	7.24
gf001	1	12/25/2010	11:34:05 AM	1.15	7.32	6.78
gf001	1	12/25/2010	11:34:10 AM	1.3	7.51	6.47
gf001	1	12/25/2010	11:34:15 AM	1.26	7.66	5.86
gf001	1	12/25/2010	11:34:20 AM	1.03	7.85	5.13
gf001	1	12/25/2010	11:34:22 AM	0.73	7.89	5.67
gf001	1	12/25/2010	11:34:25 AM	0.69	8.12	5.21

Trip Data

User ID	Trip ID	Departure date	Departure time	Arrival date	Arrival time	Mode
gf001	1	12/25/2010	11:33:00 AM	12/25/2010	12:02:00 PM	Walk
gf001	2	12/25/2010	12:07:00 PM	12/25/2010	12:32:00 PM	Walk
gf001	3	12/25/2010	4:03:00 PM	12/25/2010	4:12:00 PM	Car
gf001	4	12/25/2010	9:56:00 PM	12/25/2010	10:02:00 PM	Walk
gf001	5	12/25/2010	10:02:00 PM	12/25/2010	10:09:00 PM	Car

Data Cleaning

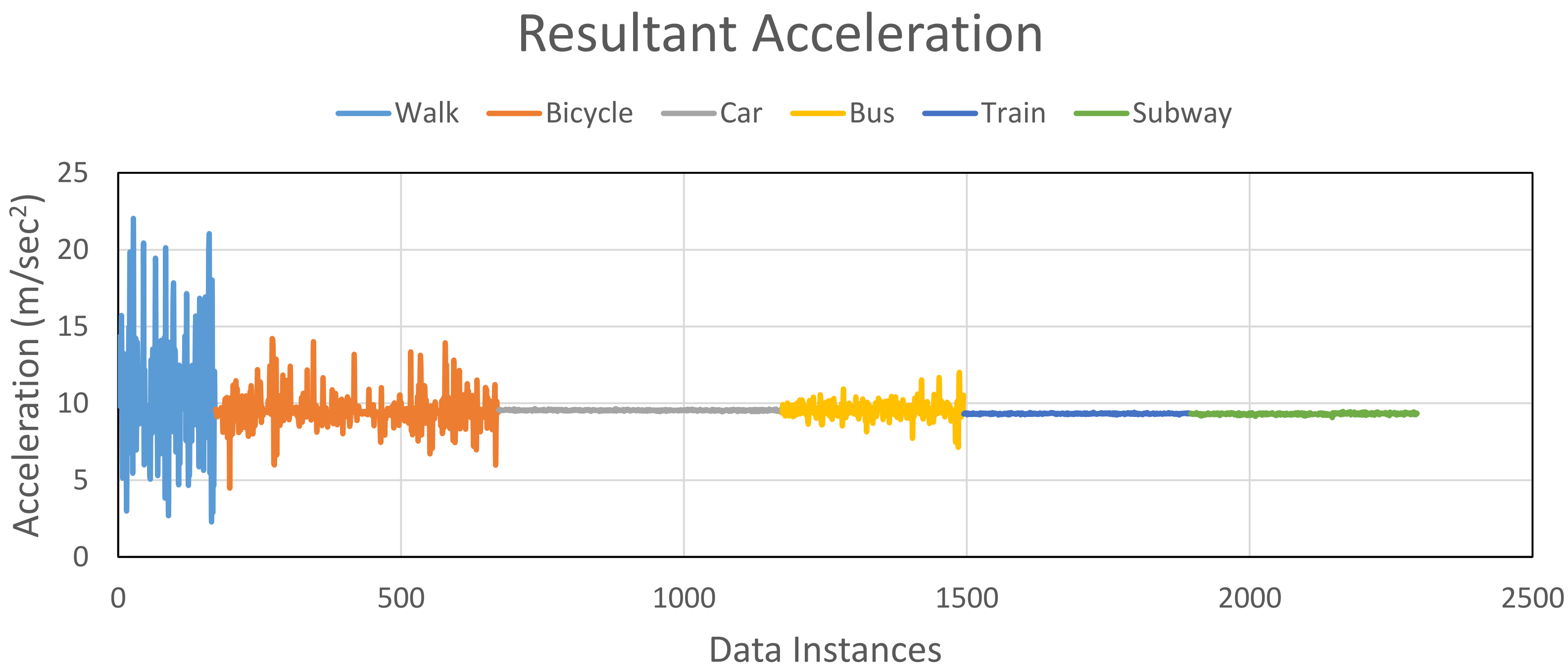
- Get rid of duplicate data
- Clean the outliers
- Keep only relevant data
- Scan data for abnormalities
    - 70 minutes long trip marked as “Walk”

Feature Engineering

- Feature engineering is the process of using domain knowledge of the data to create features that make machine learning algorithms work.
- If done correctly, it increases the predictive power of machine learning algorithms by creating features from raw data that help facilitate the machine learning process.

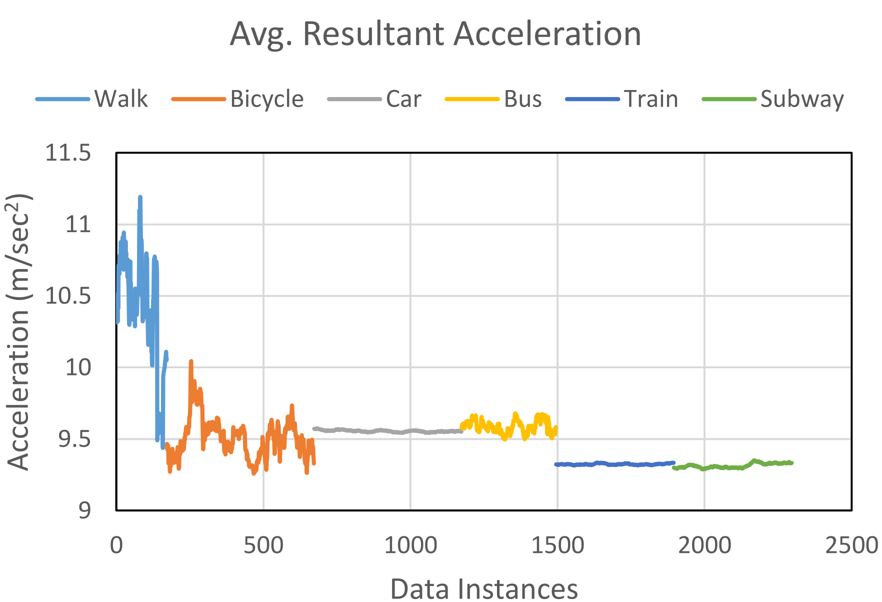
Feature Extraction

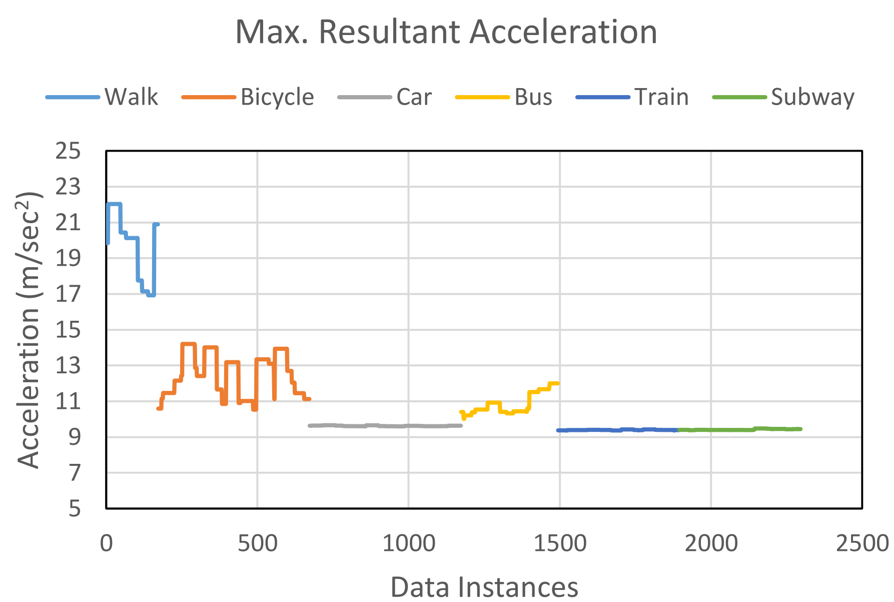
- Acceleration data can be recorded by using accelerometer.
- For a single data instance, it is difficult to identify the mode.



Feature Engineering

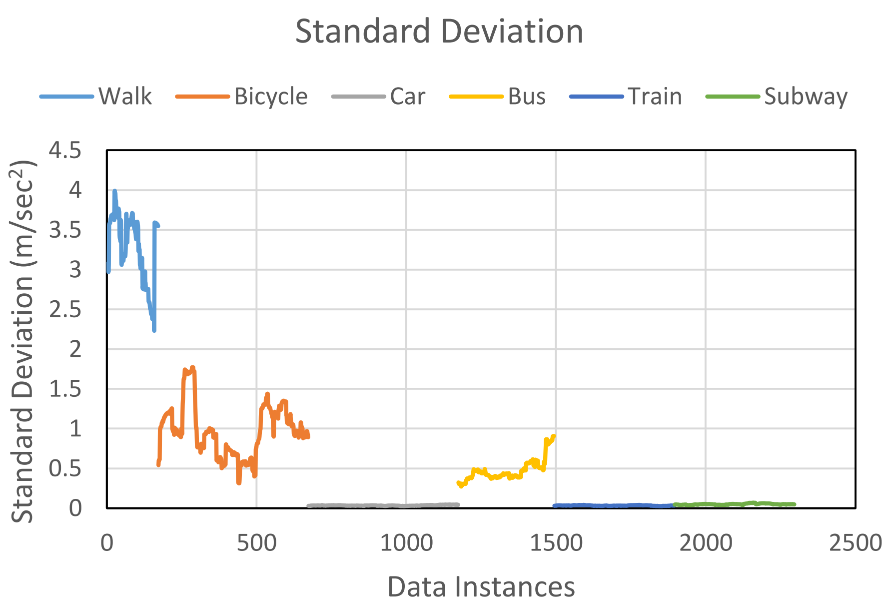
- But extracted features can assist in the identification.

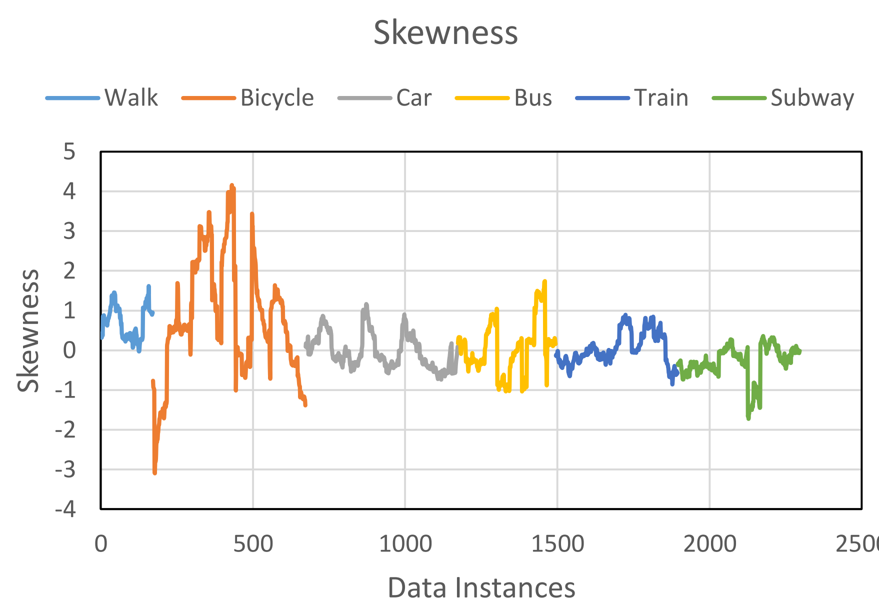




Feature Engineering

- Additional features can also be extracted.
- No single feature can distinguish among all modes.





Feature Engineering

Time Domain

- Resultant Acceleration (m/sec^2)
- Maximum Resultant Acceleration (m/sec^2)
- Average Resultant Acceleration (m/sec^2)
- Standard Deviation (m/sec^2)
- Skewness (measure of symmetry)
- Kurtosis (measure of flatness)

Frequency Domain

- Wavelet Transformations
- Fourier Transformations

Discrete Domain

- Symbolic String Representations

Training & Testing Data

- Training Data: Part of data used to train algorithm
- Testing Data: Part of data used to test an already trained algorithm

Study	**Percentage of data used for training**
Nick et al. (2010)	90
Tragopoulou et al. (2014)	80
Lester et al. (2006), Nitsche et al. (2012)	75
Nham et al. (2008)	70
Abdulazim et al. (2013)	65
Figo et al. (2010)	50

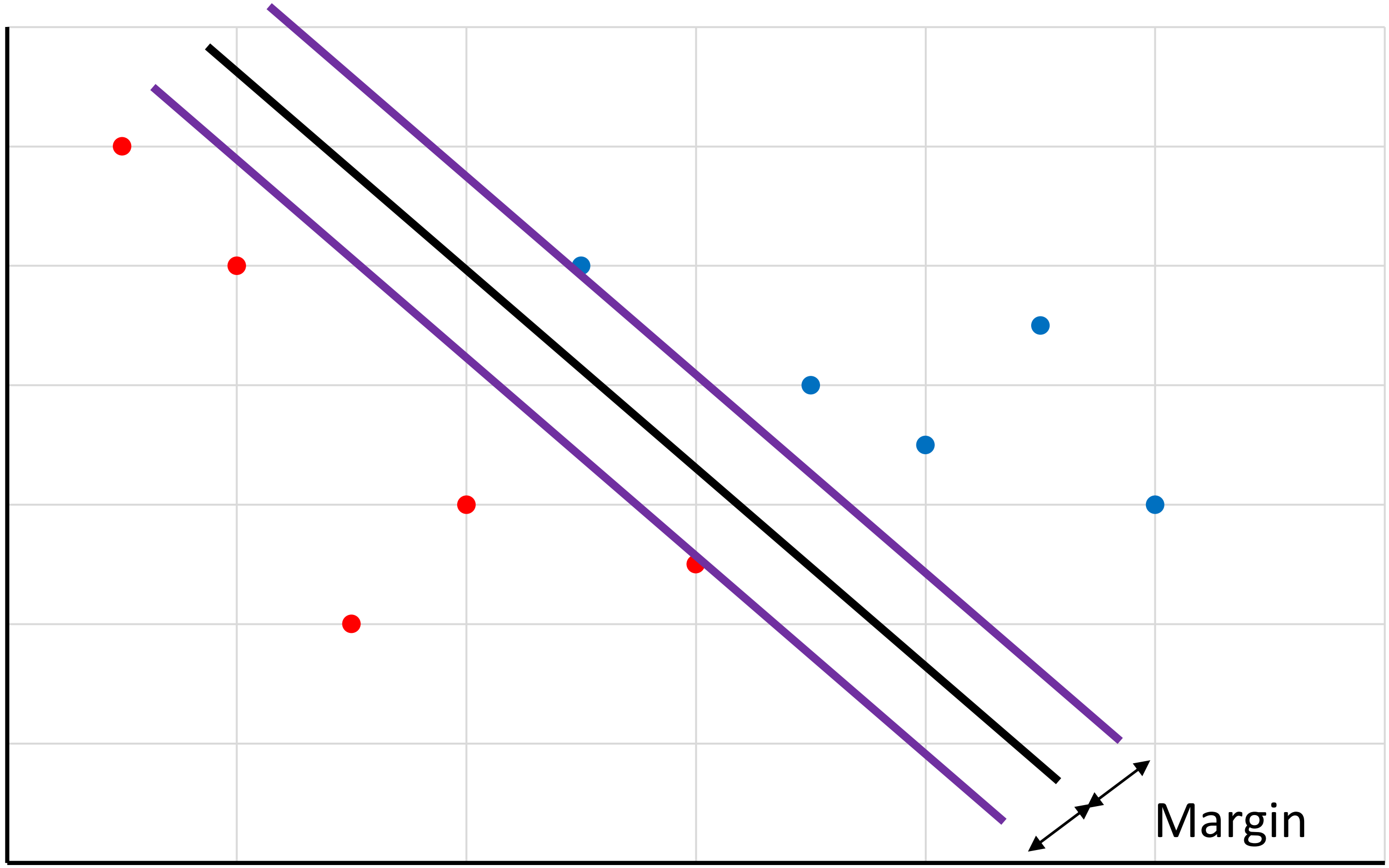
Classification Algorithms

- Support Vector Machines (SVM)
- Decision Tree (DT)
- Random Forest (RF)

Support Vector Machines

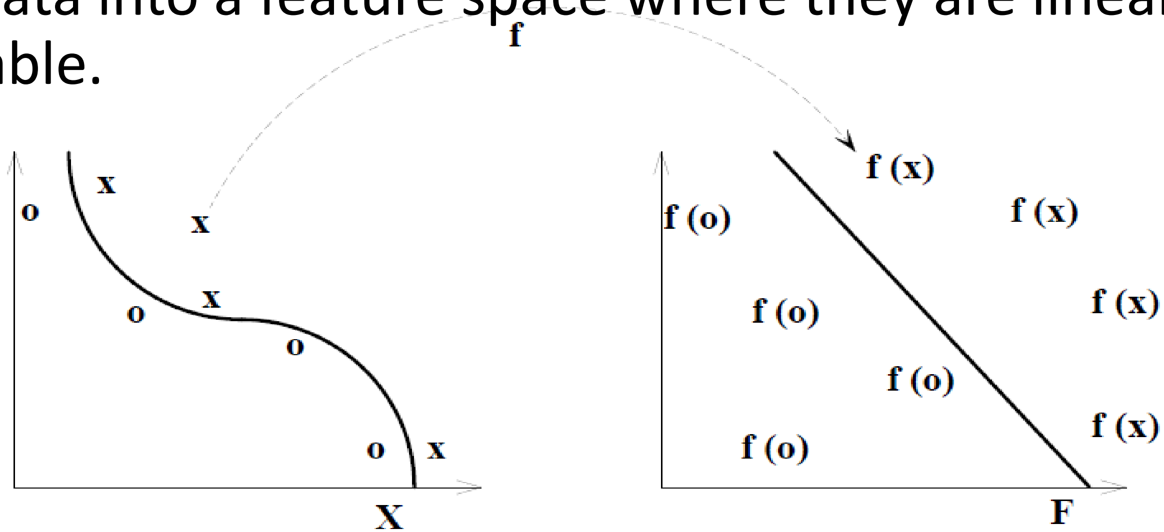
- Among the best "off-the-shelf" supervised learning algorithms.
- Off-the-shelf: A method that can be applied directly to data without requiring a great deal of time-consuming data preprocessing or careful tuning of the learning procedure.
- SVM is a two-class classifier which forms a separating hyperplane.
- When a set of training data containing class labels is supplied to SVM, it outputs an optimal hyperplane which then classifies new examples.

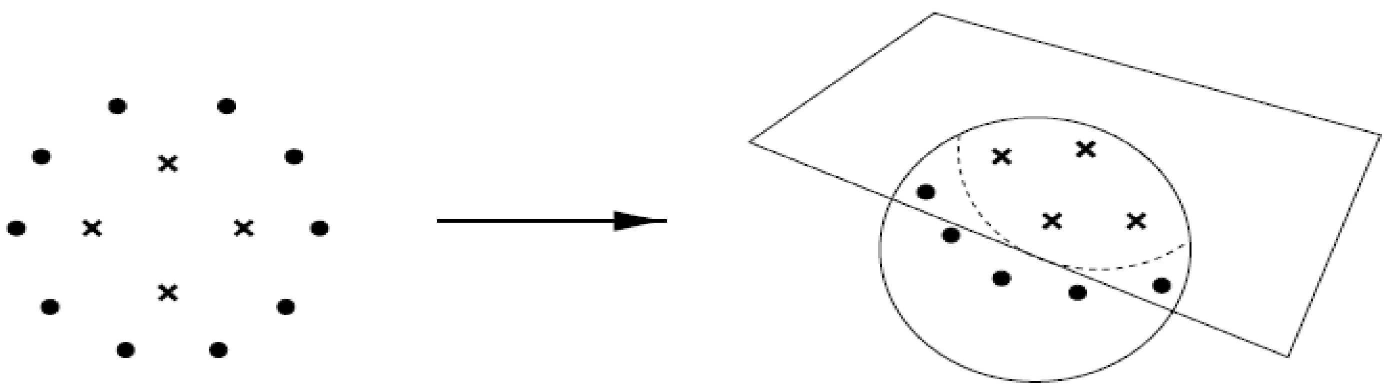
Support Vector Machines



Support Vector Machines

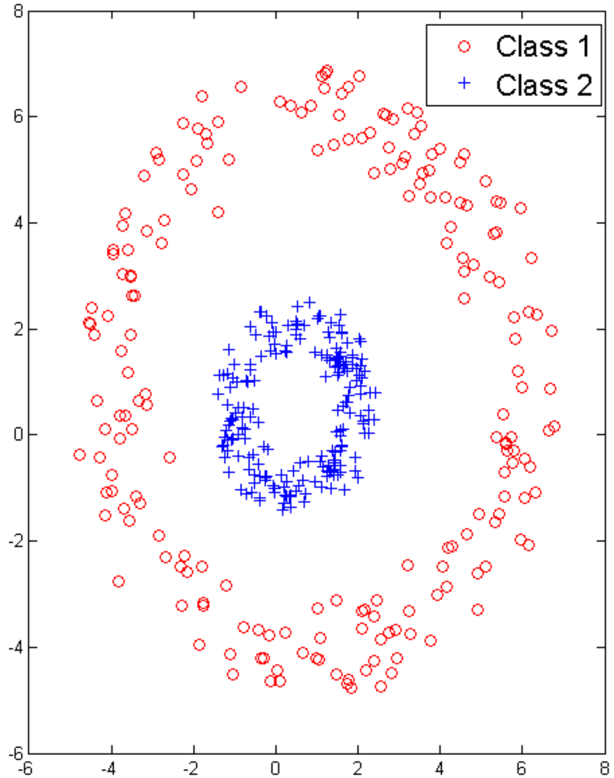
- Map data into a feature space where they are linearly separable.

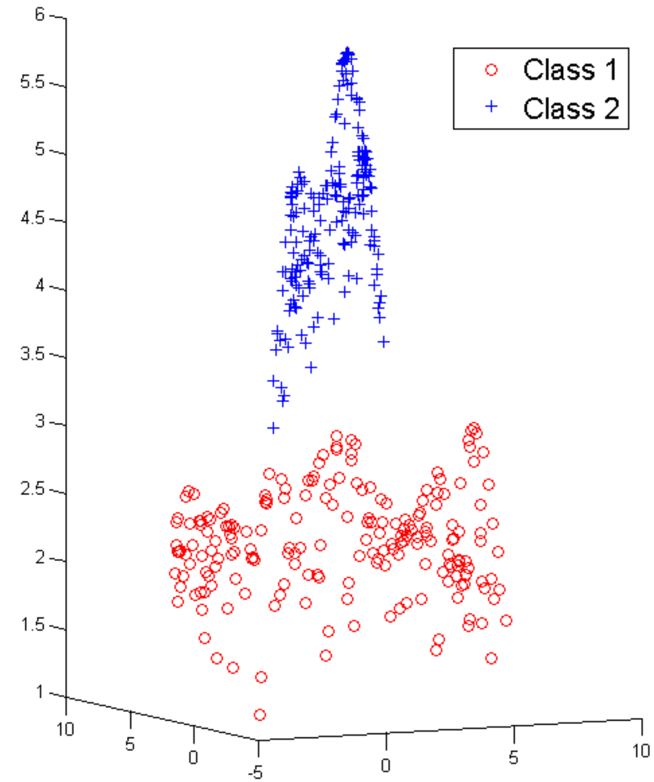




From Presentations by Nello Cristianini and Jason Weston respectively

Support Vector Machines





Decision Trees

- Decision trees repeatedly split the dataset in order to arrive at a final outcome.
- The split is made into branch-like segments and these segments progressively form an inverted tree, which originates from the starting node called the root.
- The root is the only node in the tree which does not have an incoming segment.
- The tree terminates at the decision nodes, also known as leaves or terminals.
- All the other nodes present within the tree are called internals or test nodes.

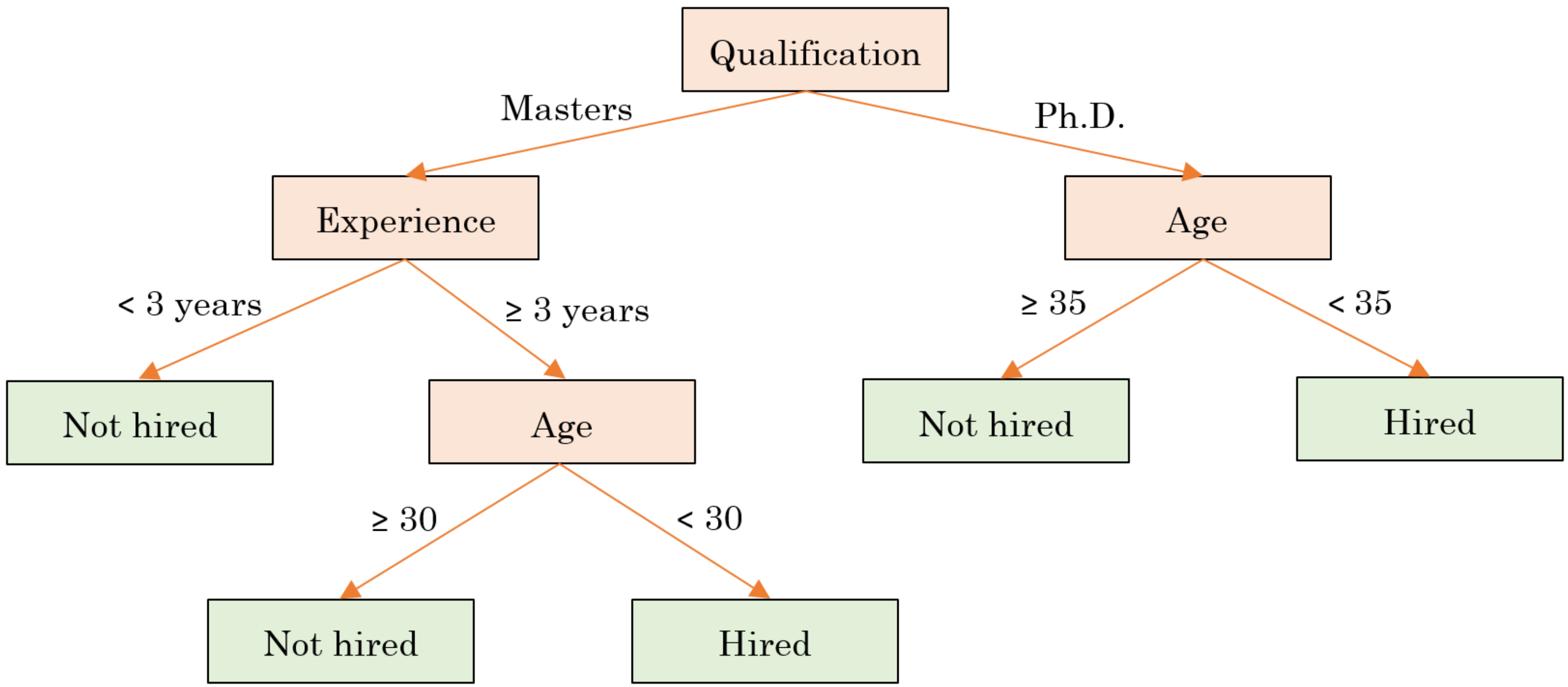
Decision Trees

- The variables or features associated with the data are used to make each split.
- At each node, the variables are tested to determine the most suitable variable to make the split.
- This testing is repeated on reaching the next node and progressively forms a tree.
- Each terminal node corresponds to a target class.

Example

- Suppose that a person A, aged 33 years and having a Ph.D. in his field along with an experience of 2 years, wants to apply for a job at a company.
- He has the hiring history of the company and knows that the important points checked in the interview are qualification, age and experience (these are the variables or features associated with the applicants).
- Based on the known data (training data) he develops a decision tree, in order to check whether he will be able to get the job or not.

Example



Qualification
Masters
Ph.D.
Experience
Age
< 3 years
≥ 3 years
≥ 35
< 35
Not hired
Age
Not hired
Hired
≥ 30
< 30
Not hired
Hired

Random Forest

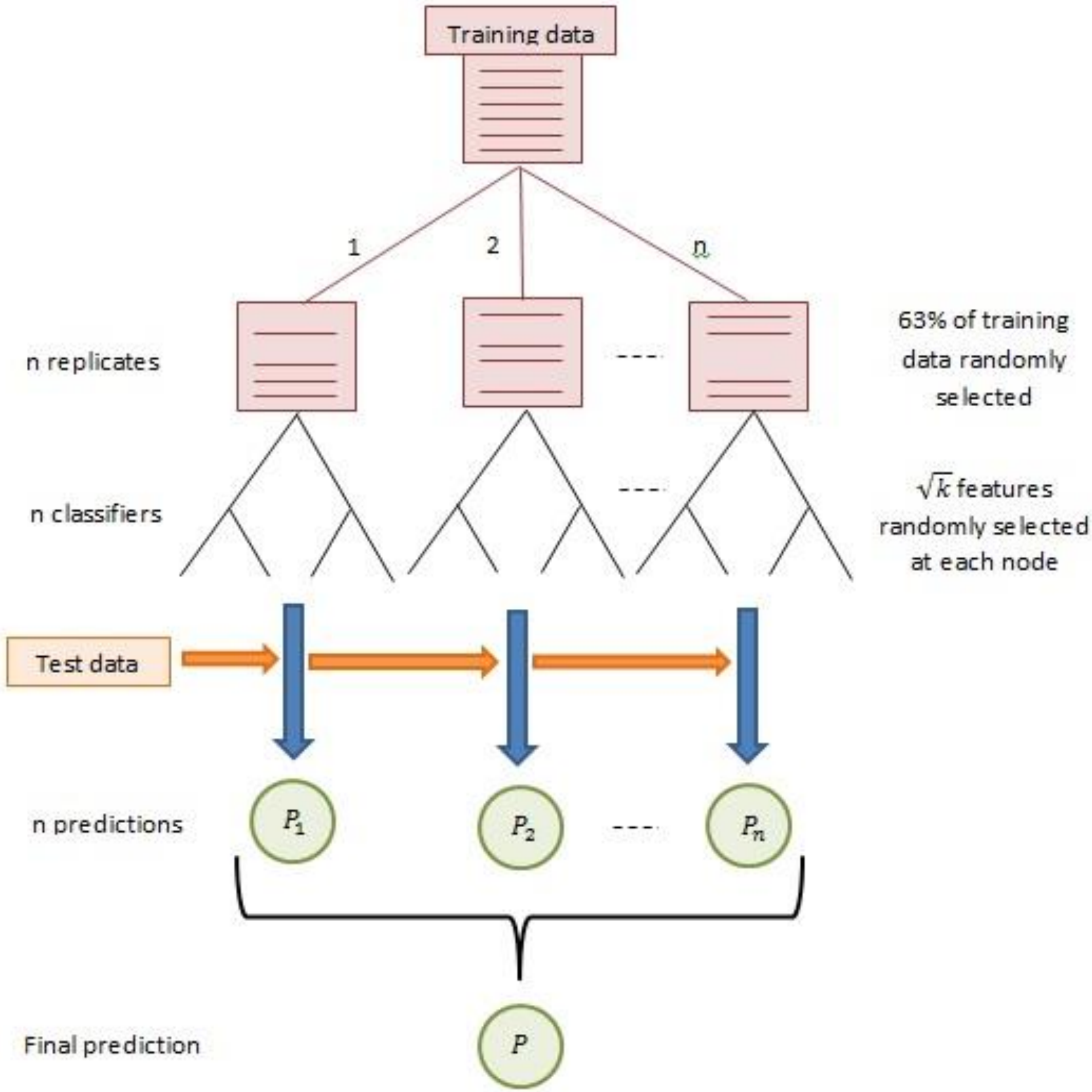
- A bootstrap sample of the dataset is used to independently grow individual trees, and the majority vote is taken to conclude the final prediction.
- In addition to using a randomly selected bootstrap sample of the data for growing each tree in the forest, randomness was introduced in the splitting of nodes.
- In standard trees, each node in the classification or regression tree is split using the best split among all variables.
- In random forests, each node is split using the best among a subset of the variables randomly selected at that node.

Random Forest

- Suppose n number of trees are grown. Each tree is generated by randomly selecting nearly 63% of the given training data.
- The sample data is therefore different for each tree. The remaining 37% data, known as out of bag (OOB) data, is used to estimate the error rate.
- The trees are fully grown without any requirement of pruning, which is one of the advantages of random forest.

Random Forest

- At each node a subset of variables or features is selected and the most suitable feature among them is used for the split.
- The size of subset is a variable which is generally taken as $\sqrt{k}$ where k is the total number of features.
- Once the forest is grown by using the labelled training dataset, the test data is introduced for the prediction.
- The individual predictions by the trees are aggregated to conclude the final prediction result (i.e. majority vote for classification and average for regression).


Training data
1
2
n
n replicates
63% of training data randomly selected
n classifiers
$\sqrt{k}$ features randomly selected at each node
Test data
n predictions
P_1
P_2
P_n
Final prediction
P

Concluding Remarks

- Machine Learning is invading Civil Engineering
- Applications are endless
- Travel Mode Detection
- Trip Purpose Detection
- Pile Bearing capacity Prediction
- Rainfall Runoff Prediction

awais.shafique@ucp.edu.pk

facebook.com/**m.awaisshafique**

Jp.linkedin.com/in/awaisshafique/

www.researchgate.net/profile/Muhammad_Shafique19