Machine Learning and Applications

Examination Scheme

Subject Code	Subject	Teaching Scheme			Examination Scheme					Total Marks	Credits
		Lecture	Practical	Tutorial	In-Sem	TW	PR	OR	End-Sem		
414453	Information and Cyber Security	3	--	--	30	--	--	--	70	100	3
414454	Machine Learning and Applications	4	--	--	30	--	--	--	70	100	4
414455	Software Design and Modeling	3	--	--	30	--	--	--	70	100	3
414456	Elective-I	3	--	--	30	--	--	--	70	100	3
414457	Elective -II	3	--	--	30	--	--	--	70	100	3
414458	Computer Laboratory-VII	--	4	--	--	50	50	--	--	100	2
414459	Computer Laboratory-VIII	--	4	--	--	50	--	50	--	100	2
414460	Project Phase-I	--	--	2	--	--	--	50	--	50	2
414461	Audit Course-V	--	--	--	--	--	--	--	--	Grade	

Machine Learning & Applications

Course Objectives:

- Understanding Human learning aspects.
- Understanding primitives and methods in learning process by computer.
- Understanding the nature of problems solved with Machine Learning

Machine Learning & Applications

Course Outcomes:

- Model the learning primitives.
- Build the learning model.
- Tackle real world problems in the domain of Data Mining and Big Data Analytics, Information Retrieval, Computer vision, Linguistics and Bioinformatics.

Unit I INTRODUCTION TO MACHINE LEARNING

- Introduction: What is Machine Learning, Examples of Machine Learning applications, Training versus Testing, Positive and Negative Class, Cross-validation.
- Types of Learning: Supervised, Unsupervised and Semi-Supervised Learning.
- Dimensionality Reduction: Introduction to Dimensionality Reduction, Subset Selection, Introduction to Principal Component Analysis.

Unit II Classification

- Binary and Multiclass Classification: Assessing Classification Performance, Handling more than two classes, Multiclass Classification-One vs One, One vs Rest Linear Models: Perceptron, Support Vector Machines (SVM), Soft Margin SVM, Kernel methods for non-linearity

Unit III REGRESSION AND GENERALIZATION

- Regression: Assessing performance of Regression – Error measures, Overfitting and Underfitting, Catalysts for Overfitting, VC Dimensions
- Linear Models: Least Square method, Univariate Regression, Multivariate Linear Regression, Regularized Regression – Ridge Regression and Lasso
- Theory of Generalization: Bias and Variance Dilemma, Training and Testing Curves Case Study of Polynomial Curve Fitting.

Unit IV LOGIC BASED AND ALGEBRAIC MODELS

- Distance Based Models: Neighbors and Examples, Nearest Neighbor Classification, Distance based clustering algorithms – K-means and K-medoids, Hierarchical clustering.
- Rule Based Models: Rule learning for subgroup discovery, Association rules mining – Apriori Algorithm, Confidence and Support parameters.
- Tree Based Models: Decision Trees, Minority Class, Impurity Measures – Gini Index and Entropy, Best Split.

Unit V PROBABILISTIC MODELS

- Conditional Probability, Joint Probability, Probability Density Function, Normal Distribution and its Geometric Interpretation, Naïve Bayes Classifier, Discriminative Learning with Maximum Likelihood. Probabilistic Models with Hidden variables: Expectation-Maximization methods, Gaussian Mixtures

Unit VI TRENDS IN MACHINE LEARNING

- Ensemble Learning: Combining Multiple Models, Bagging, Randomization, Boosting, Stacking
- Reinforcement Learning: Exploration, Exploitation, Rewards, Penalties
- Deep Learning: The Neuron, Expressing Linear Perceptron as Neurons, Feed Forward Neural Networks, Linear Neurons and their Limitations, Sigmoid, Tanh and ReLU Neurons

Unit I INTRODUCTION TO MACHINE LEARNING

- Introduction: What is Machine Learning, Examples of Machine Learning applications, Training versus Testing, Positive and Negative Class, Cross-validation.
- Types of Learning: Supervised, Unsupervised and Semi-Supervised Learning.
- Dimensionality Reduction: Introduction to Dimensionality Reduction, Subset Selection, Introduction to Principal Component Analysis.

Introduction

- Age of Big Data
- Buzz words ……..
- Machine learning, Data science, Data mining, Data analysis, Statistical learning , pattern discovery.
  - They all fall under the same umbrella which is Learning from data.

Importance

- Data today is everywhere.
- For example,
- Google processes 24 petabytes of data per day.
- Facebook processes ten millions of photo every hour.
- YouTube, we have about one hour of video uploaded every second.
- Twitter, about 400 million tweets per day.
- And in astronomy, for example, satellites data is in hundreds of petabytes.
- It is estimated that by 2020, the digital universe will reach 44 zettabytes of data.
- That is 44 trillion gigabytes.

- So data comes in different types and flavors.
- Data could be text, it could be numbers, click streams, graphs, tables, images, transactions, videos, and sometimes all of the above

- The data science process consists of five main steps.
- The first one is data collection.
- The second step is data preparation.
- The third step is exploratory data analysis.
- Machine learning is the core step in data science in which we deploy machine learning methods and statistics methods to get knowledge and to learn models from the data.
- Evaluate the Model

- Machine learning is the systematic study of algorithms and systems that improve their knowledge or performance with experience.

When to use Machine Learning

- You can not code the rules
- You can not scale
- Dynamic environment

Definition

- Tom Mitchell gave a “well-posed” definition that has proven more useful to engineering types
- *“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.”*

Applications

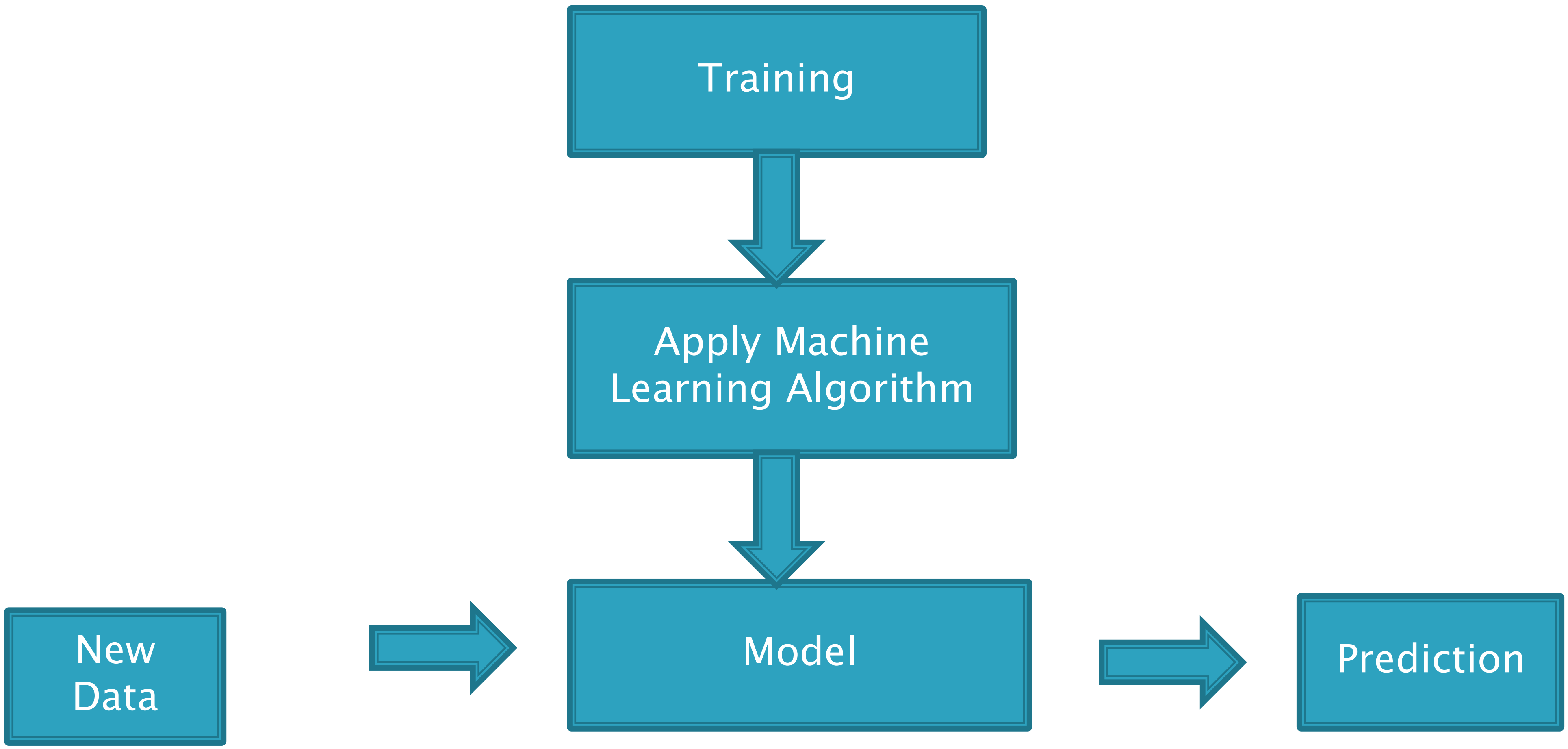
- Learning Associations
  - Basket Analysis
- Classification
  - Spam Filters
  - Loan Approval
  - Pattern Recognition(handwritten characters)
  - Face recognition
  - Analyzing Images and MRI images(medical diagnosis)
  - Speech recognition
  - Biometric recognition

- Regression
  - Prediction of the value(price of used car)
- Reinforcement Learning
  - Games
  - Robotic Navigation
-

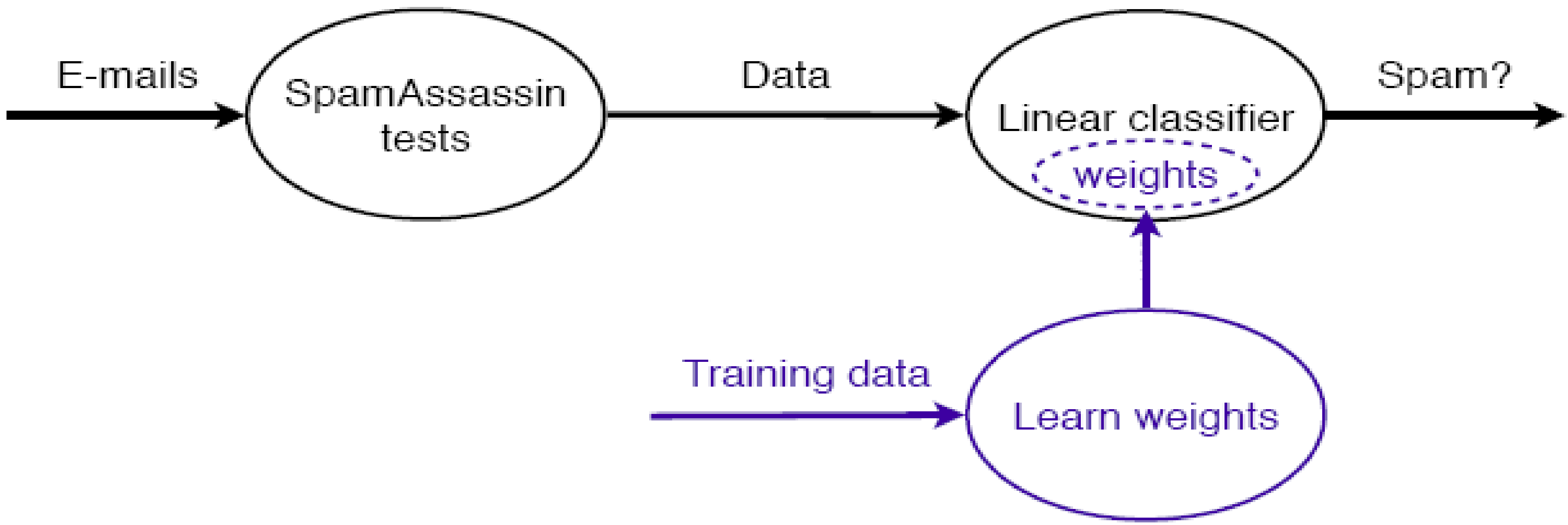
Applications

- Search Engines
- Handwritten character recognition
- Spam Filters
- Image Recognition
- Analyzing Images and MRI images
- Credit card fraud detection
- And many more…….

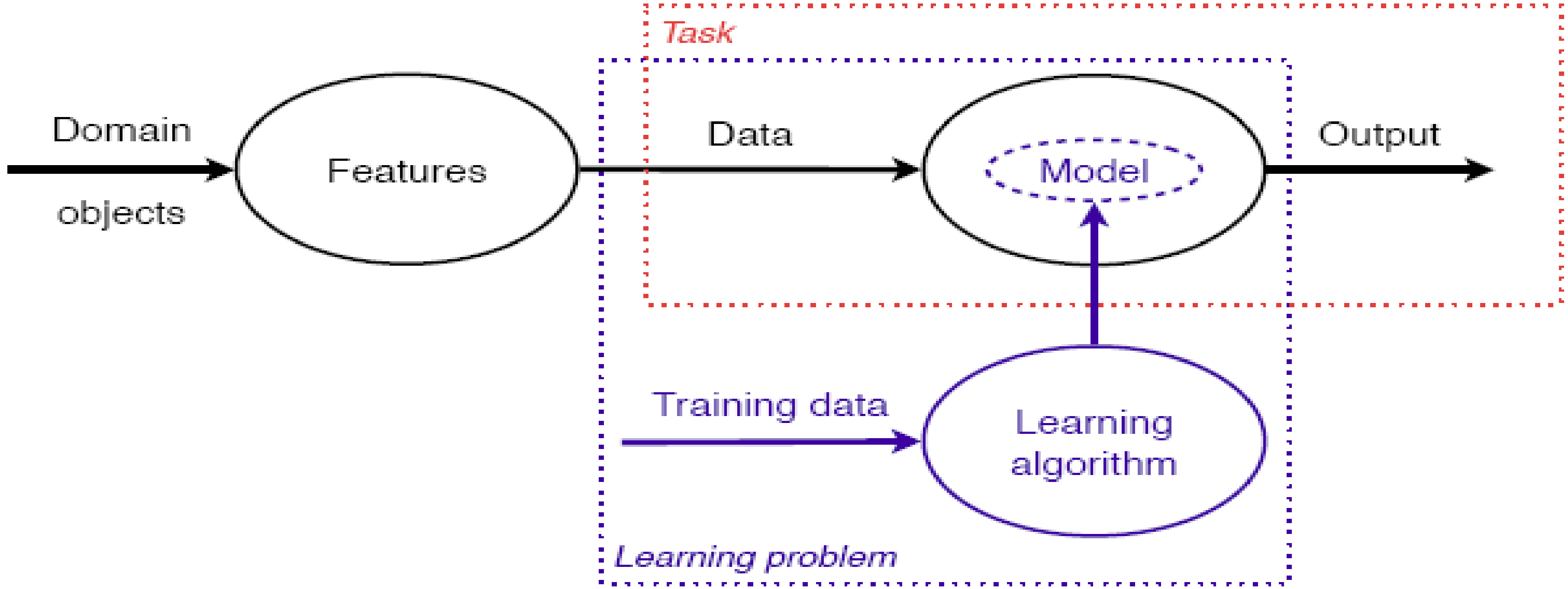
Machine Learning Process Flow



Machine learning for spam filtering



How machine learning helps to solve a task

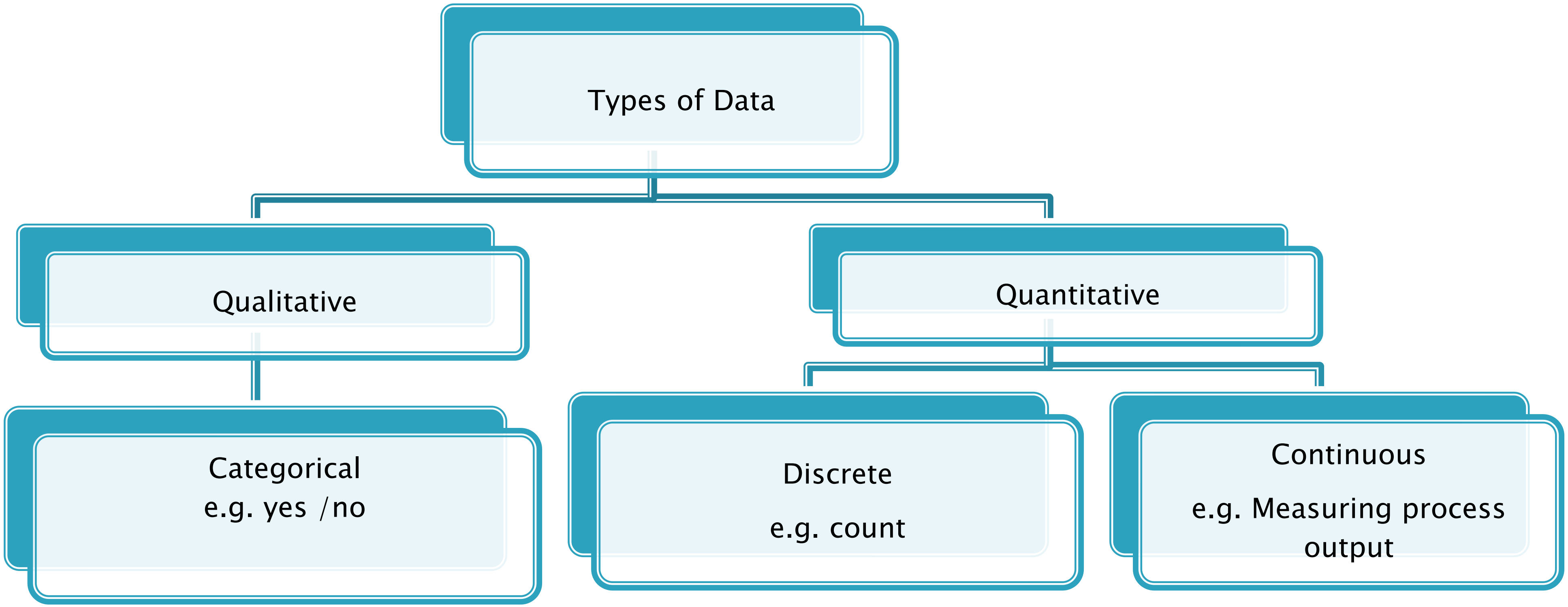


- A task (red box) requires an appropriate mapping – a model – from data described by features to outputs.
- Obtaining such a mapping from training data is what constitutes a learning problem (blue box).

Learning -Types

- Supervised
  - Labeled data
  - Positive and negative class
- Unsupervised
  - Unlabeled data
- Semi-supervised

- Reinforcement
  - Output of system is sequence of actions.


Types of Data
Qualitative
Quantitative
Categorical
e.g. yes /no
Discrete
e.g. count
Continuous
e.g. Measuring process output

Scales of Measurement

- Nominal Scale
  - Qualitative /Categorical Data
  - Names, Color, Gender etc.
  - Order does not matter
- Ordinal Scale
  - Ranking
  - Order is important
  - Differences can not be measured
- Interval Scale
  - Order Matters
  - Differences can be measured
  - No absolute zero
- Ratio Scale
  - Order Matters
  - Differences can be measured
  - Absolute zero

Features

- Important building blocks of datasets
- Preparing the proper input dataset, compatible with the machine learning algorithm
- Improves the performance
- Feature Transformation
  - Handling Missing Values
  - Normalization
  - Label Encoding

Training versus Testing

- Machine needs to be trained by explicitly feeding it data that has correct answers attached to it.
- For this we need the data set which we call Training data set
- To check the performance of model we use Test data set.
- Validation data set

- Task
  - Abstract representation of a problem
  - Ex- two or more classes
- Features
  - A Language in which we describe relevant objects in our domain.
- Model
  - Output of machine learning algorithm applied to training data

Curse of Dimensionality

- Features are very important in any Machine Learning Algorithm
- Curse of Dimensionality means that error increases with increase in number of features
- Increase in dimensionality affects the performance of the algorithm
- Solution
  - Dimensionality Reduction
    - Feature Selection
      - Forward Selection
      - Backward Selection
    - Feature Extraction
      - PCA , LDA

Dimensionality Reduction

- In most learning algorithms, the complexity depends on the number of input dimensions, *d, as well as on the size of the data sample, N*
- For reduced memory and computation, decreasing *d also decreases the* complexity of the inference algorithm during testing.
- Save the cost of extracting
- Simpler models are more robust on small datasets

Cross-validation

- Replication requirement
- To generate number of training and validation set pairs from dataset
- Repeated use of the same data split differently is called cross validation
- Generate K training/validation set pairs from the given data set.

$$\{\mathcal{T}_i, \mathcal{V}_i\}_{i=1}^{K}$$

K-fold cross validation

- In this the dataset is divided randomly into K equal sized parts X*i, i = 1, . . . , K.*
- To generate each pair, keep one of the *K* parts out as the validation set and combine the remaining *K − 1 parts to* form the training set. Doing this *K times, each time leaving out another* one of the *K parts out, we get K pairs:*

$$\begin{aligned}
\mathcal{V}_1 &= \mathcal{X}_1 & \mathcal{T}_1 &= \mathcal{X}_2 \cup \mathcal{X}_3 \cup \cdots \cup \mathcal{X}_K \\
\mathcal{V}_2 &= \mathcal{X}_2 & \mathcal{T}_2 &= \mathcal{X}_1 \cup \mathcal{X}_3 \cup \cdots \cup \mathcal{X}_K \\
&\vdots \\
\mathcal{V}_K &= \mathcal{X}_K & \mathcal{T}_K &= \mathcal{X}_1 \cup \mathcal{X}_2 \cup \cdots \cup \mathcal{X}_{K-1}
\end{aligned}$$

- To get a better idea about the process that underlies the data and this allows knowledge extraction.
- When data can be represented in a few dimensions without loss of information, it can be plotted and analyzed visually for structure and outliers.

- Two main methods for reducing dimensionality
  - Feature Selection
    - Finding k of the d dimensions which provide the most information
    - Subset Selection method
  - Feature extraction
    - finding a new set of k dimensions that are combinations of the original dimensions.
    - PCA
    - LDA

Subset Selection

- Focus is on in finding the best subset of the set of features.
- The best subset contains the least number of dimensions that most contribute to accuracy.
- We discard the remaining, unimportant dimensions.
- Two approaches
  - Forward selection
  - Backward selection

Forward Selection

- Start with no variable
- Add them one by one
- At each step add the one that decreases the error most
- Repeat until any further addition does not decrease the error

- F : set of feature set of input dimensions x_i , $i=1,2,\ldots d$
- E(F) : the error incurred on validation sample when only F inputs are used
- In sequential forward selection
  - $F= \phi$
  - Train the model for all possible x_i
  - Calculate $E(F \cup x_i)$ and choose such input that causes the least error
-

$$j = \arg\min_i E(F \cup x_i)$$

and we

add x_j to F if $E(F \cup x_j) < E(F)$

- This algorithm is called Wrapper approach

Backward Selection

- Start with all variable
- Remove them one by one
- At each step remove the one that decreases the error most
- Repeat until any further removal increases the error significantly

Backward Selection

$$j = \arg\min_i E(F - x_i)$$

and we

remove x_j from F if $E(F - x_j) < E(F)$

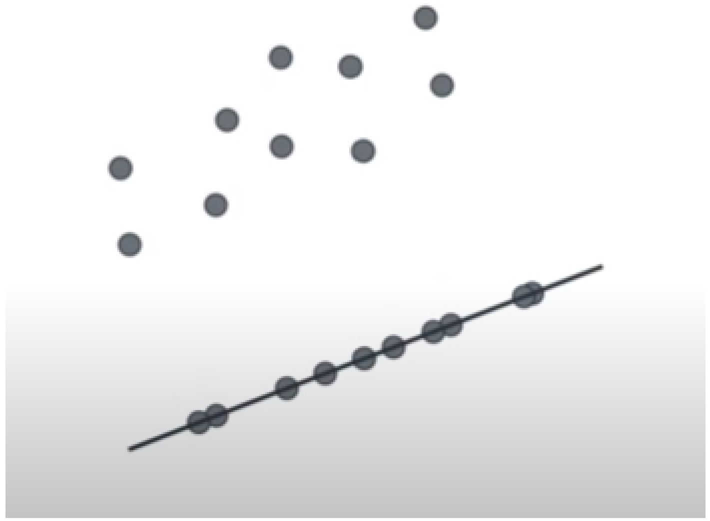
Feature Extraction

- Finding a new set of k dimensions that are combinations of the original dimensions.
- PCA
- LDA

Introduction to Principal Component Analysis

- Best known feature construction method
- Principal component analysis is a method of extracting important variables (in form of components) from a large set of variables available in a data set.
- New features are constructed as linear combinations of given features

- It extracts low dimensional set of features from a high dimensional data set with a motive to capture as much information as possible.
- With fewer variables, visualization also becomes much more meaningful.
- PCA is more useful when dealing with 3 or higher dimensional data.



Size
Number of rooms
Number of bathrooms
Schools around
Crime rate

- Dimensionality reduction method
- Steps
  - Data Preparation of the Data (Standardization)
    - Subtract mean from each variable(center the data)
    - Produces data set with mean is zero.
    - Scale the Data(divide each variable by its standard deviation)
    - Scale function in R
  - Calculate the covariance/correlation matrix
    - Measures how dimensions vary w.r.t. each other
    - Cor function in R

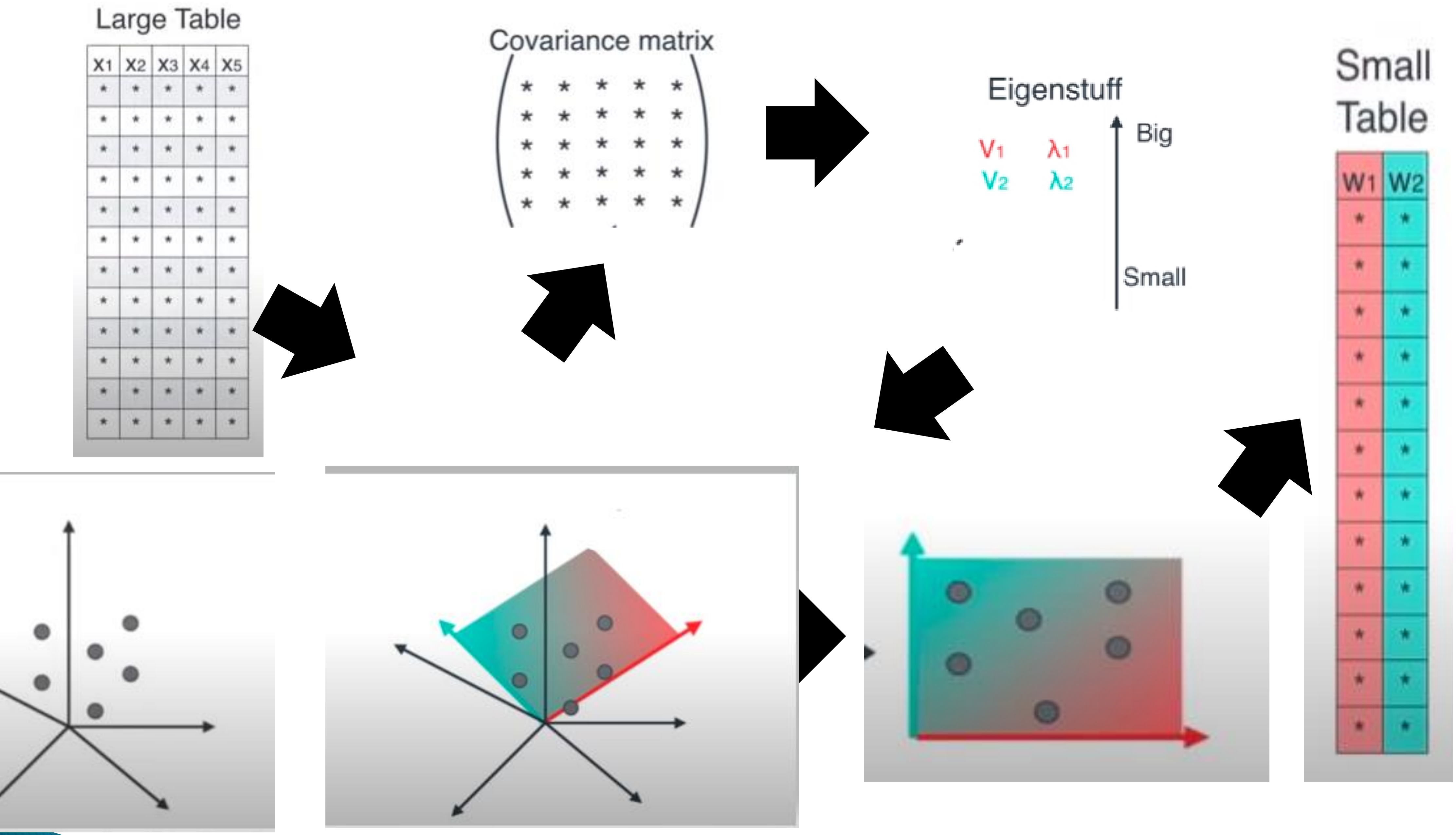
- $cov(x,x) = \sigma_{xx}^2 = \frac{\sum_i (x_i - \mu_x)(x_i - \mu_x)}{n-1}$
- $cov(y,y) = \sigma_{yy}^2 = \frac{\sum_i (y_i - \mu_y)(y_i - \mu_y)}{n-1}$
- $cov(x,y) = \sigma_{xy}^2 = \frac{\sum_i (x_i - \mu_x)(y_i - \mu_y)}{n-1}$

$$\begin{pmatrix} \mathrm{Var(X)} & \mathrm{Cov(X,Y)} \\ \mathrm{Cov(X,Y)} & \mathrm{Var(Y)} \end{pmatrix}$$

  - Calculate eigenvectors /eigenvalues
  - Eigen function in R
  - Highest eigen value corresponds to first principal component

- Compute the Data Set
  - Transpose the eigenvectors
  - Transpose the adjusted data
  - new data= eigenvectors_t * adjusted_data_t

- Construct Covariance matrix
- Compute eigenvectors of the COV matrix
- Eigenvectors corresponding to largest eigenvectors are used to reconstruct the large fraction of the data set
- Eigenvectors form principal axes onto which the data values are projected
- Retain only those that account for most of the variance


Large Table
X1 X2 X3 X4 X5
Covariance matrix
Eigenstuff
V1 λ1
V2 λ2
Big
Small
Small Table
W1 W2

- prcomp()
- princomp()